Robust comparisons of socio-economic well-being

by

Tahsin Fazle Mehdi

The thesis is submitted in part fulfillment of the requirements for the award of
the degree of Doctor of Philosophy by Publications
of the University of Portsmouth

July 27, 2020

Declaration

'Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.'

Word count: 8,666 words for the commentary

Abstract

This body of research presents robust approaches for measuring and comparing socio-economic well-being across distributions, and stresses the importance of the need for more accurate statistical inference. It contains six separate, but closely related research papers categorized into two broad themes: Theme 1: Income-based well-being indicators (Paper 1A, Paper 1B, and Paper 1C), and Theme 2: Multidimensional well-being indicators (Paper 2A, Paper 2B, and Paper 2C). Theme 1 research papers develop asymptotic frameworks for comparing inequality and poverty using income as the well-being indicator. Paper 1A and Paper 1B contribute to the economics literature by providing an alternative, and in some cases, more powerful way to undertake hypothesis testing on income distributions that does not require the derivation of complex covariance structures. Paper 1C develops the asymptotic framework for comparing poverty between distributions that share a common relative poverty line—it is argued that this approach permits for more valid comparisons as opposed to alternate approaches which assume separate poverty lines for each distribution. Theme 2 research papers go beyond material standard of living, and examine well-being in a more generalized multidimensional setting using objective as well as subjective welfare indicators. Paper 2A and Paper 2B utilizes the data-driven technique of stochastic dominance efficiency to assess the equal-weighting schemes of Save the Children UK's Child Development Index and OECD's Better Life Index, which are two composite indices used for monitoring policy and making cross-country comparisons. Paper 2C examines well-being with respect to the labour market. Inspired by international literature, over 20 indicators are utilized to assess job quality in Canada, an OECD country for which a comprehensive job quality study has not been done in the past. Paper 2C fills this gap in the literature and contributes to Canadian labour policies with respect to precarious work.

Contents

Acknowledgements

This document was prepared under the supervision of Professor Lester C. Hunt whom I would like to thank for all the support and guidance. His thoughtful comments and suggestions lead to considerable improvement of this commentary. I would also like to thank my family for their endless support and encouragement.

List of papers submitted

Theme 1: Income-based well-being indicators

Paper 1A

Mehdi, T. (2013). Weighted empirical likelihood-based inference for quantiles under stratified random sampling. *Economics Bulletin, 33*, 2437–2442.

Paper 1B

Mehdi, T. & Stengos, T. (2014). Empirical likelihood-based inference for the generalized entropy class of inequality measures. *Economics Letters, 123*, 54–57.

Paper 1C

Mehdi, T. (2017). Poverty comparisons with common relative poverty lines. *Communications in Statistics - Theory and Methods, 46*, 2029–2036.

Theme 2: Multidimensional well-being indicators

Paper 2A

Mehdi, T. (2019). Stochastic dominance approach to measuring child development. *Child Indicators Research, 12*, 1567–1588.

Paper 2B

Mehdi, T. (2019). Stochastic dominance approach to OECD's Better Life Index. *Social Indicators Research, 143*, 917–954.

Paper 2C

Chen, W.-H. & Mehdi, T. (2019) Assessing job quality in Canada: A multidimensional approach. *Canadian Public Policy, 45*, 173–191.

Introduction

There is growing consensus that individual well-being is irreducible to a single factor (e.g., Boarini & D'Ercole, 2013; Fleurbaey, 2012). The well-being of individuals is something that is intrinsic to a prosperous nation. However, there are differences of opinion as to what exactly constitutes "well-being" as pointed out by Bourguignon & Chakravarty (2003) and Boarini & D'Ercole (2013).

At the root of it, there are essentially two strands of literature that deal with socio-economic well-being: philosophical and empirical. Philosophical discussions surrounding well-being address the importance of individual differences in capabilities, multivariate nature of what it means to be "happy", balance of materialistic and non-materialistic factors, and distribution of opportunities (e.g., Sen, 1976). The empirical spectrum of the literature considers such factors and focuses on quantifying or "putting a number" on well-being (e.g., Maasoumi, 1999; Zheng, 2001). This body of research is aimed at the latter.

In the absence of an optimal measure of well-being, researchers who focus on the empirical side of the literature, have and continue to develop indicators aimed at quantifying it (e.g., OECD, 2017). Such ongoing works have attracted the attention of international organizations such as the Organisation for Economic Co-operation and Development (OECD), Save the Children, United Nations (UN), and World Bank, all of which became the forefront for developing well-being measures. Given the plethora of well-being measures that are available, the need for more reliable and accurate ways of assessing them becomes vital.

Indicators of well-being can be either objective or subjective. The former is usually measured by income, consumption, or wealth. Subjective indicators are those based on individual perceptions (e.g., self-reported life satisfaction, or self-reported health) and capture aspects of quality of life unaccounted for by objective indicators such as income (see, e.g., Boarini et al., 2012; Boarini & D'Ercole, 2013).

Although there is recognition that poverty and inequality—important aspects of well-being—are multi-faceted phenomena, it was primarily income that gave rise to the concepts (e.g., Atkinson, 1970, 1987; Davidson & Duclos, 2000; Duclos et al., 2006; Thompson, 2012; Zheng, 2001). Poverty places emphasis solely on the "poor" while inequality involves measuring how even or uneven resources (e.g., income) are distributed in society (see, e.g., Bourguignon & Chakravarty, 2003; Foster, 1984; Sen, 1976; Zheng, 1997).

Early works on poverty such as Foster et al. (1984) and Sen (1976) focused mainly on the identification of the poor. That is, they were mainly concerned with where to set the poverty line—an income level below which an individual is deemed poor. There are two types of poverty lines: absolute and relative. The former is a fixed income level usually determined by experts in conjunction with public officials. Relative poverty lines are based on some fraction of the mean or a quantile of the underlying distribution (e.g., 50% of the mean, 50% of the median, etc.).

There have been several classes of poverty measures put forth in the literature (e.g., Watts, 1968; Sen, 1976; Clark et al., 1981; Foster et al., 1984) and it is not often clear which measures to use when analyzing poverty outcomes. However, such concerns may be alleviated by taking an axiomatic approach to poverty measurement. That is, the choice of poverty measures can be partially guided by some well-known axioms of poverty set out in the literature (e.g., Sen, 1976). The first widely known axiom is the "focus axiom" which requires poverty measures to be insensitive to incomes above the poverty line. The second axiom is the "monotonicity axiom" which states that if a person below the poverty line experiences an increase in their income, the poverty measure should decrease accordingly. The third axiom known as the "transfer axiom" requires the poverty measure to decrease after a progressive income transfer (and increase after a regressive income transfer).

Another aspect of poverty measures to consider is whether or not they are decomposable. That is, whether or not the poverty measure is additively separable. The Foster et al. (1984) class of poverty measures are of this type and an appealing feature of additively separable poverty indices is that they allow analysts to study the quantitative and qualitative assessment of the effect of changes in subgroup poverty on total poverty. This is useful when studying poverty outcomes along geographic, ethnic, or other lines.

While identification of the poor is important, recent works such as Biewen (2002), Davidson & Duclos (2000), and Davidson (2007) have highlighted the need for more accurate statistical inference when measuring poverty, especially since such metrics are typically estimated from household surveys which can have rather small sample sizes, especially in developing countries (see, e.g., Biewen, 2002).

Inequality is something that is closely related to poverty but it is conceptually different and has its own strand of literature. The concept can be subject to different economic interpretations. Economic inequality, in particular, mainly refers to inequality in access to economic resources. And "economic resources" are typically measured by income so income inequality indices are summary statistics for measuring dispersion of incomes in a given society. Much like poverty, there have been numerous income inequality indices proposed in the literature (Cowell, 2000). A simple way to measure income inequality is to sort individuals according to their income and then calculate ratios of the different income quantiles such as the 90th percentile to the 10th to track the interdecile income gap. Another simple approach is to track the share of income held by different income groups like the top 1%, top 10%, bottom 50%, etc. Some examples of indices include the Gini (a ratio based on the Lorenz curve), coefficient of variation, or the generalized entropy class of inequality measures. And similar to poverty, income inequality indices can also be chosen based on axiomatic grounds. The first widely known axiom is the "transfer principle" (also known as the Pigou-Dalton principle, e.g., Pigou, 1912; Dalton, 1920) which requires the inequality index to decrease when a progressive transfer occurs (i.e., transfer of income from a rich person to a poor person). The second axiom is "scale-invariance" which requires the inequality measure to be invariant to scaling of the original income vector (i.e., doubling every person's original income level should not change the inequality index). The third axiom of income inequality indicators is "translation invariance" which require the index to be invariant to uniform additions or subtraction to original income levels. The fourth axiom is the "principle of population" which requires the inequality measure to be invariant to replications of the original population (see, e.g., Cowell & Kuga, 1981).

Much like poverty measures, inequality measures can also have the decomposability feature. However, this feature as it pertains to inequality measures differs from poverty measures. Decomposable inequality measures, such as the generalized entropy class of measures, have the desirable property that the total measured inequality can be decomposed into their "within" and "between" group components. This feature is particularly useful for analysts interested in studying inequality amongst certain subgroups of a population. The decomposability feature allow analysts to quantify exactly what portion of total inequality are due to inequality within those subgroups and what portion are due to inequality between those subgroups.

Income-based approaches to measuring poverty and inequality ultimately rely on income quantiles as a starting point since quantiles indicate an individual's position relative to others in an income distribution. For example, the interdecile ratio of the 90th percentile income to the 10th percentile income is a common indicator of income inequality (e.g., OECD, 2010). As such, there is a need for reliable and accurate statistical inference for estimates derived from sample data.

Prior to producing indicators, it is essential to be familiar with how the data was collected. Conceptually, simple random sampling is the most straightforward household survey design. But often, complex survey designs are favoured in order to alleviate sampling bias (i.e., guard against over-sampling or under-sampling of certain populations or households).

Stratification and clustering are two of the more popular complex household survey designs. Each type of design can be done either through single-stage or multi-stage sampling. The complexity induced by such survey designs invalidate the typical standard errors synonymous with classical statistics. Careful adjustments must be made to the estimates, depending on the survey design and weights, in order to draw statistical inferences on indicators. There has been a wide array of research devoted to accurate statistical inference on estimates derived from complex surveys (e.g., Gross, 1980; Deaton, 1997; Chen & Sitter, 1999; Bhattacharya, 2005, 2007). Such works highlight the methodological considerations that must be made when dealing with data arising from complex survey designs and show the consequences of ignoring the special survey structures. When measuring income inequality, in particular, it is quite important to be aware of potential sorting that goes on between rich and poor areas and ignoring this in a survey design could result in misleading estimates and conclusions (Bhattacharya, 2007).

Theme 1, which comprises Paper 1A, Paper 1B, and Paper 1C, presents statistical inference methods for robust income-based measures of poverty and inequality. Theme 1 explores the following research questions:

1i) How can hypothesis tests regarding quantiles that arise from stratified random samples be undertaken?

1ii) How can inequality between income distributions be reliably compared?

1iii) Can poverty comparisons between distributions be made if the poverty line depends on the pooled distribution?

Paper 1A introduces an empirical likelihood-based statistical inference method for testing hypotheses regarding quantiles arising from stratified random samples—one of the most common household sampling schemes (see Chen & Sitter, 1999; Francisco & Fuller, 1991; Zhong & Rao, 2000). Paper 1B builds on empirical likelihood and proposes a method for comparing a wide class of decomposable inequality indices among two populations. Paper 1C shifts the focus to poverty which concerns only the portion of the income distribution below a chosen poverty line. While existing approaches for making relative poverty comparisons between distributions assume separate poverty lines defined for each of the distributions (e.g., Zheng, 2001), Paper 1C develops an asymptotic framework for making poverty comparisons between distributions that share a common relative poverty line (e.g., 50% of the pooled median income level).

There is growing recognition that well-being is a manifestation of more than income per capita or material standards of living as pointed out in works such as Atkinson & Bourguignon (1982), Fleurbaey (2012), and Maasoumi (1999). There is mounting evidence that individual happiness may not necessarily be tied to income (e.g, Dutta & Foster, 2013) and quality of life is paramount to social well-being. In light of this, international organizations have started initiatives aimed at better measuring well-being (e.g., Stiglitz et al., 2009; OECD, 2017; Alkire, 2020). Some examples of multidimensional indices include UN's Human Development Index and Multidimensional Poverty Index, World Bank's World Governance Indicators, OECD's Better Life Index (BLI), and Save the Children UK's Child Development Index (CDI). Such indices are constructed for a select set of countries for which data is available, and then these indices are used for measuring and monitoring social progress, and cross-country well-being comparisons (e.g., OECD, 2017; Save the Children UK, 2012).

Multivariate approaches to well-being entail selecting a set of indicators which measure well-being and then assessing the dimensions separately without imposing any hierarchy across the dimensions, or by aggregating them into a single summary index. However, multivariate approaches to measuring well-being have their own set of limitations and challenges (Alkire et al., 2015; Alkire, 2020). Identification of the dimensions of well-being involves value judgment so it can be a rather arduous process. Even if the dimensions are agreed upon, it is not often clear how to communicate those indicators to the public. The "dashboard" approach, which involves assessing dimensions individually without giving preference to any one dimension, suffers from the possible heterogeneity of indicators. Since indicators maybe designed to measure vastly different dimensions of well-being, it becomes harder to communicate the disjoint nature of the indicators without a clear hierarchy system for the dimensions.

According to Stiglitz et al. (2009), one of the reasons why gross domestic product (GDP) came to be such a powerful and featured headline figure is that it is a single indicator that allows simple comparisons of socio-economic performance cross-nationally and across time. International composite indices have become highly influential tools for assessing and comparing achievements across countries and across time, and go some ways towards addressing the heterogeneity and communications challenges of the "dashboard" approach. Unlike that approach, composite indices do impose relative (non-negative) weights on each indicator and then convert the individual dimensional scores into a single real number which can be seen as appealing and easy to communicate, but there is nothing that makes such indices inherently more sophisticated than alternative approaches to multivariate measurement of well-being. A common disadvantage shared by both composite indices and the "dashboard" approach is that neither accounts for the joint distribution of the indicators. The weights imposed on the indicators govern trade-offs across aggregate dimensions and give a sense of hierarchy amongst the dimensions. However, such normative choices can be viewed as very demanding and have been called into question (e.g., Ravallion, 2011; Alkire et al., 2015). Nonetheless, there has been a proliferation of literature related to composite indicators just over the past two decades alone according to meth.

A significant number of multidimensional well-being indices are aggregated by equally weighting the achievement scores which is tantamount to taking the arithmetic mean of the indicators (see, e.g, Save the Children UK, 2008; OECD, 2017). Indeed, a single measure that synthesizes information on broad aspects of well-being is appealing for benchmarking policies and assessing welfare, but how each dimension is weighted or prioritized and the reliability of the orderings they produce remains a point of contention (see Greco et al., 2018; Maasoumi, 1999). While it is difficult to communicate disjoint indicators designed to measure very different aspects of well-being, it is straight-forward to rank countries on the basis of just a single indicator. What is not so straight-forward is comparing socio-economic performance of countries on the basis of composite indices which have normative judgments underlying them brought on by the arbitrary choice of weights imposed on each dimension governing their hierarchy.

Recognizing the multidimensional nature of well-being, Theme 2, which comprises Paper 2A, Paper 2B, and Paper 2C, delves into well-being by considering multiple indicators rather than just income. Theme 2 posits the following research questions:

2i) How sensitive are multidimensional well-being indices to the choice of weights?

2ii) Can any country achieve a good index score just by varying the indicator weights?

2iii) Is there a "best" way to select weights for the various dimensions of well-being?

2iv) Can multidimensional well-being be measured without weights?

Different weighting schemes can result in drastically different composite indices and ultimately very different rankings depending on the unit of analysis (e.g., countries, institutions, etc.). There are a number of different methods available for assessing weighting schemes and rank robustness (e.g., Alkire et al., 2015; Greco et al., 2018). Generalized mean aggregation methods shed light on the interactions between different well-being dimensions by letting researchers tune the substitutability and complementarity amongst dimensions (e.g., Pinar, 2019). Then there are methods aimed at checking sensitivity of weights to rank reversals. For example, Foster et al. (2013) examines empirically cases where the initial vector of a given composite index is changed that result in rank reversals and cases where the ranking is robust to plausible weight changes. Similarly, there are techniques that posit reasonable ranges for weights and begin by examining the "corners" where full weight is shifted to a single dimension and then weights are altered (e.g., Seth & McGiillivray, 2018). Since selection of well-being dimensions is inherently subjective, there is another strand of literature that considers imposing dimensional weights on the basis of societal preferences (e.g., Yang, 2018). While much of the discussion up to this point has been focused on exogenous treatment of weights, there are data-driven methods like principal components analysis (e.g., Biswas & Caliendo, 2002), factor analysis, I-distance (e.g., Markovic et al., 2016), or data envelopment analysis (e.g., Mizobuchi, 2014) that treat the weights as endogenous. Such methods are sometimes seen as more "objective" but they are only seen as such due to the fact that data-driven procedures are agnostic and rely solely on the raw indicator data.

The data-driven method known as stochastic dominance efficiency (SDE) is used to assess the official equal-weighting schemes of the popular CDI and BLI composite indices in Paper 2A and Paper 2B, respectively. The CDI, developed by Save The Children UK, is a composite index for measuring child development and has been produced periodically for nearly 100 countries since 2008. The BLI, developed by the OECD, is a general measure of composite well-being limited to OECD countries and has been published annually since 2011. SDE stems from the idea of stochastic dominance which is a robust ranking method that entails comparing entire cumulative distribution functions. While stochastic dominance is for making pairwise comparisons of distributions, SDE extends the method by permitting comparison of a linearly-weighted benchmark distribution against all other possible linear weighting schemes. Unlike other methods, those based on stochastic dominance (like SDE) have the advantage of making use of all moments of the underlying index distributions and SDE, specifically, permits comparison of a benchmark weighting scheme (e.g., equal-weighting) against all other possible weighting strategies. The implicit (endogenous) weights generated by SDE analysis allows the evaluation of two extreme weighting scenarios: a best-case scenario where the weights are allocated such that more countries achieve better measured outcomes based on the least variable combination of indicators, and a worst-case scenario where weights are distributed in a way that decreases measured cross-country performance. This should not be seen as synonymous with maximizing or minimizing true well-being as there is no way of quantifying actual well-being. Thus, SDE should be seen as a guidance tool for assessing the sensitivity and tendencies of actual measurable indicator data which are thought to approximate well-being. SDE assesses composite indices and allows researchers to identify the indicators of measured well-being leading to collective improvements and indicators which are difficult to improve—something that would be of great interest for economic and social policy. This permits greater transparency by revealing exactly how the imposed weights govern the trade-offs and inter-dependency amongst the dimensions. The end goal is not to prescribe a specific set of weights or preferences for well-being dimensions but rather it is to provide greater transparency and to underscore the need for more scrutiny of composite indicators. SDE in the context of multidimensional well-being introduces a novel approach and results from Paper 2A and Paper 2B, combined with earlier works of Pinar et al. (2013, 2017), Pinar (2015), Pinar et al. (2015), and Pinar et al. (2019) demonstrate promising progress for the field.

Paper 2C examines a specific aspect of well-being: work. Work-life balance is a significant part of well-being and more emphasis is increasingly being placed on the quality of jobs as opposed to the quantity of jobs (e.g., Drobnic et al., 2010). Organizations such as the OECD and Eurofound frequently publish job quality indicators for its member countries (see OECD, 2014; Eurofound, 2016). Eurofound administers the European Working Conditions Survey to a number of European countries in order to gather a unified set of indicators with which to assess job quality. Inspired by Eurofound's framework, Paper 2C takes a "dashboard" approach and utilizes data from the 2016 Canadian General Social Survey (CGSS) to construct over 20 job quality indicators for Canada—an OECD country for which a detailed job quality study has never been done. Paper 2C fills this gap in the international literature and makes a contribution to Canadian labour policies pertaining to precarious work.

Conditions under which this body of research was undertaken

Paper 1A and Paper 1B were the end results of research into economic inequality undertaken in the Department of Economics and Finance at University of Guelph. Professor Thanasis Stengos contributed to Paper 1B by providing editorial assistance. Paper 1A and Paper 1B support the department's reputation for research excellence in econometrics and economic measurement of well-being. Paper 1C was completed in the Department of Economics at University of Waterloo and contributes to the department's strengths in applied economics research.

Research presented in Paper 2A and Paper 2B were undertaken in the Department of Economics at Ryerson University and contributes to the department's research in the areas of econometrics and international economics. Paper 2C was the result of a collaboration with Wen-Hao Chen at Statistics Canada and contributes to one of the agency's mandates which entails analyzing and disseminating statistical information to the public. Paper 2C also contributes to the Government of Canada's broader policies surrounding the labour market as evidenced by the fact that Paper 2C became a topic of discussion in the Canadian House of Commons (see May, 2019).

Theme 1: Income-based well-being indicators

Paper 1A

Ever since the seminal work of Atkinson (1970), there has been a proliferation of literature surrounding income inequality—the measurement of how evenly or unevenly income is distributed in society (e.g., Cowell, 1989, 2011; Maasoumi, 1999; Davidson & Duclos, 2000; Thompson, 2010). Accurately characterizing and assessing income distributions has long been a subject of great interest among researchers and policy makers (e.g., Brachman et al., 1996; Maasoumi, 1999; McDonald, 1984). One of the most basic ways of characterizing an income distribution is to compute income quantiles which divide the distribution into equally-numbered units ranked according to income level (e.g., Gross, 1980). These threshold levels of income can then be compared with other distributions or across time to gauge whether they have gone up or down.

Since researchers rarely have access to population level data, they rely on representative samples drawn from the population. Although random sampling is arguably the most well-known sampling technique, other methods such as stratified random sampling have become the sampling method of choice for many national statistical agencies (see Cochran, 1977). Since there may be variation or heterogeneity among subgroups of populations (e.g., socio-demographics, neighbourhoods, etc.), stratified random sampling can aid in producing more accurate estimates (see, e.g., Zheng, 2001). Since the strata have different proportional representations in the population, calculating sample statistics is a bit more involved than simple random sampling. The strata sampling fractions must be accounted for in order to produce accurate estimates for proper statistical inference (see, e.g., Zhong & Rao, 2000).

Paper 1A considers weighted empirical likelihood as a method of inference for income quantiles arising from stratified random samples. Paper 1A contributes to the survey sampling literature by offering a more accurate way to make inferences on stratified quantiles. It also contributes to the literature on income distribution and inequality which rely on estimates of quantiles for examining and monitoring inequality (e.g., Zheng, 2001).

Empirical likelihood is a data-driven non-parametric counterpart to maximum likelihood. It is a statistical inference procedure that does not require the derivation of any asymptotic covariance structure for hypothesis testing (see Owen, 1988, 1990, 1991, 2001). As explained in Owen (2001), forgoing the calculation of complex covariances also allows empirical likelihood to excel in small samples. Unlike the mean of a distribution, inference on quantiles require the estimation of densities of the underlying distribution in order to compute the covariance (i.e., the standard error of a quantile is not a straightforward calculation). And estimation of densities require the selection of an exogenous bandwidth (smoothing parameter) which introduces more complexity (see, e.g., Li & Racine, 2007).

Paper 1A illustrates, through a Monte Carlo study, that the 95% confidence intervals for any given income quantile are indeed accurate even with small sample sizes. Paper 1A also shows that the weighted empirical likelihood approach, under some mild conditions, outperforms Woodruff's (1952) method for stratified random sample quantiles. Paper 1A's contribution is an important one for economics for two reasons: (1) it provides for an accurate data-driven procedure for making inferences on a quantile (e.g., the median income level) drawn from a stratified random sample without having to estimate distribution densities, and (2) it bridges the gap between economics and statistics by bringing more attention to empirical likelihood as a valid method of inference which is more often used by statisticians than economists (e.g., Fu et al., 2009; Wu, 2004a,b).

Paper 1B

Robust well-being comparisons entail examining multiple socio-economic indicators. While income quantiles offer a simple and effective way for comparing well-being across distributions, there exist other scalar measures in the literature which have been developed based on axiomatic grounds (e.g., Cowell, 2000; Thompson, 2010).

Paper 1B employs empirical likelihood in a two-sample setting and proposes a statistical inference procedure for comparing inequality outcomes between two populations. The procedure is developed for the generalized entropy class of inequality measures which satisfy the most widely accepted set of axioms in the income inequality literature (see Cowell, 2011; Thompson, 2010). Paper 1B tests income inequality between any two distributions and allows for a vector of inequality indicators to be tested. As pointed out by Davidson & Duclos (2000), when comparing welfare across distributions, judgments regarding which distribution is better or worse off should only be made when a wide class of indices point towards the same conclusion. Relying on single scalar measures can be problematic since different measures can point to contradictory conclusions and lead to rank reversals when comparing distributions. By allowing for multiple inequality indicators to be tested, Paper 1B alleviates this problem. Prior to Paper 1B, entropy measures of inequality could only be compared using the delta method (t–test) which require the computation of standard errors. Paper 1B shows that the empirical likelihood approach not only avoids the use of standard errors, but the approach also has greater sampling power when faced with relatively small samples.

Through a Monte Carlo study, Paper 1B illustrates that under some mild regularity conditions, empirical likelihood outperforms the traditional delta method in terms of statistical size and power. Paper 1B offers a statistically powerful test for comparing income inequality, which has significant importance and relevance in economics given that the discipline often has to rely on small samples, especially in developing countries and rural areas. Paper 1B also offers a new inference tool for economists for assessing inequality, who may not be aware of empirical likelihood techniques rooted in mathematics and statistics.

Paper 1C

Other than income inequality, the other leading topic of interest among all economists invariably is poverty—which focuses on the portion of the income distribution below the poverty line. Alleviating and eradicating poverty remains an ambitious goal for policy makers across the world (it is one of the UN's sustainable development goals; UN, 2015).

Since poverty measurement places full emphasis on the portion of the distribution that fall below the poverty line, the specification of the poverty line itself becomes a point of debate (see, e.g., Zheng, 2001; Thompson, 2012). This becomes even more problematic when comparing poverty between two subgroups of a population. A decision has to be made as to whether the two groups should have two distinct poverty lines or one common poverty line (see, e.g., Thompson, 2012; Zheng, 2001).

Since absolute poverty lines are fixed and therefore independent of the underlying distribution, statistical inference is rather straightforward as the asymptotic covariance structure is simple. However, more countries have adopted the use of relative poverty lines—those based on some fraction of the mean or a quantile from the underlying income distribution (see, e.g., Davidson & Duclos, 2000; Zheng, 2001). International organizations and local governments are increasingly using such poverty lines to report poverty estimates (e.g., OECD and the UN) since unlike absolute poverty lines, relative poverty lines do not depend on value judgments of policy makers regarding what is viewed as a necessity to achieve a minimum standard of living (see Zheng, 2001).

When using relative poverty lines, inference on poverty measures require the derivation of the asymptotic covariance structure since the poverty estimate itself relies on an estimate of the mean or quantile of the underlying income distribution (i.e., the standard error is not a straightforward calculation). Zheng (2001) developed the asymptotic framework for exactly such cases and his work is relied upon by poverty researchers interested in hypothesis tests involving relative poverty measures. Under Zheng's (2000) approach, poverty comparisons can be made but under the assumption that the two distributions are independent, and therefore each has its own poverty line.

Paper 1C adapts Zheng's (2000) approach to the case where interest is on comparing poverty outcomes between distributions that share the *same* relative poverty line. For example, if there is interest in comparing poverty outcomes among males and females that arise from the same distribution, Paper 1C argues that it is much more sensible to set a common poverty line rather than males and females having differing sets of poverty lines (this argument is supported by Thompson, 2012). Suppose a poverty line is set equal to 50% of the median income level. Under Zheng's (2000) approach, poverty estimates for males would be based on 50% of the male median income, and poverty estimates for females would be based on 50% of the female median income. Under Paper 1C's approach, the poverty line would be set equal to 50% of the pooled median income from males and females.

Prior to Paper 1C, the conventional practice was to allow each group to have its own poverty line even though this could lead to potentially invalid comparisons, or set a common absolute poverty line which is not dependent on the underlying distributions (e.g., Kakwani, 1993). That is, there was no asymptotic framework for comparing poverty levels between two groups under a common relative poverty line. Consider males and females in a certain population with female incomes much more evenly distributed but still lower than male incomes. Females would have a lower poverty line than males if we consider a relative poverty line such as 50% of the median female income. When calculating the poverty measures, it is conceivable that an incongruous conclusion could be reached that males have higher poverty than females (see Thompson, 2012).

Using asymptotic theory and partial guidance from the work of Zheng (2001), Paper 1C derives the covariance matrix for testing poverty levels between two distributions with a commonly set relative poverty line. Unlike Thompson's (2012) test which can only detect equality ($=$) or inequality ($\neq$) of poverty outcomes, the test proposed in Paper 1C is able to detect whether one subgroup actually exhibits weakly higher ($\geq$) or weakly lower ($\leq$) levels of poverty than another. The testing procedure shows good size and power properties even under small samples. Paper 1C illustrates the testing procedure by comparing poverty outcomes among males and females using data from the 2012 American Community Survey and finds that, indeed, when the two subgroups are treated as having separate poverty lines (e.g., Zheng, 2001), one invariably comes to the conclusion that males have higher poverty even though females have lower incomes than males in the sample. But under Paper 1C's approach of a common relative poverty line, the results are reversed in that females exhibit higher poverty.

Since there are often differences of opinion as to the choice of poverty line (e.g., Thompson, 2012), Paper 1C's test is robust in the sense that it allows the researcher to set and simultaneously test multiple poverty lines (e.g., one could be 40% of the median, another could be 50% of the median, another could be 60% of the median, etc.). Paper 1C advanced the literature on economic poverty and raises issues regarding the choice of poverty lines. Paper 1C's major contribution to the literature is that policy makers no longer have to rely on constructing different poverty lines for different groups of the population. Instead, a vector of commonly set relative poverty lines can be placed into Paper 1C's framework and the outcomes maybe compared without worrying about the differences in characteristics of the distributions.

Theme 1 summary

It is a common assertion now that individual well-being is multidimensional. Income remains among the strongest indicators of well-being since it often permeates so many aspects of it (see, e.g., Barrett

& Donald, 2003; Davidson & Duclos, 2000). Drawing from the statistics literature, Theme 1 develops statistical inference procedures for quantiles arising from complex survey schemes and for comparing inequality between distributions, using the data-driven non-parametric technique of empirical likelihood which does not impose any assumption on the underlying distributions (e.g., Owen, 2001). Theme 1 also makes a contribution to the poverty literature by proposing a statistical test for comparing poverty between distributions that share a common relative poverty line (such as 50% of the median income level of the pooled distributions). Theme 1 highlighted the importance of accurate and robust statistical inference procedures in the well-being literature and Monte Carlo studies confirm the statistical reliability of the methodologies presented in Theme 1.

Theme 2: Multidimensional well-being indicators

Paper 2A

While income-based measures have their respective merits and are important aspects of individual well-being, they do not capture subjective aspects of quality of life (e.g., environment quality, leisure, personal relationships, health, etc.) (see, e.g., Boarini et al., 2012; Decancq, 2017; Fleurbaey, 2012; Mizobuchi, 2014). Subjective well-being indicators have recently been elevated in importance as they relate strongly to quality of life (e.g., OECD, 2017; Monika, 2018). Life satisfaction, a subjective indicator of well-being, is a manifestation of multiple life domains and is rather difficult to synthesize into just one metric. Nonetheless, aggregating several dimensions into a single measure provides policy makers with a simple tool which is easy to communicate, and allows them to benchmark and monitor progress in composite well-being. One of the most pressing issues when it comes to multidimensional well-being indices is how to weight the individual component indicators (e.g., Biswas & Caliendo, 2002; Decancq & Lugo, 2013; Lorenz et al., 2017).

The CDI, developed by Save the Children UK, is a composite index for measuring and monitoring the well-being of children across various countries. The CDI is composed of 3 equally-weighted dimensions: health, education, and nutrition. Paper 2A uses consistent tests of SDE proposed by Scaillet & Topaloglou (2010), and finds that equally weighting the dimensions is not stochastic dominance efficient—there exists other weighting vectors that result in more countries achieving better measured outcomes.

SDE is a direct extension of stochastic dominance which is a robust method of ranking distributions that account for all moments of the underlying distributions and is quite commonly used in making pairwise comparisons of welfare (e.g., Barrett & Donald, 2003; Davidson & Duclos, 2000). SDE extends the concept by making it possible to compare a benchmark distribution of multiple aggregated variables against all possible linear combinations of those variables (see Scaillet & Topaloglou, 2010). There has been burgeoning literature on the use of SDE for well-being comparisons and ranking in general (e.g., Pinar et al., 2013, 2015; Pinar, 2015; Pinar et al., 2017, 2019).

In Paper 2A, the official equal-weighting scheme of the CDI is considered as the "benchmark" distribution and is statistically tested against all other possible weighting combinations. Unlike other methods such as generalized mean aggregation, preference-based approaches, principal components analysis, factor analysis, or data envelopment analysis which only consider the first few moments of the distribution, stochastic dominance utilizes all moments of the underlying distribution and SDE reveals what type of weighting scheme it takes to induce the most optimistic scenario where more countries achieve better measured outcomes based on the least variable combination of components. SDE also reveals what it takes to achieve the most pessimistic scenario where more countries achieve worse measured outcomes. This is not synonymous with actually maximizing or minimizing well-being (no method can actually measure or assess true well-being). But rather SDE is a data-driven method and as such it provides intuition into the tendencies of the index data which are thought to partially represent well-being. Applying SDE reveals which measured aspects or dimensions of well-being have driven the most overall improvements and which aspects are hindering more countries from achieving better measured outcomes. Another advantage of SDE is that under the most optimistic scenario, the weights are allocated such that most countries achieve a better measured outcome so in a sense it is the measured outcomes that are "optimized" under a fixed weighting scheme for every country. Sometimes, national policy makers take issue with the fact that equal-weighting masks achievements in dimensions they think they excel in. SDE partially addresses such concerns by optimizing most countries' indices.

In order to induce the most optimistic scenario where more countries achieve higher index scores, Paper 2A finds that relatively more weight needs to be shifted towards education. This finding suggests that the upper bound of the education sub-index is more achievable compared to health or nutrition. In other words, more countries find it easier to achieve better index scores when education receives more weight than the other dimensions of the CDI.

At the other extreme, shifting more weight towards health induces the most pessimistic scenario where index scores are worsened for more countries. This suggests that compared to education, improvements in the health sub-index have not kept pace with what is observed in the education sub-index data.

Paper 2A provides re-weighted rankings using the SDE weights from the two extreme scenarios (i.e., best-case and worst-case). Countries that excel in all aspects of the CDI (i.e., health, education, and nutrition) show little deviation in rankings since they achieved a balanced set of measured outcomes across the different aspects of the CDI. Again, the key word here is "measured" because the CDI, just like any other composite index, provides an approximation of well-being and thus the country rankings should also be viewed as approximate. Countries that posted poor scores in education, generally moved down in ranking. These findings are in line with Pinar et al. (2013) who found education to be the dominant indicator for UN's Human Development Index. Paper 2A's findings help shed light for policy makers who may be unaware of their nation's strengths and weaknesses when it comes to indicators that are supposed to partially represent child development outcomes, and allows policy makers to better target areas in need of improvement.

Paper 2A takes SDE from the finance literature (Scaillet & Topaloglou, 2010) and applies it in an economic setting. SDE is rooted in the finance literature where it is used for portfolio optimization, but Paper 2A's contribution along with those of Pinar et al. (2013, 2017), Pinar (2015), Pinar et al. (2015), and Pinar et al. (2019) demonstrate the possibilities of SDE as a tool for assessing weighting schemes in economics, particularly when it comes to assessing the choice of weights in constructing composite indices.

Paper 2B

While the CDI measures and monitors the well-being of children, The OECD's BLI is aimed at measuring and monitoring multidimensional well-being for the broader population. Introduced in 2011, the BLI is by far one of the most well-known multidimensional indices and several researchers have assessed its efficacy (e.g., Decancq, 2017; Kasparian, 2012; Lorenz et al., 2017; Markovic et al., 2016; Mizobuchi, 2014; Monika, 2018; Nikolaev, 2014; von Reumont et al., 2017). It comprises 11 dimensions with each dimension consisting of 1 to 4 indicators. Countries are assigned a score in each dimension by taking the arithmetic average of the component indicators (i.e., equally-weighting the indicators).

Paper 2B not only analyzes each of the 11 dimensions separately but also constructs an aggregate BLI index that encompasses all 24 BLI indicators, something not done by the OECD. The BLI allows its website (`www.oecdbetterlifeindex.org`) users to create their own BLI by adjusting the normalized weight assigned to each of the dimensions. But in constructing the overall BLI index in Paper 2B, it is argued that due to embedding effects (e.g., von Reumont et al., 2017), it would be more beneficial to allow users to assign importance to dimensional *indicators* as opposed to just the dimensions alone. A preference for dimension A over dimension B does not necessarily imply a preference for all indicators in dimension A over those of dimension B.

In constructing the aggregate composite indicator, Paper 2B actually weights all 24 indicators as opposed to weighting the 11 dimensions. BLI country rankings by dimension, are reported in Paper 2B under the two extreme scenarios. SDE reveals that due to unbalanced achievements (as measured by the BLI), most OECD countries experience drastic shifts in rankings, depending on how the indicators, thought to be representative of well-being, are actually weighted. The rankings serve as a tool for policy makers looking to improve their relative standing among OECD countries by setting targets that address their weaknesses. It lets policy makers see which aspects of well-being their country is lacking in and which ones they excel in. By focusing on the weak indicators identified by SDE, countries can move towards improving their relative standing in future editions of the BLI.

Paper 2C

One of the dimensions of the BLI is *jobs*. Work-life balance and job satisfaction are important aspects of well-being as pointed out by Drobnic et al., 2010, OECD (2014), and OECD (2017). Chen & Hou (2019) investigate the effects of unemployment on well-being and establishes an association between an individual's own employment status and their life satisfaction. The amount of time spent working during a person's lifetime is not trivial and therefore working-time quality becomes a significant factor of individual well-being.

There is growing recognition that employment should not only be judged by the number of jobs created or how much these jobs pay but they should also be judged by the quality of jobs created (see, e.g., Drobnic et al., 2010). While much of Europe has consistently published a unified set of job quality indicators, countries like Canada have lagged behind, primary due to a lack of unified data source (such as that of Eurofound). With increasing concern regarding so-called "precarious" work (e.g., term, contract, seasonal) in Canada, a lack of job quality indicators make it difficult to assess working conditions (May, 2019).

Paper 2C not only fills a gap in the academic literature but also offers indicators and analyses for guiding policy. Paper 2C uses the Eurofound framework (Eurofound, 2016) to develop job quality indicators from the 2016 CGSS which had a module similar to Eurofound. Paper 2C adopts a multidimensional approach to measuring job quality in Canada by examining 6 broad dimensions used by Eurofound.

By constructing indicators for Canada similar to Eurofound, Paper 2C opens up possibilities for international comparisons of job quality with Canada. Paper 2C makes an important contribution to the Canadian labour literature and policy by focusing on indicators other than wages. Paper 2C serves as an important policy guide as evidenced by the fact that it has been recently cited in an official Canadian government report on precarious jobs from the House of Commons (May, 2019).

Paper 2C not only presents descriptive statistics of job quality but also illustrates the possibilities of incorporating machine learning into economics. Paper 2C borrows a technique known as latent class analysis from computer science and provides predictions of workers likely to be in good or poor quality jobs. With the advent of technology and increasing computing power, disciplines such as computer science have dramatically boosted the capabilities of machine learning algorithms. Through the use of a machine learning technique, Paper 2C takes a step towards bridging this gap between economics and computer science.

Theme 2 summary

With increased interest in developing well-being measures within a multidimensional framework (e.g., Boarini & D'Ercole, 2013; OECD, 2017), there is a growing need for methods aimed at aggregating and assessing such measures. Multidimensional well-being measures, as the name suggests, comprise two or more dimensions related to well-being. The final index is typically computed by taking a weighted average of the dimensions. The dimensional weights are often contested given differences in people's preferences with regards to the different aspects of well-being. Theme 2 assesses OECD's BLI and Save the Children UK's CDI, which are two indices that exemplify multivariate well-being. Rather than advocating for a single weighting vector, SDE is used to examine the type of weighting scheme necessary to collectively improve or worsen the outcome scores as measured by the respective indices. By uncovering dimensions most or least responsible for improving trends in measured outcomes over time, this allows policy makers to better target areas in need of improvement. Theme 2 made an additional contribution to the well-being literature by developing multidimensional indicators for assessing an important aspect of well-being: work. The indicators were aimed at assessing job quality in Canada which is an OECD country that, up until now, lacked internationally comparable job quality indicators.

Conclusion

This body of research presents robust approaches for making comparisons of socio-economic well-being outcomes by blending theory and applications. Theme 1 research papers propose novel statistical inference methods for comparing inequality and poverty outcomes using income as the welfare measure. Theme 2 research papers addressed the growing recognition that well-being manifests along multiple dimensions rather than wealth alone.

Measuring and monitoring well-being—whether it is univariate or multivariate—is important for policy makers since the economic success of a nation ultimately comes down to the welfare of its citizens. With increasing data availability in the age of "big data", governments and international organizations continue to develop indicators for approximating individual well-being. Given the proliferation of composite indices for measuring well-being, there is a growing need for more reliable and accurate statistical methods for assessing such measures. Theme 1 and Theme 2 research papers recognize this and highlight the importance of robust statistical inference and accurate estimation.

Although multidimensional approaches to measuring well-being are now more pervasive, income-based measures still have their place in the literature given the reliance on metrics such as GDP and the fact that income is thought to permeate many aspects of well-being. Measuring well-being ultimately comes down to first collecting reliable data. And data is typically drawn from household surveys which can have complex designs. Stratified sampling is one such design. Theme 1 presents a cohesive body of research aimed at better measuring and comparing inequality and poverty outcomes using income as a measure of well-being, with the following questions in mind:

1i) How can hypothesis tests regarding quantiles that arise from stratified random samples be undertaken?

1ii) How can inequality between income distributions be reliably compared?

1iii) Can poverty comparisons between distributions be made if the poverty line depends on the pooled distribution?

To answer question (1i), Theme 1 begins by proposing an empirical likelihood approach for making inferences on income quantiles drawn from stratified random samples. Since empirical likelihood relaxes distributional assumptions regarding the underlying data and does not rely on covariances, it avoids such complexities synonymous with stratified sample quantiles. Question (1ii) is addressed by applying the empirical likelihood method for making comparisons of distributions using a wide class of inequality measures that satisfy the most widely accepted axioms of inequality in the literature. Reliable well-being comparisons entail comparing multiple measures. The proposed methodology is reliable and robust in the sense that it permits simultaneous comparison of multiple inequality measures, while at the same time it avoids having to compute complex covariance structures. Shifting the focus from inequality to poverty, question (1iii) is answered with the conclusion of Theme 1 which develops an asymptotic framework for testing whether one distribution has higher or lower poverty than another given a common relative poverty line—which was not possible in the past. The use of a shared poverty line between distributions enables more valid comparisons of poverty outcomes.

Building on Theme 1, Theme 2 goes beyond material standard of living and aims at assessing well-being measures that incorporate more than just income indicators, with these questions in mind:

2i) How sensitive are multidimensional well-being indices to the choice of weights?

2ii) Can any country achieve a good index score just by varying the indicator weights?

2iii) Is there a "best" way to select weights for the various dimensions of well-being?

2iv) Can multidimensional well-being be measured without weights?

Since multidimensional indices are usually the weighted sum of all the components, the choice of weights becomes a point of contention. A distributional robust ranking method known as stochastic dominance efficiency is used to assess the official equal-weighting schemes of Save the Children UK's Child Development Index and OECD's Better Life Index, which are two examples of composite welfare measures. Question (2i) is answered by presenting a best-case and worst-case weighting scenario for each index to reveal the aspects of well-being responsible for improving measured outcomes and aspects hindering improvements across countries. Analyses show that multivariate well-being measures are highly influenced by choice of weights and policy makers need to be aware of such sensitivities when basing policy targets on multivariate well-being measures. The answer to question (2ii) is: no. Although going from the equal-weight to the best-case or worst-case scenario can drastically change index scores and country rankings, a country must excel in at least one of the dimensions to achieve a relatively good index score. In response to (2iii), an optimal way of selecting weighting vectors does not exist for multidimensional well-being measures given that there is often disagreement as to what dimensions should be included and individuals assign different relative importance to dimensions. In the absence of an optimal measure, methods such as SDE do, however, go some ways towards assessing existing empirical procedures. To answer question (2iv), Theme 2 concludes by offering a deeper look at one of the dimensions of well-being: work. The assessment is done by examining the joint distribution of the indicators rather than aggregating indicators into a single index. Canada is an OECD country which until now, lacked data on job quality. Constructing a suite of Canadian job quality indicators, largely consistent with those used by several European countries, unlocks the potential for international comparability. It also marks the first comprehensive job quality study in Canada based on wage as well as non-wage indicators

With surging interest on measuring and assessing societal well-being, there exists a growing need for more accurate and reliable methods of measurement and inference. This body of research which blends theory and applications, has important implications for researchers as well as policy makers who rely on welfare measures for setting policy targets. The power and breadth of the empirical likelihood technique was showcased and its application to inequality and poverty analysis illustrates the versatility of the method. Robust poverty comparisons can now be made between distributions by setting a common poverty line relative to the combined distribution. The technique of stochastic dominance efficiency shows potential for economic applications, especially with regards to assessing multidimensional welfare indices which have proliferated with the advent of richer and better data availability.

References

Alkire, S. (2020). Multidimensional poverty measures as policy tools. In V. Beck, H. Hahn, & R. Lepenies (Eds.) *Dimensions of Poverty*. Cham: Philosophy and Poverty, vol 2. Springer.

Alkire, S., Foster, J. E., Seth, S., Santos, M. E., Roche, J., & Ballon, P. (2015). *Multidimensional poverty measurement and analysis*. Oxford: Oxford University Press.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*, 244–263.

Atkinson, A. B. (1987). On the measurement of poverty. *Econometrica*, *55*, 749–764.

Atkinson, A. B., & Bourguignon, F. (1982). The comparison of multi-dimensioned distributions of economic status. *The Review of Economic Studies*, *49*, 183–201.

Barrett, G. F., & Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, *71*, 71–104.

Bhattacharya, D. (2005). Asymptotic inference from multi-stage samples. *Journal of Econometrics*, *126*, 145–171.

Bhattacharya, D. (2007). Inference on inequality from household survey data. *Journal of Econometrics*, *137*, 674–707.

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *Journal of Econometrics*, *108*, 317–342.

Biswas, B., & Caliendo, F. (2002). A multivariate analysis of the Human Development Index. *Indian Economic Journal*, *49*, 96–100.

Boarini, R., Comola, M., Smith, C., Manchin, R., & de Keulenaer, F. (2012). *What Makes for a Better Life?: The Determinants of Subjective Well-Being in OECD Countries - Evidence from the Gallup World Poll*. OECD Statistics Working Papers, No. 1815-2031, Paris.

Boarini, R., & D'Ercole, M. M. (2013). Going beyond GDP: An OECD perspective. *Fiscal Studies*, *34*, 289–314.

Bourguignon, F., & Chakravarty, S. R. (2003). The measurement of multidimensional poverty. *Journal of Economic Inequality*, *1*, 25–49.

Brachman, K., Stich, A., & Trede, M. (1996). Evaluating parametric income distribution models. *Allgemeines Statistisches Archiv*, *80*, 285–298.

Chen, J., & Sitter, R. R. (1999). A pseudo empirical likelihood approach to the effective use of auxilary information in complex surveys. *Statistica Sinica*, *9*, 385–406.

Chen, W.-H., & Hou, F. (2019). The effect of unemployment on life satisfaction: A cross-national comparison between Canada, Germany, the United Kingdom and the United States. *Applied Research in Quality of Life*, *14*, 1035–1058.

Clark, S., Hemming, R., & Ulph, D. (1981). On indices for the measurement of poverty. *Economic Journal*, *91*, 515–526.

Cochran, W. G. (1977). *Sampling techniques (third edition)*. New York: John Wiley & Sons Inc.

Cowell, F. A. (1989). Sampling variance and decomposable inequality measures. *Journal of Econometrics*, *42*, 27–41.

Cowell, F. A. (2000). Measurement of inequality. In A. B. Atkinson, & F. Bourguignon (Eds.) *Handbook of Income Distribution, Volume I*. Amsterdam: North Holland.

Cowell, F. A. (2011). *Measuring Inequality, third ed.*. Oxford: Oxford University Press.

Cowell, F. A., & Kuga, K. (1981). Inequality measurement: An axiomatic approach. *European Economic Review*, *15*, 287–305.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, *30*, 348–361.

Davidson, R. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*, *141*, 141–166.

Davidson, R., & Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and the measurement of poverty and inequality. *Econometrica*, *68*, 1435–1464.

Deaton, A. (1997). *The analysis of household surveys: A microeconometric approach to development policy*. Baltimore: The John Hopkins University Press.

Decancq, K. (2017). Measuring multidimensional inequality in the OECD member countries with a distribution-sensitive Better Life Index. *Social Indicators Research*, *131*, 1057–1086.

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, *32*, 7–34.

Drobnic, S., Beham, B., & Prag, P. (2010). Good job, good life? Working conditions and quality of life in Europe. *Social Indicators Research*, *99*, 205–225.

Duclos, J.-Y., Sahn, D. E., & Younger, S. D. (2006). Robust multidimensional poverty comparisons. *The Economic Journal*, *116*, 943–968.

Dutta, I., & Foster, J. (2013). Inequality of happiness in the U .S .: 1972–2010. *The Review of Income and Wealth*, *59*, 393–415.

Eurofound (2016). Sixth european working conditions survey–overview report. Luxembourg: Publications Office of the European Union.

Fleurbaey, M. (2012). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, *47*, 1029–1075.

Foster, J., Greer, J., & Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, *52*, 761–766.

Foster, J. E. (1984). On economic poverty: A survey of aggregate measures. *Advances in Econometrics*, *3*, 215–251.

Foster, J. E., McGiillivray, M., & Seth, S. (2013). Composite indices: Rank robustness, statistical association, and redundancy. *Econometric Reviews*, *32*, 35–36.

Francisco, C. A., & Fuller, W. A. (1991). Quantile estimation with a complex suvey design. *The Annals of Statistics*, *19*, 454–469.

Fu, Y., Wang, X., & Wu, C. (2009). Weighted empirical likelihood inference for multiple samples. *Journal of Statistical Planning and Inference*, *139*, 1462–1473.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2018). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, (pp. 1–34).

Gross, S. T. (1980). Median estimation in sample surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, (pp. 181–184).

Kakwani, N. (1993). Statistical inference in the measurement of poverty. *The Review of Economics and Statistics*, *75*, 632–639.

Kasparian, J. (2012). OECD's 'better life index': Can any country be well ranked? *Journal of Applied Statistics*, *39*, 2223–2230.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice*. Princeton: Princeton University Press.

Lorenz, J., Brauer, C., & Lorenz, D. (2017). Rank-optimal weighting or "how to be best in the OECD Better Life Index?". *Social Indicators Research*, *134*, 75–92.

Maasoumi, E. (1999). Multidimensional approaches to welfare analysis. In J. Silber (Ed.) *Handbook of income inequality measurement*. Dordrecht and New York: Klewer Academic Publishers.

Markovic, M., Zdravkovic, S., Mitrovic, M., & Radojicic, A. (2016). An iterative multivariate post hoc i-distance approach in evaluating OECD Better Life Index. *Social Indicators Research*, *126*, 1–19.

May, B. (2019). *Precarious work: Understanding the changing nature of work in Canada*. House of Commons of Canada, Ottawa, Canada.

McDonald, J. (1984). Some generalized functions for the size distribution of income. *Econometrica*, *52*, 647–663.

Mizobuchi, H. (2014). Measuring world better life frontier: A composite indicator for OECD Better Life Index. *Social Indicators Research*, *118*, 987–1007.

Monika, G. T. (2018). The weight of weighting - an empirical study based on the OECD better life index. *The Business and Management Review*, *9*, 443–450.

Nikolaev, B. (2014). Economic freedom and quality of life: Evidence from the OECD's Your Better Life Index. *The Journal of Private Enterprise*, *29*, 61–96.

OECD (2010). Income inequality. In *OECD Factbook 2010: Economic, Environmental and Social Statistics*. Paris: OECD Publishing.

OECD (2014). How good is your job? Measuring and assessing job quality in oecd employment outlook 2014. Paris: OECD Publishing.

OECD (2017). How's life? 2017: Measuring well-being. Paris: OECD Publishing.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, *75*, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*, 90–120.

Owen, A. B. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, *19*, 1725–1747.

Owen, A. B. (2001). *Empirical likelihood*. Boca Raton: Chapman & Hall/CRC.

Pigou, A. C. (1912). *Wealth and Welfare*. London: Macmillan.

Pinar, M. (2015). Measuring world governance: Revisiting the institutions hypothesis. *Empirical Economics*, *48*, 747–778.

Pinar, M. (2019). Multidimensional well-being and inequality across the European regions with alternative interactions between the well-being dimensions. *Social Indicators Research*, *144*, 31–72.

Pinar, M., Milla, J., & Stengos, T. (2019). Sensitivity of university rankings: Implications of stochastic dominance efficiency analysis. *Education Economics*, *27*, 75–92.

Pinar, M., Stengos, T., & Topaloglou, N. (2013). Measuring human development: A stochastic dominance approach. *Journal of Economic Growth*, *18*, 69–108.

Pinar, M., Stengos, T., & Topaloglou, N. (2017). Testing for the implicit weights of the dimensions of the Human Development Index using stochastic dominance. *Economics Letters*, *161*, 38–42.

Pinar, M., Stengos, T., & Yazgan, M. E. (2015). Measuring human development in the MENA region. *Emerging Markets Finance and Trade*, *51*, 1179–1192.

Ravallion, M. (2011). On multidimensional indices of poverty. *Journal of Economic Inequality*, *9*, 235–248.

Save the Children UK (2008). *The Child Development Index: Holding governments to account for children's wellbeing*.

Save the Children UK (2012). *The Child Development Index 2012: Progress, challenges, and inequality*.

Scaillet, O., & Topaloglou, N. (2010). Testing for stochastic dominance efficiency. *Journal of Business & Economic Statistics*, *28*, 169–180.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, *44*, 219–231.

Seth, S., & McGiillivray, M. (2018). Composite indices, alternative weights, and comparison robustness. *Social Choice and Welfare*, *51*, 657–679.

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *The measurement of economic performance and social progress revisited*. Reflections and overview Commission on the Measurement of Economic Performance and Social Progress, Paris.

Thompson, B. S. (2010). Statistical inference for vector measures of inequality and poverty. *Journal of Economic Inequality*, *8*, 451–462.

Thompson, B. S. (2012). Empirical likelihood-based inference for poverty measures with relative poverty lines. *Econometric Reviews*, *32*, 513–523.

UN (2015). Transforming our world: The 2030 agenda for sustainable development. United Nations.

von Reumont, L., Schob, R., & Hetschko, C. (2017). *Embedding effects in the OECD Better Life Index*. Beitrage zur Jahrestagung des Vereins fur Socialpolitik 2017: Alternative Geld-und Finanzarchitekturen - Session: Political Economy II, No. E20-V1.

Watts, H. (1968). *An economic definition of poverty*. New York: In: Moynihan, D.P. (Ed.), On Understanding Poverty. Basic Books.

Wu, C. (2004a). Some algorithmic aspects of the empirical likelihood method in survey sampling. *Statistica Sinica*, *14*, 1057–1069.

Wu, C. (2004b). Weighted empirical likelihood inference. *Statistics & Probability Letters*, *66*, 67–79.

Yang, L. (2018). Measuring well-being: A multidimensional index integrating subjective well-being and preferences. *Journal of Human Development and Capabilities*, *18*, 456–476.

Zheng, B. (1997). Aggregate poverty measures. *Journal of Economic Surveys*, *11*, 123–162.

Zheng, B. (2001). Statistical inference for poverty measures with relative poverty lines. *Journal of Econometrics*, *101*, 337–356.

Zhong, B., & Rao, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxilary population information. *Biometrika*, *87*, 929–938.

Appendix A: Papers submitted for PhD by Publications

Paper 1A

Economics Bulletin

Volume 33, Issue 3

Weighted empirical likelihood-based inference for quantiles under stratified random sampling

Tahsin Mehdi
University of Guelph

Abstract

A growing body of literature is emerging on empirical likelihood methods for complex surveys. These works largely focus on the population mean. We propose a weighted empirical likelihood approach as a method of inference for quantiles under stratified random sampling, which is one of the most popular complex survey designs. A simulation study substantiates our proposed methodology.

Citation: Tahsin Mehdi, (2013) "Weighted empirical likelihood-based inference for quantiles under stratified random sampling", *Economics Bulletin*, Vol. 33 No. 3 pp. 2437-2442.
Contact: Tahsin Mehdi - tmehdi@uoguelph.ca.
Submitted: July 26, 2013. **Published:** September 19, 2013.

1 Introduction

Ever since the pioneering work of Owen (1988, 1990), there has been a proliferation of literature on empirical likelihood (EL), which is a powerful nonparametric statistical tool. An advantage of this method is that one does not need to assume anything about the underlying distribution of the data. The EL ratio is entirely data driven. The main focus of Owen (1988) was to construct confidence intervals for a population mean given a single sample of independent and identically distributed (iid) observations. For a detailed overview of EL, see Owen (2001).

EL methods for complex surveys are yet to be investigated to their full extent. Chen and Sitter (1999) and Zhong and Rao (2000) were among the first to consider EL in the context of complex survey designs. Complex survey designs pose additional difficulties for the conventional EL approach. Asymptotic results from conventional EL are not directly applicable to complex surveys as special types of constraints may need to be imposed, depending on the survey design. Generally, the EL ratio in such cases will not have the same calibration as in the case of simple random sampling. Also, existing computational procedures may not be readily applicable. To alleviate such problems, Fu et al. (2008) introduced a weighted empirical likelihood method and developed an unified approach for making inferences on population means in the presence of multiple samples. One of the cases they consider is stratified random sampling where the focal point of interest is on the overall population mean. Their approach relies on the augmentation of the special types of constraints induced by stratified samples.

Though a large body of literature exists regarding inference on population means, quantiles have received relatively less attention. This is especially true in the case of complex survey designs. Such designs could conceivably give rise to multiple distribution functions (as is the case with stratified random sampling) instead of just one. Thus, deriving asymptotic expressions for quantiles can get quite tedious and sometimes may not even be possible. For simple random sampling, (Owen, 2001, Ch. 3.6) provides a good introduction to EL methods for quantiles.

The promising results of Fu et al. (2008) warrant further research into the weighted empirical likelihood approach. Drawing upon their work, we propose a weighted empirical likelihood-based inference method for quantiles under stratified random sampling. Our results rely on very similar asymptotic expansions.

The rest of this article is organized as follows. In Section 2, we introduce our proposed methodology and establish some asymptotic results under stratified random sampling. In Section 3, we present a Monte Carlo study which assesses the accuracy of the confidence interval obtained from our method. Section 4 concludes.

2 Inference for a quantile

Suppose that a population is divided into k mutually exclusive strata of known sizes $N_1, ..., N_k$. The weight associated with the ith stratum is $w_i = N_i/N$, where $N = \sum_{i=1}^k N_i$ is the overall population size. Let $\{Y_{ij}, j = 1, ..., n_i\}$, $i = 1, ..., k$, be k independent samples of size n_i extracted from the strata and let $n = \sum_{i=1}^k n_i$ be the pooled sample size. Assume the strata sampling fraction n_i/N_i is negligible so that $\{Y_{ij}, j = 1, ..., n_i\}$ is regarded as an iid sample generated by the continuous random variable Y_i with distribution function F_i. The overall distribution function is then given by $F(y) = \sum_{i=1}^k w_i F_i(y)$. Let Q^α denote the α-quantile of F. This quantile is implicitly characterized by $F(Q^\alpha) = \alpha$. Our focal point of interest is on constructing confidence intervals for Q^α given α.

The weighted empirical log-likelihood (WEL) function of Fu et al. (2008) is given by

$$l_w(F_1, ..., F_k) = \sum_{i=1}^{k} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(p_{ij}), \tag{1}$$

where p_{ij} is the probability associated with Y_{ij}. The formulation of (1) was motivated using the argument of Chen and Sitter (1999). See Fu et al. (2008) for more on this.

An advantage of using the WEL function is that the usual large sample properties of EL can be established under the special type of constraints induced by stratified samples. If the constraints are reformulated in a suitable way, computational procedures are also readily available.

To construct a confidence interval for Q^α, we maximize (1) subject to $p_{ij} > 0$ and

$$\sum_{j=1}^{n_i} p_{ij} = 1, \quad i = 1, ..., k, \tag{2}$$

$$\sum_{i=1}^{k} w_i \sum_{j=1}^{n_i} p_{ij} 1_{Y_{ij} \leq Q^\alpha} = \alpha, \tag{3}$$

where $1_{(\cdot)}$ is an indicator function which evaluates to one if the argument $(\cdot)$ is true, and zero otherwise. Constraint (3) identifies the quantile Q^α and its use can be justified by arguments similar to (Owen, 2001, Ch. 3.6). Since $E_i(1_{Y_i \leq Q^\alpha}) = F_i(Q^\alpha)$, where E_i denotes the expectation under distribution F_i, constraint (3) indeed identifies Q^α.

To construct confidence intervals for Q^α, we require the asymptotic distribution of the WEL ratio which Fu et al. (2008) defines as

$$r_w(Q^\alpha) = \sum_{i=1}^{k} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(n\hat{p}_{ij}),$$

where $\hat{p}_{ij}$ given by (7) solve the maximization problem. Assume $n_i/n \to x \neq 0$, so that it is unnecessary to distinguish between $O(n^{-1/2})$ and $O(n_i^{-1/2})$, and between $o(n^{-1/2})$ and $o(n_i^{-1/2})$. The following theorem establishes the asymptotic distribution of $r_w(Q^\alpha)$ at $Q^\alpha = Q_0^\alpha$.

Theorem 2.1. *Suppose $\{Y_{ij}, j = 1, ..., n_i\}$ is an iid sample, with finite variance, from $F_i, i = 1, ..., k$, and the k samples are independent of each other. If Q_0^α is the α-quantile of the overall distribution function F, then $-2r_w(Q_0^\alpha)/c \xrightarrow{d} \chi^2_{(1)}$, where the scaling constant c is given by* (12).

Proof. Our proof follows very closely the proof of Fu et al. (2008) for stratified sampling. For ease of notation and without loss of generality, consider $k = 3$. Constraints (2) and (3) can be reformulated as

$$\sum_{i=1}^{3} w_i \sum_{j=1}^{n_i} p_{ij} = 1, \tag{4}$$

$$\sum_{i=1}^{3} w_i \sum_{j=1}^{n_i} p_{ij} \boldsymbol{Z}_{ij} = \boldsymbol{\eta}, \tag{5}$$

where the vector-valued variables $\boldsymbol{Z}_{ij}$ and $\boldsymbol{\eta}$ are given by

$$\begin{aligned} \boldsymbol{Z}_{1i} &= (1, 0, 1_{Y_{1i} \leq Q^\alpha})', \\ \boldsymbol{Z}_{2i} &= (0, 1, 1_{Y_{2i} \leq Q^\alpha})', \\ \boldsymbol{Z}_{3i} &= (0, 0, 1_{Y_{3i} \leq Q^\alpha})', \end{aligned}$$

and

$$\boldsymbol{\eta} = (w_1, w_2, \alpha)'.$$

Equation (5) can be re-written as

$$\sum_{i=1}^{3} w_i \sum_{j=1}^{n_i} p_{ij} \boldsymbol{u}_{ij} = \boldsymbol{0}, \tag{6}$$

where $\boldsymbol{u}_{ij} = \boldsymbol{Z}_{ij} - \boldsymbol{\eta}$. The reformulation of constraints (2) and (3) ensure that the probabilities in each of the stratum sum to unity. The maximization of (1) subject to (4) and (6) can be carried out using the Lagrange multiplier technique. For a given Q^α, it can be shown that the optimized probabilities are

$$\hat{p}_{ij}(Q^\alpha) = \frac{1}{n_i(1 + \boldsymbol{\lambda}' \boldsymbol{u}_{ij})}, \tag{7}$$

where the vector-valued Lagrange multiplier $\boldsymbol{\lambda}$ is the solution to

$$\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{\boldsymbol{u}_{ij}}{1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij}} = \mathbf{0}. \tag{8}$$

The above equation can be solved using the algorithm described in Wu (2004). Rewriting the numerator $\boldsymbol{u}_{ij}$ in (8) as $\boldsymbol{u}_{ij}[(1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij}) - \boldsymbol{u}'_{ij}\boldsymbol{\lambda}]$, equation (8) can be expressed as

$$\left[\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \frac{\boldsymbol{u}_{ij}\boldsymbol{u}'_{ij}}{1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij}}\right] \boldsymbol{\lambda} = \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \boldsymbol{u}_{ij}. \tag{9}$$

Noting that $\sum_{j=1}^{n_i} [n_i(1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij})]^{-1} = 1$, for $i = 1, 2, 3$, the order of $\boldsymbol{\lambda}$ is related to the right-hand side of (9), which can be written as

$$\boldsymbol{U} = \sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \boldsymbol{u}_{ij} = \left(0, 0, \hat{F}(Q^\alpha) - \alpha\right)', \tag{10}$$

where (for $k = 3$) $\hat{F}(Q^\alpha) = \sum_{i=1}^{3}(w_i/n_i)\sum_{j=1}^{n_i} 1_{Y_{ij} \le Q^\alpha}$. Since $\alpha = F(Q^\alpha)$, it immediately follows that $\boldsymbol{U} = O_p(n^{-1/2})$ (component-wise) when $Q^\alpha = Q_0^\alpha$.

Letting $\boldsymbol{D} = \sum_{i=1}^{3}(w_i/n_i)\sum_{j=1}^{n_i} \boldsymbol{u}_{ij}\boldsymbol{u}'_{ij}$ and noting that it is $O_p(1)$, from (9) we have that $\boldsymbol{\lambda} = O_p(n^{-1/2})$. The finite variance assumption allows us to have $\max_{ij}|\boldsymbol{u}_{ij}| = o_p(n^{1/2})$ and $\boldsymbol{\lambda}'\boldsymbol{u}_{ij} = o_p(1)$ uniformly over all i and j (see Owen, 2001, Ch. 11.1). An asymptotic expression for the Lagrange multiplier is obtained as

$$\boldsymbol{\lambda} = \boldsymbol{D}^{-1}\boldsymbol{U} + o_p(n^{-1/2}). \tag{11}$$

The WEL ratio function at Q_0^α is

$$r_w(Q_0^\alpha) = -\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij}).$$

Using a second order Taylor expansion on $\log(\cdot)$, we obtain the following asymptotic expansion of the WEL ratio,

$$\begin{aligned}
-2r_w(Q_0^\alpha) &= 2\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log(1+\boldsymbol{\lambda}'\boldsymbol{u}_{ij}) \\
&= 2\sum_{i=1}^{3} \frac{w_i}{n_i} \sum_{j=1}^{n_i} \log\left(\boldsymbol{\lambda}'\boldsymbol{u}_{ij} - \frac{1}{2}\boldsymbol{\lambda}'\boldsymbol{u}_{ij}\boldsymbol{u}'_{ij}\boldsymbol{\lambda}\right) + o_p(n^{-1}) \\
&= \boldsymbol{U}'\boldsymbol{D}^{-1}\boldsymbol{U} + o_p(n^{-1}) \\
&= d_{33}(\hat{F}(Q_0^\alpha) - \alpha)^2 + o_p(n^{-1}),
\end{aligned}$$

where the last step is a consequence of (10) and d_{33} is the last (third for $k = 3$) diagonal element of $\boldsymbol{D}^{-1}$. If we let

$$c = d_{33} \sum_{i=1}^{3} \frac{w_i^2}{n_i - 1} F_i(Q_0^\alpha)(1 - F_i(Q_0^\alpha)), \tag{12}$$

it immediately follows that $-2r_w(Q_0^\alpha)/c$ will have a limiting χ^2 distribution with one degree of freedom. □

The scaling constant c involves the true distribution function F_i and quantile Q_0^α. Replacing Q_0^α by its weighted sample quantile $\hat{Q}^\alpha = \hat{F}^{-1}(\alpha)$ and F_i by its empirical counterpart $\hat{F}_i(\hat{Q}^\alpha) = n_i^{-1}\sum_{j=1}^{n_i} 1_{Y_{ij} \le \hat{Q}^\alpha}$ will not affect the limiting distribution of the test statistic.

Under the WEL approach, a $100(1-\rho)\%$ confidence interval for Q_0^α can be constructed as $\{Q^\alpha | -2r_w(Q^\alpha)/c < \chi^{2,\rho}_{(1)}\}$, where $\chi^{2,\rho}_{(1)}$ is the ρ-quantile from the χ^2 distribution with one degree of freedom. The ratio $r_w(Q^\alpha)$ is computable for any Q^α such that Q^α is in the convex hull formed by the overall sample. A bootstrap calibration of the confidence interval is also a possibility. See Fu et al. (2008) for details.

3 Simulation study

To assess the finite sample performance of our proposed methodology, we now present the results of some Monte Carlo simulations. As a benchmark, we consider the approach of Woodruff (1952) who basically suggested constructing confidence intervals for quantiles of complex surveys by inverting the confidence intervals of the distribution function. Sitter and Wu (2001) found this method to be quite reliable even in the moderate to extreme tail regions of distributions.

We consider a population divided into three strata with weights $0.50, 0.30$, and 0.20. The samples for the strata are independently generated from three lognormal distribution functions with means and standard deviations $(1.5, 0.3)$, $(2, 0.4)$, and $(2.1, 0.4)$. We use pooled sample sizes of $n = 50, n = 100$, $n = 200$, and construct 95% confidence intervals for seven different quantiles. For each specification, we conduct $5,000$ simulations. Table 1 reports the simulated coverage probability (CP), lower tail error rates (L), upper tail error rates (U), and the average length (AL) of the intervals. With the exception of the case where $n = 50$ and $\alpha = 0.05$, both confidence intervals seem to have excellent coverage rates even in the tails of the distribution. Interestingly, the tail error rates of the WEL interval seem to be much more balanced than Woodruff's. In the moderate to extreme tail regions (i.e., $\alpha = 0.05, 0.10, 0.90, 0.95$), WEL tends to slightly outperform Woodruff as WEL's coverage probabilities are closer to the nominal level of 95%. This is not true for all instances but the "overall picture" gives WEL a slight advantage. The quantiles towards the center of the distribution do not pose much problems (which is expected). The WEL intervals are roughly on par with Woodruff's.

Table 1: Simulated coverage and tail error rates for 95% confidence intervals

(n_1, n_2, n_3)		$Q^{0.05}$	$Q^{0.10}$	$Q^{0.25}$	$Q^{0.50}$	$Q^{0.75}$	$Q^{0.90}$	$Q^{0.95}$
Woodruff Confidence Interval for Quantile Q^{α}								
$(20, 20, 10)$	CP	82.94	94.58	95.48	95.52	95.42	96.24	94.86
	L	11.34	0.78	1.08	1.52	1.56	1.92	1.02
	U	5.72	4.64	3.44	2.96	3.02	1.84	4.12
	AL	0.91	1.34	1.28	1.56	2.66	5.30	7.51
$(40, 40, 20)$	CP	93.90	94.72	95.04	95.42	95.12	95.98	96.64
	L	0.98	0.68	1.38	1.84	1.96	1.72	1.72
	U	5.12	4.60	3.58	2.74	2.92	2.30	1.64
	AL	1.01	0.93	0.88	1.10	1.84	3.39	5.85
$(80, 80, 40)$	CP	94.62	94.36	94.92	95.26	94.70	95.60	95.60
	L	0.98	1.62	1.74	1.78	2.22	2.06	2.16
	U	4.40	4.02	3.34	2.96	3.08	2.34	2.24
	AL	0.73	0.62	0.62	0.77	1.28	2.31	3.59
WEL Confidence Interval for Quantile Q^{α}								
$(20, 20, 10)$	CP	87.34	94.38	95.56	94.56	94.74	94.50	94.00
	L	11.34	3.06	2.38	2.38	2.20	2.52	1.88
	U	1.32	2.56	2.06	3.06	3.06	2.98	4.12
	AL	1.09	1.30	1.28	1.57	2.65	5.08	7.36
$(40, 40, 20)$	CP	96.08	95.10	94.80	95.18	94.78	95.26	94.88
	L	1.84	2.78	2.64	2.32	2.56	1.94	2.18
	U	2.08	2.12	2.56	2.50	2.66	2.80	2.94
	AL	1.06	0.88	0.87	1.10	1.84	3.37	5.43
$(80, 80, 40)$	CP	94.64	94.78	95.10	95.18	94.48	95.22	94.68
	L	3.18	3.04	2.40	2.24	2.56	2.32	2.52
	U	2.18	2.18	2.50	2.58	2.96	2.46	2.80
	AL	0.66	0.61	0.62	0.77	1.28	2.28	3.44

4 Conclusion

Following up on the work of Fu et al. (2008), we proposed a weighted empirical likelihood-based inference method for quantiles in the presence of a stratified random sampling design. Our method is very easy to implement as computational routines are readily available. Through simulations, we were able to show that the confidence intervals obtained from our method perform just as well (and slightly better in some cases) as the popular method of Woodruff (1952). Thus, the WEL approach is a perfectly reliable method of inference for quantiles arising from stratified random samples.

So far, we have limited ourselves to inferences on a single measure (i.e., one quantile or one mean). But one may be interested in making simultaneous inference on multiple quantiles or means. The nature of complex surveys make the asymptotics much more difficult in such cases. Our work along with the work of Fu et al. (2008) provide partial guidance for future research into inference for a vector of measures.

References

Chen, J. and Sitter, R. R. (1999) “A pseudo empirical likelihood approach to the effective use of auxilary information in complex surveys” *Statistica Sinica* **9**, 385-406.

Cochran, W. G. (1977) *Sampling Techniques*, 3rd ed., John Wiley and Sons Inc.

Francisco, C. A. and Fuller, W. A. (1991) “Quantile estimation with a complex suvey design” *The Annals of Statistics* **19**, 454-469.

Fu, Y., Wang, X. and Wu, C. (2008) “Weighted empirical likelihood inference for multiple samples” *Journal of Statistical Planning and Inference* **139**, 1462-1473.

Gross, T. (1980) “Median estimation in sample surveys” in *Proceedings of Section on Survey Research Methods, American Statistical Association*, 181-184

Owen, A. B. (1988) “Empirical likelihood ratio confidence intervals for a single functional” *Biometrika* **75**, 237-249.

Owen, A. B. (1990) “Empirical likelihood ratio confidence regions” *The Annals of Statistics* **18**, 90-120.

Owen, A. B. (2001) *Empirical Likelihood*, Chapman and Hall/CRC.

Qin, J. and Lawless, J. (1994) “Empirical likelihood and general estimating equations” *The Annals of Statistics* **22**, 300-325.

Sitter, R. R. and Wu, C. (2001) “A note on Woodruff confidence intervals for quantiles” *Statistics & Probability Letters* **52**, 353-358

Woodruff, R. (1952) “Confidence intervals for medians and other position measures” *Journal of the American Statistical Association* **47**, 635-646.

Wu, C. (2004) “Some algorithmic aspects of the empirical likelihood method in survey sampling” *Statistica Sinica* **14**, 1057-1069.

Wu, C. (2005) “Algorithms and R codes for the pseudo empirical likelihood method in survey sampling” *Survey Methodology* **31**, 239-243.

Wu, C. and Rao, J. N. K. (2006) “Pseudo-empirical likelihood ratio confidence intervals for complex surveys” *The Canadian Journal of Statistics* **34**, 359-375.

Zhong, B. and Rao, J. N. K. (2000) “Empirical likelihood inference under stratified random sampling using auxilary population information” *Biometrika* **87**, 929-938.

Paper 1B

Economics Letters 123 (2014) 54–57

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Empirical likelihood-based inference for the generalized entropy class of inequality measures

Tahsin Mehdi, Thanasis Stengos *

Department of Economics and Finance, University of Guelph, Guelph, Ontario N1G 2W1, Canada

HIGHLIGHTS

- We propose an empirical likelihood-based method of inference for the generalized entropy class of inequality measures.
- We conduct a Monte Carlo study to assess the size and power of our proposed test.
- Simulations show that our method matches the performance of the delta method, and in some cases outperforms it.
- We apply our method to some Canadian household income data for illustrative purposes.

ARTICLE INFO

Article history:
Received 29 October 2013
Accepted 17 January 2014
Available online 28 January 2014

JEL classification:
C1
C12
I3

Keywords:
Empirical likelihood
Inequality measures
Entropy class

ABSTRACT

We propose an empirical likelihood-based method of inference for comparing inequality between two populations. A series of Monte Carlo experiments are used to assess our method's finite sample performance. We illustrate our approach using some Canadian household income data.

1. Introduction

Ever since the work of Atkinson (1970), there has been significant research interest in economic inequality and poverty. Although the measurement of inequality and poverty are important, statistical inference for such measures have gained considerable interest in recent years. The work of Kakwani (1993), Zheng (2001), Biewen (2002) and Davidson and Flachaire (2007) serve to highlight the importance of statistical inference in measuring inequality and poverty rather than just the incidence.

The growing body of literature surrounding the theory of inequality measurement has been accompanied by increasing availability of income data distribution which have armed researchers with the capability to conduct more sophisticated analyses. Statistical inference for inequality measures was largely neglected until the work of Cowell (1989). Recently, Thompson (2010) derived the asymptotic properties of vector measures of inequality (and poverty). He argued that since there is often no "best" measure of inequality or poverty, multiple measures could be used.

Our method of inference relies on empirical likelihood (EL), a powerful nonparametric statistical method pioneered by Owen (1988, 1990). An advantage of empirical likelihood is that no assumptions are needed regarding the underlying distribution of the data. Thompson (2013) used the approach for making inference on poverty measures which utilize relative poverty lines. His main focus was to compare poverty between two subgroups of a population that share a common poverty line. We depart from focusing on poverty measures and turn our attention to inequality measures (more specifically, we limit our focus to the generalized entropy class of inequality measures).

The remainder of this paper is organized as follows. In Section 2, we provide a brief overview of inequality measures. In Section 3, we present our methodology. In Section 4, we examine the finite sample performance of our method using a Monte Carlo

* Corresponding author. Tel.: +1 519 824 4120x53917; fax: +1 519 763 8497.
E-mail addresses: tmehdi@uoguelph.ca (T. Mehdi), tstengos@uoguelph.ca (T. Stengos).

0165-1765/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.econlet.2014.01.015

simulation. In Section 5, we demonstrate the practicality of our method using an empirical application.

2. Inequality measures

In this section, we provide a basic overview of the measurement of inequality. For a more thorough treatment of the literature, see Cowell (2011) or Cowell (2000). Before proceeding, we need to introduce some notation. Following Thompson (2010), we generalize our approach for vector measures of inequality. Let $Y = (Y_1, \ldots, Y_J)'$ be a random vector whose value is determined by a set of attributes (e.g., income, education, etc.) for an individual from a certain population. In the case where we are interested in only one attribute but we want to consider J distinct measures, we will have $Y_j = Y_k$ for all j, k. Let F_j be the distribution function of Y_j.

There are several different scalar measures of inequality that exist in the literature. We focus exclusively on the generalized entropy class of measures which fulfill the most widely accepted axioms including decomposability (see, e.g., Cowell, 2000).[1] For the random vector Y_j, such measures can be written as $I_j = E_j(h_j(y_j, \mu_j, \alpha_j))$ where E_j denotes expectation under distribution F_j, $h_j(y_j, \mu_j, \alpha_j)$ is some real-valued function, μ_j is the mean of F_j, and α_j is an exogenous parameter (and thus its choice is subjective). Formally, we have

$$I_j = \int h_j(y_j, \mu_j, \alpha_j) dF_j(y_j),$$

where $\mu_j = \int y_j dF_j(y_j)$ and

$$h_j(y_j, \mu_j, \alpha_j) = \begin{cases} [(y_j/\mu_j)^{\alpha_j} - 1]/(\alpha_j^2 - \alpha_j) & \alpha_j \neq 0, 1 \\ -\log(y_j/\mu_j) & \alpha_j = 0 \\ y_j \log(y_j/\mu_j)/\mu_j & \alpha_j = 1. \end{cases}$$

Let $\mu = (\mu_1, \ldots, \mu_J)'$ be the vector of means. A vector of inequality measures can be written as $I = (I_1, \ldots, I_J)'$.

3. Empirical likelihood-based inference

The empirical likelihood method was first brought to the forefront by Owen (1988, 1990). It is a nonparametric method of inference and an alternative to the bootstrap. For an extensive overview, see Owen (2001).

The basic framework can be explained as follows. Let $y_1, \ldots, y_n$ be independent observations with common distribution function F_0. For any distribution function F, let $p_i \geq 0$ be the probability associated with y_i, with $\sum_{i=1}^n p_i = 1$. Define $L(F) = \Pi_{i=1}^n p_i$ as the nonparametric likelihood function for F. Maximizing $L(F)$, subject to the constraints on p_i, yields $p_i = n^{-1}$. In other words, the nonparametric likelihood function attains its maximum when equal weight is placed on each observation.

Let $\theta_0 = T(F_0)$ be a J-dimensional parameter vector for some function T. Analogous to the parametric likelihood case, inferences about θ_0 can be made using the empirical likelihood ratio $L(F)/L(\hat{F})$, where $\hat{F}$ is the empirical distribution function.

Next, suppose we have r estimating functions $g(Y; \theta) = (g_1(Y; \theta), \ldots, g_r(Y; \theta))$ such that $E_F(g(Y; \theta)) = 0$. The main purpose of such functions is to identify the parameters of the problem. The profile empirical likelihood ratio function can then be written as

$$\mathcal{R}(\theta) = \max \left\{ \prod_{i=1}^n np_i \middle| p_i \geq 0, \ \sum_{i=1}^n p_i = 1, \ \sum_{i=1}^n p_i g(y_i; \theta) = 0 \right\}.$$

Under mild regularity conditions, it can be shown that $-2\log \mathcal{R}(\theta_0) \xrightarrow{d} \chi^2_{(J)}$.[2] Details on the computation of the profile likelihood ratio function can be found in Owen (2001, Chapter 3.14).[3]

Our main focus in this paper is to compare inequality between two distinct populations.[4] To distinguish between the two populations, let superscripts A and B hereby indicate association with population A and B, respectively. If we let $D_0 = (D_{1,0}, \ldots, D_{J,0}) = (I^B_{1,0} - I^A_{1,0}, \ldots, I^B_{J,0} - I^A_{J,0})$, we can test the null hypothesis that $D_0 = D$. Usually, applied researchers would be most interested in testing the null hypothesis that $I_0^A = I_0^B$, which is equivalent to testing $D_0 = 0$. To apply the empirical likelihood-based inference method to the generalized entropy class of inequality measures, we need to encode the parameters of our problem into suitable estimating functions.

Given that we are interested in comparing two populations, the profile empirical likelihood ratio function is

$$\begin{aligned} \mathcal{R}(\theta^A, \theta^B) = \max \Bigg\{ & \prod_{i=1}^{n^A} n^A p_i^A \prod_{i=1}^{n^B} n^B p_i^B \Bigg| p_i^A \geq 0, p_i^B \geq 0, \\ & \sum_{i=1}^{n^A} p_i^A = 1, \ \sum_{i=1}^{n^B} p_i^B = 1, \\ & \sum_{i=1}^{n^A} p_i^A g(y_i^A; \theta^A) = 0, \ \sum_{i=1}^{n^B} p_i^B g(y_i^B; \theta^B) = 0 \Bigg\}, \end{aligned}$$

where $\theta^A = (\mu^A, I^A)$, $\theta^B = (\mu^B, I^A, D)$, and the estimating functions are

$$g(Y^A; \theta^A) = \begin{pmatrix} Y_1^A - \mu_1^A \\ \vdots \\ Y_J^A - \mu_J^A \\ h_1(Y_1^A, \mu_1^A, \alpha_1) - I_1^A \\ \vdots \\ h_J(Y_J^A, \mu_J^A, \alpha_J) - I_J^A \end{pmatrix},$$

and

$$g(Y^B; \theta^B) = \begin{pmatrix} Y_1^B - \mu_1^B \\ \vdots \\ Y_J^B - \mu_J^B \\ h_1(Y_1^B, \mu_1^B, \alpha_1) - I_1^A - D_1 \\ \vdots \\ h_J(Y_J^B, \mu_J^B, \alpha_J) - I_J^A - D_J \end{pmatrix}.$$

Since we are only interested in conducting hypotheses on D, the remaining parameters in the μ^A, μ^B and I^A vectors are regarded as "nuisance" parameters. Following Owen (1990), we can "profile out" such parameters by maximizing over them. So the empirical likelihood ratio function for D is

$$\mathcal{R}(D) = \max_{\mu^A, \mu^B, I^A} \mathcal{R}(\mu^A, \mu^B, I^A, D).$$

To compute $\mathcal{R}(D)$ for any vector D, we can follow Owen (1990) and use a nested algorithm which involves an "inner" and "outer"

[1] The Atkinson class of inequality measures, and the Gini index are some of the other well established measures of inequality.

[2] A bootstrap calibration is also possible (see Owen, 2001, Chapter 3.3).

[3] Computational routines for several statistical packages are available on Owen's website: http://www-stat.stanford.edu/~owen/empirical/.

[4] There have been numerous studies done on empirical likelihood for the two population case (see, e.g., Wu and Yan, 2012).

stage. The former involves solving $\mathcal{R}(\mu^A, \mu^B, I^A, D)$ given candidate values for μ^A, μ^B and I^A while the latter involves maximizing $\mathcal{R}(\mu^A, \mu^B, I^A, D)$ over the choices of μ^A, μ^B and I^A. The initial candidate values for such parameters could simply be the sample estimates. By maximizing over these parameter vectors, we reduce the empirical likelihood ratio function into a function of only D which is J-dimensional. Therefore, we have $-2 \log \mathcal{R}(D_0) \xrightarrow{d} \chi^2_{(J)}$.

4. Simulation evidence

To assess the size and power of our empirical likelihood-best test, we now present the results of some Monte Carlo experiments. All our hypothesis tests concern a single random variable (i.e., $Y_j^A = Y_k^A$ and $Y_j^B = Y_k^B$ for all j, k), and a single scalar measure of inequality with $\alpha = 0$.

To calibrate the size of our proposed method, we consider two parametric distributions: the gamma distribution and the Singh–Maddala distribution. The cumulative distribution function for the gamma distribution is given by $F(y) = \gamma(a_2, y/a_1)/\Gamma(a_2)$, where a_1 is a scale parameter, a_2 is a shape parameter, $\gamma(\cdot)$ is the gamma function, and $\Gamma(\cdot)$ is the incomplete gamma function. The cumulative distribution function for the Singh–Maddala distribution is $F(y) = 1 - (1 + b_1 y^{b_2})^{-b_3}$, where b_1 is a scale parameter, and b_2 and b_3 are shape parameters. Following McDonald (1984), we set $a_2 = 2.1557$ for the gamma distribution, and $b_2 = 1.697$ and $b_3 = 8.368$ for the Singh–Maddala distribution which closely mimic the 1980 US income distribution. The scale parameters for both distributions are set to unity.[5] Given these specifications, the true inequality measures under the gamma and Singh–Maddala distributions are 0.2495227 and 0.2488523, respectively.

Samples for both populations are generated from the same distribution. We test the null hypothesis that $D_0 = 0$, which is true. The nominal size of the test is set to 5%. Rejection frequencies for 100,000 independent trials, for sample sizes varying from $(n^A, n^B) = (100, 100)$ to $(n^A, n^B) = (500, 500)$, are reported in Table 1. As a benchmark, we also report the rejection frequencies for the delta method.[6] Although the empirical likelihood-based method seems to over reject in small samples, it is evident that the errors in rejection probability (difference between the simulated and nominal rejection rates) subside as the sample sizes are increased. Even when we have just 300 observations on each population, these errors are less than one-half of one percentage point for both distributions.

To assess the power of our testing methodology, we conduct two Monte Carlo experiments. For the first experiment, we consider generating both samples from the gamma distribution. The second experiment involves generating both samples from the Singh–Maddala distribution. The shape parameters, of population A, for the gamma and Singh–Maddala distribution remain set to their previous values from the size simulation. We vary our choices for the shape parameters of population B so that D_0, the difference in inequality measures between population B and population A, is pre-specified to be −0.10, −0.05, 0.05, and 0.10.

As in the size simulations, we set the nominal size to 5% and test the null hypothesis that $D_0 = 0$ (which is false in all cases here). We consider sample sizes of $(n^A, n^B) = (100, 100)$ and $(n^A, n^B) = (500, 500)$, and conduct 100,000 independent trials. To correct for size distortions, the critical value of the test statistic is set equal to its 95,000th largest value from the size simulation. Rejection frequencies along with our choices of shape parameters (for population B) for the first and second experiment are reported in Table 2 and Table 3, respectively. We also report the (size-corrected) rejection frequencies for the delta method. From Table 2, it is apparent that the two testing methodologies perform almost identically for the first experiment (though the empirical likelihood-based method holds a slim advantage with 100 observations on each population). But when the samples are both drawn from the Singh–Maddala distribution as in Table 3, the empirical likelihood-based method clearly outperforms the delta method when $D_0 = 0.05, 0.10$. In fact, with 500 observations, our method has rejection frequencies that are more than five percentage points higher than those of the delta method. Overall, we can conclude that the empirical likelihood-based approach certainly matches the performance of the delta method, and in some cases outperforms it.

Table 1
Rejection frequencies for size simulation.

Distribution	n^A, n^B				
	EL				
	100, 100	200, 200	300, 300	400, 400	500, 500
Gamma	5.769	5.403	5.250	5.171	5.067
Singh–Maddala	6.215	5.615	5.291	5.281	5.118
	Delta method				
Gamma	4.751	4.897	4.938	4.954	4.896
Singh–Maddala	4.705	4.982	4.919	4.986	4.908

Table 2
Rejection frequencies for power simulation with gamma distributions.

D_0	a_2^B	n^A, n^B			
		EL		Delta method	
		100, 100	500, 500	100, 100	500, 500
−0.10	3.501874	71.115	99.992	70.562	99.991
−0.05	2.660860	19.841	70.527	19.708	70.056
0.05	1.818233	14.964	53.615	14.865	54.100
0.10	1.576528	38.405	96.440	38.312	96.530

Notes: a_2^B is the shape parameter for population B.

Table 3
Rejection frequencies for power simulation with Singh–Maddala distributions.

D_0	b_2^B	b_3^B	n^A, n^B			
			EL		Delta method	
			100, 100	500, 500	100, 100	500, 500
−0.10	2.28	6.32	62.853	99.951	64.610	99.957
−0.05	2.01	4.45	17.243	64.067	17.913	64.098
0.05	2.09	1.38	11.143	39.256	9.118	33.982
0.10	2.08	1.17	25.120	83.727	20.146	77.966

Notes: b_2^B and b_3^B are the shape parameters for population B.

5. Empirical application

In this section, we illustrate our proposed methodology using a "real-world" application where we compare income inequality, using after-tax income data, between two populations in Canada: non-immigrants and immigrants.

We obtain our data from the Survey of Labour and Income Dynamics (SLID) for 2009. To reduce the level of heterogeneity

[5] Zheng (2001) used the Singh–Maddala distribution with the same parameter specifications to analyze the asymptotic properties of decomposable poverty measure estimates with relative poverty lines. Thompson (2013) used both distributions with the same parameters to assess the size and power of his testing procedure for comparing poverty measures between two subgroups of a population.

[6] The delta method can be used to obtain variances of inequality measure estimates belonging to the generalized entropy class (see, e.g., Thompson, 2010).

Table 4
Sample estimates for empirical application.

Population	μ	I_1	I_2	I_3
Non-immigrants	25,019	0.547714	0.417437	0.378834
Immigrants	17,132	0.675557	0.501120	0.449005

within samples, we restrict our analysis to single (never married) individuals who reside in urban areas with a population of 500,000 or greater. For the purposes of this illustration, we consider only those immigrants who have been in Canada for 19 years or less. Our sample consists of 3093 non-immigrants and 393 immigrants (i.e., $n^A = 3093$, and $n^B = 393$).

We consider three distinct scalar measures of inequality by setting $\alpha_1 = 0, \alpha_2 = 0.5$, and $\alpha_3 = 1$. The sample estimates of the population means and inequality measures are reported in Table 4. The nominal size of the test is set to 5%. Given three measures of inequality, the appropriate degrees of freedom for the null distribution is three. In testing $D_0 = 0$, our test statistic is determined to be $-2 \log \mathcal{R}(0) = 9.851083$ which exceeds the 95th percentile of the $\chi^2_{(3)}$ distribution (i.e., 7.814728). Thus, our null hypothesis is rejected leading us to conclude that income inequality between non-immigrants and immigrants in Canada may not be equal.

References

Atkinson, A.B., 1970. On the measurement of inequality. J. Econom. Theory 12, 244–263.
Biewen, M., 2002. Bootstrap inference for inequality, mobility and poverty measurement. J. Econometrics 108, 317–342.
Cowell, F.A., 1989. Sampling variance and decomposable inequality measures. J. Econometrics 42, 27–41.
Cowell, F.A., 2000. Measurement of inequality. In: Atkinson, A.B., Bourguignon, F. (Eds.), Handbook of Income Distribution, Vol. I. North-Holland, Amsterdam.
Cowell, F.A., 2011. Measuring Inequality, third ed. Oxford University Press, Oxford.
Davidson, R., Flachaire, E., 2007. Asymptotic and bootstrap inference for inequality and poverty measures. J. Econometrics 141, 141–166.
Kakwani, N., 1993. Statistical inference in the measurement of poverty. Rev. Econ. Stat. 75, 632–639.
McDonald, J., 1984. Some generalized functions for the size distribution of income. Econometrica 52, 647–663.
Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.
Owen, A.B., 1990. Empirical likelihood ratio confidence regions. Ann. Statist. 18, 90–120.
Owen, A.B., 2001. Empirical Likelihood. Chapman and Hall/CRC, Boca Raton.
Thompson, B.S., 2010. Statistical inference for vector measures of inequality and poverty. J. Econ. Inequal. 8, 451–462.
Thompson, B.S., 2013. Empirical likelihood-based inference for poverty measures with relative poverty lines. Econometric Rev. 32, 513–523.
Wu, C., Yan, Y., 2012. Empirical likelihood inference for two-sample problems. Stat. Interface 5, 345–354.
Zheng, B., 2001. Statistical inference for poverty measures with relative poverty lines. J. Econometrics 101, 337–356.

Paper 1C

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2017, VOL. 46, NO. 4, 2029–2036
http://dx.doi.org/10.1080/03610926.2015.1040504

Poverty comparisons with common relative poverty lines

Tahsin Mehdi

Department of Economics, University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT

Relative poverty lines are increasingly being used in poverty comparison studies. Existing methods assume that the distributions being compared are distinct with independent relative poverty lines. However, this practice may be problematic when comparing two subgroups of a population. We follow up on a recent proposal for the usage of common relative poverty lines in such cases, and develop a test for comparing poverty between subgroups of a single population, using inequality restrictions. Monte Carlo experiments are conducted in order to examine the size and power of our proposed test. We illustrate our procedure using some U.S. household income data.

ARTICLE HISTORY
Received 9 September 2014
Accepted 26 March 2015

KEYWORDS
Bootstrap inference; Income distribution; Relative poverty.

MATHEMATICS SUBJECT CLASSIFICATION
91B82; 62F03

1. Introduction

Inception of poverty measurement into economic literature was largely propagated by the work of Sen (1976) who brought the topic to the forefront. As a result, there has been a paradigm shift in the way poverty is viewed and measured. Over the past two decades, great advancements in inference for poverty measures have been made due to works such as Kakwani (1993), Bishop et al. (1995), Bishop et al. (1997), Davidson and Duclos (2000), Zheng (2001), and Biewen (2002).

Poverty measurement typically entails setting or defining a *poverty line* (i.e., a threshold level of income or some other measure of individual welfare that divides the "poor" and "non poor"). A poverty line can be specified as one that is absolute or relative. The former is usually determined by analysts who base their decision on their own interpretation of economic factors. Such poverty lines need to be revised frequently to reflect changes in economic conditions. To alleviate this issue, an alternative is to specify the poverty line as a fraction of either the mean or some quantile of the underlying income distribution. These are known as relative poverty lines and they have gained momentum in recent years.[1]

In this article, we focus exclusively on decomposable (additively separable) poverty measures that utilize relative poverty lines.[2] Asymptotic properties of such measures remained rather elusive, until the work of Zheng (2001). His work is closely related to that of Preston (1995) who also considered relative poverty lines but only for the headcount ratio (i.e., the proportion of the population with income levels below the poverty line). Zheng's main focus was

CONTACT Tahsin Mehdi tmehdi@uwaterloo.ca Department of Economics, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

[1] For example, the Organisation for Economic Co-operation and Development (OECD) often reports poverty rates for its member countries by setting poverty lines to 40%, 50%, and 60% of the national median income level.

[2] Such poverty measure estimates themselves require an estimate of the underlying poverty line and therefore the sampling variance of the poverty line must be taken into consideration.

on comparing poverty between two distinct populations each with its own relative poverty line. He developed a procedure for testing the null hypothesis that $P^{\mathrm{B}} \geq P^{\mathrm{A}}$, where P^{G} is a vector of poverty measures for population G, for G = A, B.

Recently, Thompson (2013) proposed a novel, empirical likelihood (EL)-based method of inference, for comparing poverty between two subgroups of a population that share a *common* relative poverty line. His particular focus was on testing the null hypothesis that $P^{\mathrm{B}} = P^{\mathrm{A}}$, under the assumption that the poverty line is some fraction of the "pooled" median income level. This may be sensible when comparing poverty between, say, males and females from a certain population.

Often, applied researchers may be more interested in determining if one subgroup is better (or worse) off. So we depart from Thompson's (2013) null of equality and devise a method for testing null hypotheses of the form $P^{\mathrm{B}} \geq P^{\mathrm{A}}$ (under the assumption of common relative poverty lines). We focus on quantile-based poverty lines but our test can be adapted to the case of mean-based poverty lines.

The remainder of this article is organized as follows. In Section 2, we provide a brief overview of decomposable poverty measures. In Section 3, we lay out the asymptotic framework and develop our testing procedure. In Section 4, we conduct a series of Monte Carlo experiments in order to assess the size and power of our test. In Section 5, we provide a brief illustration of our procedure using some U.S. household income data. Section 6 concludes.

2. Decomposable poverty measures

As stated earlier, we focus exclusively on decomposable poverty measures, which are a class of poverty measures that are additively separable.

Let Y be a continuous random variable whose value is determined by income or some other measure of individual welfare. Let F be its distribution function. For a given poverty line z, an individual with income y is classified as being poor if $y \leq z$, and non poor if $y > z$.

A decomposable poverty measure is one which can be written as $P = E_F[p(y; z)]$, where E_F denotes the expectation under distribution F, and $p(y; z)$ is an individual deprivation function such that $p(y; z) \geq 0$ for $y \leq z$, and $p(y; z) = 0$ for $y > z$.[3] Several deprivation functions have been proposed in the literature (see, e.g., Zheng, 2001, Table 1) and it may seem as though there is no clear guidance as to which one should be chosen. However, one can select it based on axiomatic grounds (see Zheng, 1997, for a discussion on the desirability of axioms and poverty measures).

The aim of this article is to develop a framework which will allow us to make poverty comparisons between two subgroups of a population that share a common relative poverty line. Suppose a finite population of size N consists of N_{A} individuals who belong to a subgroup, denoted A, and N_{B} individuals who belong to another subgroup, denoted B. Let F_{A} and F_{B} be the distribution functions that generate the underlying variable of interest (e.g., income) in each subgroup. The population level distribution function is simply $F(\cdot) = w_{\mathrm{A}}F_{\mathrm{A}}(\cdot) + w_{\mathrm{B}}F_{\mathrm{B}}(\cdot)$, where $w_{\mathrm{A}} = N_{\mathrm{A}}/N$ and $w_{\mathrm{B}} = N_{\mathrm{B}}/N$. We assume a common quantile-based poverty line $z = c\xi_q$ for some fraction $c \in [0, 1]$, where ξ_q is a population quantile of order q (i.e., $F(\xi_q) = q$).[4] Then the poverty measures for subgroups A and B, respectively, are just $E_{\mathrm{A}}[p(y; z)]$ and $E_{\mathrm{B}}[p(y; z)]$, where E_{G} denotes the expectation under distribution F_{G}.

[3] The restrictions on the individual deprivation function follow from the "focus axiom" (see Foster, 1984).

[4] If one is interested in a common mean-based poverty line, then $z = c\mu$ where $\mu = w_{\mathrm{A}}E_{\mathrm{A}}(y) + w_{\mathrm{B}}E_{\mathrm{B}}(y)$. In this case, the asymptotic expression (1) would need to be changed accordingly.

Since there may be conflicting opinions on the choice of c, one may want to make poverty comparisons more "robust" by considering J different poverty lines with the jth poverty line being $z_j = c_j\xi_q$ for $j = 1, \ldots, J$. Then the vector poverty measures for subgroups A and B, respectively, are $P^{\mathrm{A}} = (P_1^{\mathrm{A}}, \ldots, P_J^{\mathrm{A}}) = (E_{\mathrm{A}}[p(y; z_1)], \ldots, E_{\mathrm{A}}[p(y; z_J)])$ and $P^{\mathrm{B}} = (P_1^{\mathrm{B}}, \ldots, P_J^{\mathrm{B}}) = (E_{\mathrm{B}}[p(y; z_1)], \ldots, E_{\mathrm{B}}[p(y; z_J)])$.

3. Estimation and inference

In this section, we establish the asymptotic properties required for comparing poverty between distributions with common quantile-based poverty lines. Our work will be greatly facilitated by Zheng (2001) who laid much of the groundwork but under the assumption that each distribution has its own relative poverty line.

Let $\{y_i^{\mathrm{A}}\}_{i=1}^{n_{\mathrm{A}}}$ and $\{y_i^{\mathrm{B}}\}_{i=1}^{n_{\mathrm{B}}}$ be random *iid* draws from F_{A} and F_{B}, respectively, and let $n = n_{\mathrm{A}} + n_{\mathrm{B}}$ be the pooled sample size. Assume that $n \to \infty$ implies $n_{\mathrm{G}} \to \infty$, and $n_{\mathrm{G}}/N_{\mathrm{G}}$ is sufficiently small so that no finite population adjustment is necessary. Further, assume that the sampling proportions coincide with the population (i.e., $n_{\mathrm{G}}/n = w_{\mathrm{G}}$) The jth poverty measure for subgroup G can be consistently estimated by $\hat{P}_j^{\mathrm{G}} = n_{\mathrm{G}}^{-1} \sum_{i=1}^{n_{\mathrm{G}}} p(y_i^{\mathrm{G}}; \hat{z}_j)$, where $\hat{z}_j = c_j y_{(r)}, y_{(r)}$ is the rth order statistic of the pooled sample $y = (y^{\mathrm{A}}, y^{\mathrm{B}})'$ with $r = \left[nq\right]$, and $\left[nq\right]$ is the integer part of nq.

Let the vector of differences in poverty measures between subgroups B and A be given by $D = (D_1, \ldots, D_J) = (P_1^{\mathrm{B}} - P_1^{\mathrm{A}}, \ldots, P_J^{\mathrm{B}} - P_J^{\mathrm{A}})$. Using similar arguments as Zheng (2001, Section 4.2), an asymptotic expression for the jth difference is

$$D_j = \frac{1}{n_{\mathrm{B}}} \sum_{i=1}^{n_{\mathrm{B}}} p\left(y_i^{\mathrm{B}}; z_j\right) - \frac{1}{n_{\mathrm{A}}} \sum_{i=1}^{n_{\mathrm{A}}} p\left(y_i^{\mathrm{A}}; z_j\right) + c_j s_j (y_{(r)} - \xi_q) + o_p\left(n^{-1/2}\right) \tag{1}$$

where $s_j = a_j^{\mathrm{B}} - a_j^{\mathrm{A}} + p(z_j; z_j)[f_{\mathrm{B}}(z_j) - f_{\mathrm{A}}(z_j)]$, $a_j^{\mathrm{G}} = \partial P_j^{\mathrm{G}}/\partial z_j$, and f_{G} is the underlying density function of distribution F_{G}.

Using the Bahadur representation (see, e.g., Zheng, 2001, pp. 351), we can express the difference between the sample quantile and population quantile as

$$y_{(r)} - \xi_q = \frac{q - \hat{F}(\xi_q)}{f(\xi_q)} + o_p\left(n^{-1/2}\right) \tag{2}$$

where $\hat{F}(\xi_q) = n^{-1} \sum_{i=1}^{n} I(y_i \leq \xi_q)$, $I(\cdot)$ is an indicator function which evaluates to 1 if its argument $(\cdot)$ is true and 0 otherwise, and $f(\cdot)$ is the underlying population density function. Thus, (1) becomes

$$D_j = \frac{1}{n_{\mathrm{B}}} \sum_{i=1}^{n_{\mathrm{B}}} p\left(y_i^{\mathrm{B}}; z_j\right) - \frac{1}{n_{\mathrm{A}}} \sum_{i=1}^{n_{\mathrm{A}}} p\left(y_i^{\mathrm{A}}; z_j\right) - \frac{c_j s_j}{f(\xi_q)} \hat{F}(\xi_q) + \frac{c_j s_j q}{f(\xi_q)} + o_p\left(n^{-1/2}\right) \tag{3}$$

Under the conditions that F_{A} and F_{B} are differentiable and have finite first two moments, D can be consistently estimated by $\hat{D} = (\hat{D}_1, \ldots, \hat{D}_J) = (\hat{P}_1^{\mathrm{B}} - \hat{P}_1^{\mathrm{A}}, \ldots, \hat{P}_J^{\mathrm{B}} - \hat{P}_J^{\mathrm{A}})$, so it follows that $\sqrt{n}(\hat{D} - D)$ will converge in distribution to a normal random vector with mean vector zero and covariance matrix Σ which has typical element

$$\begin{aligned} \mathrm{Cov}(\hat{D}_j, \hat{D}_k) = {} & \{E_{\mathrm{A}}[p(y; z_j)p(y; z_k)] - P_j^{\mathrm{A}} P_k^{\mathrm{A}}\}/w_{\mathrm{A}} + \{E_{\mathrm{B}}[p(y; z_j)p(y; z_k)] - P_j^{\mathrm{B}} P_k^{\mathrm{B}}\}/w_{\mathrm{B}} \\ & - c_k s_k \{P_j^{\mathrm{B}}[1 - F_{\mathrm{B}}(\xi_q)] - P_j^{\mathrm{A}}[1 - F_{\mathrm{A}}(\xi_q)]\}/f(\xi_q) \\ & - c_j s_j \{P_k^{\mathrm{B}}[1 - F_{\mathrm{B}}(\xi_q)] - P_k^{\mathrm{A}}[1 - F_{\mathrm{A}}(\xi_q)]\}/f(\xi_q) \\ & + c_j c_k s_j s_k \{w_{\mathrm{A}} F_{\mathrm{A}}(\xi_q)[1 - F_{\mathrm{A}}(\xi_q)] + w_{\mathrm{B}} F_{\mathrm{B}}(\xi_q)[1 - F_{\mathrm{B}}(\xi_q)]\}/f^2(\xi_q), \quad \forall j, k \end{aligned}$$

In practice Σ can be consistently estimated by $\hat{\Sigma}$ which has typical element

$$\begin{aligned}
\hat{\text{Cov}}(\hat{D}_j, \hat{D}_k) = & \left[\frac{1}{n_\text{A}} \sum_{i=1}^{n_\text{A}} p(y_i^\text{A}; \hat{z}_j) p(y_i^\text{A}; \hat{z}_k) - \hat{P}_j^\text{A} \hat{P}_k^\text{A} \right] \frac{n}{n_\text{A}} \\
& + \left[\frac{1}{n_\text{B}} \sum_{i=1}^{n_\text{B}} p(y_i^\text{B}; \hat{z}_j) p(y_i^\text{B}; \hat{z}_k) - \hat{P}_j^\text{B} \hat{P}_k^\text{B} \right] \frac{n}{n_\text{B}} \\
& - c_k \hat{s}_k \{ \hat{P}_j^\text{B} [1 - \hat{F}_\text{B}(y_{(r)})] - \hat{P}_j^\text{A} [1 - \hat{F}_\text{A}(y_{(r)})] \} / \hat{f}(y_{(r)}) \\
& - c_j \hat{s}_j \{ \hat{P}_k^\text{B} [1 - \hat{F}_\text{B}(y_{(r)})] - \hat{P}_k^\text{A} [1 - \hat{F}_\text{A}(y_{(r)})] \} / \hat{f}(y_{(r)}) \\
& + c_j c_k \hat{s}_j \hat{s}_k \{ n_\text{A} \hat{F}_\text{A}(y_{(r)}) [1 - \hat{F}_\text{A}(y_{(r)})] \\
& + n_\text{B} \hat{F}_\text{B}(y_{(r)}) [1 - \hat{F}_\text{B}(y_{(r)})] \} / [n \hat{f}^2(y_{(r)})], \quad \forall j, k,
\end{aligned}$$

where $\hat{F}_\text{G}(y_{(r)}) = n_\text{G}^{-1} \sum_{i=1}^{n_\text{G}} I(y_i^\text{G} \leq y_{(r)})$, $\hat{s}_j = \hat{a}_j^\text{B} - \hat{a}_j^\text{A} + p(\hat{z}_j; \hat{z}_j)[\hat{f}_\text{B}(\hat{z}_j) - \hat{f}_\text{A}(\hat{z}_j)]$, $\hat{a}_j^\text{G} = \partial \hat{P}_j^\text{G} / \partial \hat{z}_j$, and $\hat{f}_\text{G}$ is the estimated underlying density function of F_G. A well-known method for estimating densities is kernel estimation.[5] The estimate of the underlying population density function is $\hat{f}(\cdot) = n_\text{A} n^{-1} \hat{f}_\text{A}(\cdot) + n_\text{B} n^{-1} \hat{f}_\text{B}(\cdot)$. Buskirk (1998) showed that weighted kernel density estimators (as is the case here) will indeed be consistent.

Recall from Section 1 that our focus is on testing the null hypothesis that $P^\text{B} \geq P^\text{A}$, which means subgroup B has (weakly) higher poverty than subgroup A for all poverty lines considered. We may equivalently state the null hypothesis as

$$H_0 : D \geq 0$$

The alternate hypothesis is simply the negation of H_0.

For our purposes, the relevant methods can be found in Kodde and Palm (1986) and Wolak (1989) who developed the framework for testing multivariate inequality restrictions.[6] First, we compute the Wald-type test statistic

$$W = \min_{D \geq 0} n(\hat{D} - D)' \hat{\Sigma}^{-1} (\hat{D} - D) \tag{4}$$

where the right-hand side is a quadratic programming problem. Then we obtain a p-value in order to decide whether the null should be rejected or not. Under the null, W will converge in distribution to a mixture of χ^2 distributions.

Since obtaining the critical values can be rather cumbersome, we follow Stengos and Thompson (2012) and advocate the use of the bootstrap. The procedure can be explained as follows. Given samples y^A and y^B of sizes n_A and n_B, respectively, we pool the samples and obtain $y = (y^\text{A}, y^\text{B})'$. Then the bootstrap samples y_*^A and y_*^B are generated by resampling n_A and n_B observations (with replacement) from y. Next, using the bootstrap samples, we compute the bootstrap test statistic W^* in a similar manner to W. After repeating this process a large number of times, the bootstrap p-value is the proportion of times that W^* exceeds W. A value less than the nominal size of the test should lead to the rejection of H_0.

[5] The consistency of such estimators has been rigorously established in the literature (see, e.g., Li and Racine, 2007).

[6] Zheng (2001) suggested using the approach for testing a similar type of hypothesis (but under the assumption of distinct relative poverty lines). Stengos and Thompson (2012) also use it in testing for bivariate stochastic dominance.

Table 1. Rejection frequencies for size simulation.

	n			
Distribution	100	300	500	1,000
Gamma	5.37	4.93	4.53	4.97
Singh–Maddala	4.82	4.90	4.60	4.98
Log-normal	4.91	4.88	4.88	4.98
Unit Exponential	4.94	5.48	5.15	5.03
Uniform	4.85	4.94	4.74	5.04

Note: The nominal size of the test is 5%. The poverty lines are set to 40%, 50%, and 60% of the median.

4. Simulation evidence

We now present the results from a series of Monte Carlo experiments which were used to assess the size and power of our test. The hypothesis tests here were conducted with a nominal size set to 5%. Each of our experiments were carried out using 10,000 independent trials, and sample sizes of $n_{\mathrm{A}} = n_{\mathrm{B}} = n/2$. The deprivation function is set to $p(y; z) = I(y \leq z)$ (i.e., the headcount ratio), for which $a_j^{\mathrm{G}} = 0$ (see, e.g., Zheng, 2001, Table 1).

In assessing the size of our test, five different parametric distributions are considered: gamma, Singh–Maddala, log-normal, unit exponential, and uniform. The latter four distributions were used in Zheng (2001). The former two were used in Thompson (2013). The cumulative distribution function of the gamma distribution is given by $F(y) = \gamma(a_2, y/a_1)/\Gamma(a_2)$, where a_1 is a scale parameter, a_2 is a shape parameter, $\gamma(\cdot)$ is the gamma function, and $\Gamma(\cdot)$ is the incomplete gamma function. The cumulative distribution function of the Singh–Maddala distribution is given by $F(y) = 1 - (1 + b_1 y^{b_2})^{-b_3}$, where b_1 is a scale parameter, and b_2 and b_3 are shape parameters. Following McDonald (1984), we set $a_2 = 2.1557$ for the gamma distribution, and $b_2 = 1.697$ and $b_3 = 8.368$ for the Singh–Maddala distribution, which were used to simulate 1980 U.S. income distribution. The scale parameters for both distributions are set to unity. For the log-normal distribution, the mean and standard deviation are set to 2.9372 and 0.7797, respectively, which were also used by McDonald (1984) to simulate 1980 U.S. income distribution. For the uniform distribution, we follow Zheng (2001) and specify the support as the unit interval [0, 1]. We consider three different poverty lines set to 40%, 50%, and 60% of the median (i.e., $q = 0.5$).[7]

To assess the size of the test, observations for subgroups A and B are generated from the same distribution. We test the null hypothesis that $D \geq 0$ which is (weakly) true in this case. We consider pooled sample sizes varying from $n = 100$ to $n = 1,000$ and utilize 199 bootstrap replications. Rejection frequencies reported in Table 1 indicate good size properties. Overall, we can conclude that $n = 1,000$ should be sufficient for our proposed test to be reliable. This is not a very demanding requirement at all as typical household survey datasets tend to have thousands of observations.

In assessing the power of our test, we focus exclusively on the gamma distribution. The shape and scale parameters for subgroup A remain set to their original levels from the size simulation. For subgroup B, we vary the parameters following Thompson (2013), so as to induce $D < 0$. Rejection frequencies based on 199 bootstrap replications, along with shape and scale parameters for subgroup B, are reported in Table 2. Our test exhibits good power properties. Clearly, when sample sizes are small and D is close to 0, power is quite low but with increased sample sizes it approaches full power quite quickly.

[7] As discussed earlier, quantile-based poverty lines require density estimation in calculating the covariance structure. We use kernel estimation with a Gaussian kernel and a "rule-of-thumb" bandwidth (see, e.g., Li and Racine, 2007, Ch. 1).

Table 2. Rejection frequencies for power simulation.

a_1^B	a_2^B	D_1	D_2	D_3	n 100	300	500	1,000
0.651	3.143	− 0.05	− 0.05	− 0.05	16.80	33.75	47.47	73.20
0.421	4.688	− 0.10	− 0.10	− 0.10	43.67	84.57	96.52	99.99

Note: The nominal size of the test is 5%. D_j is the jth difference in poverty measures between subgroups B and A with poverty lines set to 40%$(j = 1)$, 50%$(j = 2)$, and 60%$(j = 3)$ of the median. Both distributions are generated from the gamma distribution. The scale parameter of distribution A is set to unity and its shape parameter is set to 2.1557.

In the final experiment, we examine power again but using Zheng (2001) and Thompson (2013) as benchmarks. We consider a single poverty line set to one-half the median. In order to ensure "fairness" to Thompson's (2013) EL-based test which has the null of *equality* of poverty, we briefly depart from our null involving inequality, and test $D = 0$. Our test and that of Zheng (2001) can easily accommodate this type of hypothesis. In fact, for a single poverty line, both tests can be done using t statistics which will have limiting standard normal distributions under the null. Thompson's (2013) EL test statistic will have a limiting χ_1^2 distribution. The shape and scale parameters for subgroup A remain set to their original levels while those for subgroup B are varied using Thompson's (2013) values which induce $D = (-0.10, -0.05, 0.05, 0.10)$ while maintaining the same median for both distributions (this ensures fairness to Zheng's (2001) test). Rejection frequencies (size-corrected) for $n = 100$ and $n = 1,000$ along with the shape and scale parameter values for subgroup B are reported in Table 3. To correct for size distortions, we first conducted the trials by generating both sets of observations using the same shape and scale parameters from the size simulation. For our test and that of Zheng (2001), the lower-tail and upper-tail critical values were then set equal to the 250th and 9,750th largest test statistics, respectively. For Thompson's (2013) test, the critical value was set equal to the 9,500th largest test statistic. The results suggest that our test and that of Thompson (2013) are on par which is not at all that surprising given that both use a common relative poverty line. Zheng's (2001) test appears to be more powerful but one must remember that his test is based on the assumption of distinct relative poverty lines which may be associated with relatively lower sampling variances.

5. Illustration

In this section, we illustrate our approach using some household income data obtained from the U.S. Census Bureau's American Community Survey (ACS) for 2012. Specifically, we compare poverty between males and females.

Table 3. Rejection frequencies for power simulation with null of equality.

			W		Zheng		EL	
D	a_1^B	a_2^B	$n = 100$	$n = 1,000$	$n = 100$	$n = 1,000$	$n = 100$	$n = 1,000$
−0.10	0.421	4.688	27.90	99.43	31.37	99.70	27.89	99.49
−0.05	0.651	3.143	9.16	55.36	10.85	62.58	9.28	56.76
0.05	1.588	1.471	9.19	48.63	8.81	58.77	9.67	47.95
0.10	2.725	0.979	21.63	95.29	23.45	99.27	21.99	95.03

Note: The null here is $D = 0$ with nominal size 5%. The poverty line is set to one-half the median. Both distributions are generated from the gamma distribution. The scale parameter of distribution A is set to unity and its shape parameter is set to 2.1557.

Table 4. Descriptive statistics and estimated poverty measures for empirical illustration.

Subgroup	Mean	Median	Std. Dev.	P_1	P_2	P_3	P_1^*	P_2^*	P_3^*
Male	$44,315	$30,020	$53,210	0.1579	0.2106	0.2631	0.1980	0.2585	0.3107
Female	$34,026	$23,000	$38,945	0.1875	0.2601	0.3349	0.1522	0.2175	0.2865

Note: P_j denotes the estimated poverty measure with a poverty line set to 40%$(j = 1)$, 50%$(j = 2)$, and 60%$(j = 3)$ of the median. P_j^* denotes the jth poverty measure based on distinct poverty lines. The overall sample median is $25,800.

For the purposes of this illustration, we ignore the complex sampling structure of the survey and assume the observations are drawn through simple random sampling. The survey also includes observations from the Commonwealth of Puerto Rico which we exclude from our study. Thus, we only consider the 50 states of the United States. Furthermore, to reduce the level of heterogeneity within the sample, we consider only one person households. There are a total of 333,769 observations with 139,284 males and 194,485 females.

The U.S. government has essentially defined poverty in absolute terms for over half a century. As discussed in Section 1, absolute poverty lines are ones which are fixed and thus do not depend on the underlying income distribution. Currently, there are two basic poverty measures the United States uses. The first of which is based on poverty "thresholds" set by the U.S. Census Bureau. These thresholds are adjusted for factors like family size, number of children under 18 years of age, etc., and are most commonly used for statistical purposes. The second way the United States measures poverty is based on poverty "guidelines" issued by U.S. Department of Health and Human Services. These so-called guidelines are mainly used for administrative purposes (i.e., determining financial eligibility for federal assistance programs).[8]

Here we consider relative poverty and set poverty lines to 40%, 50%, and 60% of the median income level, which the Organisation for Economic Co-operation and Development (OECD) use to report poverty rates for its member countries. Some descriptive statistics along with estimated headcount ratios, based on common poverty lines as well as distinct poverty lines, are reported in Table 4. Despite less variability and a lower median income level for females, the conventional estimates based on distinct poverty lines are higher for males on all counts. But those based on common poverty lines suggest higher poverty for females.

Next, we employ our test as well as that of Zheng (2001) in testing some hypotheses based on the headcount ratio. We utilize 999 bootstrap replications and use the Gaussian kernel with bandwidths chosen via "rule-of-thumb" in estimating densities. The nominal size of the test is set to 5%.

Consider Zheng's (2001) test which, in this instance, sets the poverty lines for males and females based on their respective median income levels. We first test the null hypothesis that females have (weakly) higher poverty than males. We obtain a bootstrap p-value of exactly 0 which leads us to reject the null. Next, we test the null hypothesis that males have (weakly) higher poverty than females. In doing so, we obtain a bootstrap p-value of 0.6797 so we are unable to reject our second null hypothesis. The tests here indicate that males may have (weakly) higher poverty than females. If income among females is indeed more evenly distributed as the standard deviations suggest, and they also have a lower median income level, we argue that this conclusion is rather incongruous.

Now we consider our approach which, in this instance, is predicated on the poverty lines being determined by the median of the overall income distribution of males and females. In testing (weakly) higher poverty for females than males, we obtain a bootstrap p-value of

[8] More details can be found on the Institute for Research on Poverty website: http://www.irp.wisc.edu

0.6537 while in testing (weakly) higher poverty for males than females, we obtain exactly 0. Thus, we are able to reject the null of (weakly) higher poverty for males than females, but not the converse. So contrary to Zheng's (2001) test, we reach a conclusion which seems to suggest (weakly) higher poverty for females than males. The likely explanation for the contradictory conclusions is that since the estimated median and standard deviation for females are lower than males, there is more evidence of higher poverty for females based on our framework.

This brief illustration serves as a caution to applied researchers who continue the conventional practice of assuming distinct poverty lines.

6. Conclusion

In this article, we followed up on Thompson's (2013) proposal for the usage of common relative poverty lines, and developed the asymptotic framework for making poverty comparisons between subgroups of a population. While Thompson (2013) was concerned with testing equality of poverty between subgroups, we devised a method for testing inequality restrictions (i.e., whether one subgroup has (weakly) higher poverty than the other).

Simulation evidence presented here validates our approach. A series of Monte Carlo experiments suggest our test has good size and power properties. Finally, an empirical investigation was conducted using some U.S. household income data, for illustrative purposes, which demonstrated why using common relative poverty lines may be more sensible in some cases.

References

Biewen, M. (2002). Bootstrap inference for inequality, mobility and poverty measurement. *J. Economet.* 108:317–342.

Bishop, J.A., Chow, K.V., Zheng, B. (1995). Statistical inference and decomposable poverty measures. *Bull. Econ. Res.* 47:329–340.

Bishop, J.A., Formby, J.P., Zheng, B. (1997). Statistical inference and the Sen index of poverty. *Int. Econ. Rev.* 38:381–387.

Buskirk, T.D. (1998). Nonparametric density estimation using complex survey data. In: *Proceedings of the Survey Research Methods Section*. American Statistical Association.

Davidson, R., Duclos, J.-Y. (2000). Statistical inference for stochastic dominance and the measurement of poverty and inequality. *Econometrica* 68:1435–1464.

Foster, J.E. (1984). On economic poverty: A survey of aggregate measures. In: Basmann, R., Rhodes, G., eds. *Advances in Econometrics*, Volume III. Greenwich, CT: JAI Press.

Kakwani, N. (1993). Statistical inference in the measurement of poverty. *Rev. Econ. Stat.* 75:632–639.

Kodde, D.A., Palm, F.C. (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica* 54:1243–1248.

Li, Q., Racine, J.S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton: Princeton University Press.

McDonald, J. (1984). Some generalized functions for the size distribution of income. *Econometrica* 52:647–663.

Preston, I. (1995). Sampling distributions of relative poverty statistics. *Appl. Stat.* 44:91–99.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica* 44:219–231.

Stengos, T., Thompson, B.S. (2012). Testing for bivariate stochastic dominance using inequality restrictions. *Econ. Lett.* 115:60–62.

Thompson, B.S. (2013). Empirical likelihood-based inference for poverty measures with relative poverty lines. *Economet. Rev.* 32:513–523.

Wolak, F.A. (1989). Testing inequality constraints in linear econometric models. *J. Economet.* 41:205–235.

Zheng, B. (1997). Aggregate poverty measures. *J. Econ. Surv.* 11:123–162.

Zheng, B. (2001). Statistical inference for poverty measures with relative poverty lines. *J. Economet.* 101:337–356.

Paper 2A

Child Indicators Research
https://doi.org/10.1007/s12187-018-9597-5

Stochastic Dominance Approach to Measuring Child Development

Tahsin Mehdi[1]

Accepted: 12 August 2018 / Published online: 24 August 2018

Abstract
We measure global performance when it comes to child development outcomes based on Save the Children UK's Child Development Index, a composite measure comprising health, education, and nutrition. We rely on consistent tests of stochastic dominance efficiency to derive the most optimistic scenario for measured child development where more countries achieve better measured outcomes based on the least variable combination of components of the index. Using the same approach, we also derive the most pessimistic scenario where more countries achieve worse measured outcomes. This approach presents an opportunity to study the sensitivity of the Child Development Index and allows us to better understand the aspects of child development by offering a glimpse into the index components most responsible for driving or hindering improvements in measured child development across countries. To gain a sense of the evolution of the dimensions, we consider four time periods: 1995 - 1999, 2000 - 2004, 2005 - 2010, and 2011 - 2016. We find that in the most optimistic scenario for measured child development, increasing the weight attached to the education dimension over time allows more countries to achieve better measured outcomes. On the other hand, shifting majority of the weight towards health results in the most pessimistic scenario. These results indicate that improvements in children's education outcomes have outpaced health and nutrition. That is, relative to health and nutrition, more countries find it easier to achieve better education outcomes.

Keywords Multidimensional welfare · Stochastic dominance · Child development

1 Introduction

The year 2015 marked the end of United Nation's (UN) Millennium Development Goals (MDGs), and ushered in the era of Sustainable Development Goals (SDGs), an

✉ Tahsin Mehdi
tahsin.mehdi@ryerson.ca

[1] Department of Economics, Ryerson University, 350 Victoria Street, Toronto, Ontario, M5K 2K3, Canada

overarching framework for the world to *"end poverty, protect the planet and ensure prosperity for all"*. Measuring and improving the well-being of children, in particular, has always been at the forefront of public discourse. Achieving universal primary education and reducing child mortality were two of the MDGs (see, e.g., United Nations 2015a). SDGs include better health, education, and nutrition outcomes for all ages (see, United Nations 2015b). According to (World Health Organization 2007), *"economists now argue on the basis of the available evidence that investment in early childhood is the most powerful investment a country can make, with returns over the lifecourse many times the size of the original investment"*.

Neuman and Devercelli (2013) states that worldwide inequalities in child development are stark and they begin prior to birth and expand during a child's early years. There is cross-country evidence suggesting that by the time children enroll in primary school, significant gaps exist in their development. Over 200 million children under the age of 5 living in low and middle-income countries fail to reach their developmental potential as a result of poverty, nutritional deficiencies, and inadequate learning opportunities (see, Grantham-McGregor et al. 2007). The earliest years of a child's life usually presents the best window of opportunity for better outcomes later in life. Child development inequalities exist not just across countries but within countries also as evidenced by figures that show significant disparities in pre-primary enrollment between the poorest and wealthiest households (see, Neuman and Devercelli 2013).

It is now widely recognized that univariate indicators such as income per capita are inadequate for assessing quality of life. This sentiment is echoed by numerous works such as Maasoumi (1999) and List (2004), or Fleurbaey (2012). Recent decades have seen a dramatic increase in the development of multidimensional welfare indicators (e.g., UN's Human Development Index and Multidimensional Poverty Index, or OECD's Better Life Index). When it comes to assessing the well-being of children, it's no different (see, e.g., Roelen and Gassmann 2008; Trani et al. 2013; Chzhen and Ferrone 2017; Chzhen et al. 2017). Main and Bradshaw (2016) and Roelen (2017) point out that household-based monetary metrics are insufficient for capturing what it means for children to be "poor".

Building on the UN's well-known Human Development Index (HDI), Save the Children UK developed the Child Development Index (CDI) in 2008 (see, Save the Children UK 2008), a composite index that aggregates health, education, and nutrition into a summary measure.[1] Due to a lack of reliable annual data, Save the Children UK (2012) combined data from 1995 to 1999, 2000 to 2004, and 2005 to 2010 to produce three sets of CDIs for several countries. In this paper, we consider the same time periods but also combine data for 2011 to 2016 to produce a more recent set of CDIs. The health dimension of the CDI is captured by mortality rate for children under 5 (per 1,000 live births which Save the Children UK bounds to 340), the education dimension is measured by the net primary-school *non*-enrollment, and lastly nutrition is captured by the percentage of underweight children under 5. The CDI is simply the arithmetic mean of the three components, and thus is bounded by

[1] Instead of nutrition, the HDI uses standard of living as its third component.

the unit interval (or between 0 and 100 depending on whether the dimensional indicators are expressed in fractions or not). In other words, each dimension is assigned a weight of 1/3. A relatively lower CDI corresponds to better measured child development outcomes. For this paper, it is easier to use the obverse of the CDI indicators. That is, we consider the *survival* rate of children under 5 (using Save the Children UK's 340 live births as an upper bound), net primary-school *enrollment* rate, and percentage of *healthy weight* children under 5. In this regard, a relatively *higher* CDI will correspond to better measured child development.[2]

The ability of composite indices to synthesize complex social phenomena into single summary measures is indeed appealing but how the underlying components are aggregated often raise questions of efficacy. The CDI, much like the HDI, is an equally-weighted composite index (i.e., each of its dimensions are given equal weight prior to aggregation). Such a weighting scheme is just one among an infinite number of weighting strategies that can be envisioned so it becomes a major source of debate among researchers and policy makers. However, the equal-weighting scheme is not without its virtues as pointed out by Hagerty and Land (2007) who developed a mathematical model for examining weighting schemes for quality of life indicators. One of their many findings was that when the distribution of preferences with regards to weighting are unknown for a given population (which is often the case in practice), constructing an index with equal weights constitutes the "minimax" solution. That is, the equal-weighting strategy actually minimizes the maximum possible disagreement in the population (see Proposition 6, Hagerty and Land 2007).

In this article, our aim is not to discuss the superiority or inferiority of weighting schemes but rather it is to shed light on the CDI components most responsible for driving composite improvements over time and those hindering improvements across countries. For this, we rely on consistent tests of stochastic dominance efficiency (SDE) proposed by Scaillet and Topaloglou (2010), to derive the best-case and worst-case scenarios for measured child development. The former refers to a situation where more countries achieve better measured child development outcomes, while the latter refers to a situation where more countries achieve worse measured outcomes.

SDE is a data-driven aggregation procedure and can be viewed as a direct extension of stochastic dominance which is a nonparametric statistical method of ranking distributions in a robust manner taking all moments of a distribution into account. Stochastic dominance has gained significant interest across multiple disciplines. Stochastic dominance does not require the specification of an explicit social welfare functional form or probability distribution. First-order stochastic dominance satisfies all types of non-decreasing utility functions. It simply requires economic agents to be rational in the sense that more of a social "good" increases utility level while

[2] We refer to "measured" child development to emphasize that the CDI (and any other composite index for that matter) is a subjective measure of well-being. Discussions surrounding a truly optimal indicator of well-being would delve not only into the empirical realm but political and philosophical realms as well. In the absence of such an indicator, composite indicators such as the CDI serve as good analytical tools with which to assess well-being.

more of a social "bad" decreases utility level (monotonicity axiom). Until (Scaillet and Topaloglou 2010), only pairwise testing of dominance was possible (e.g., Davidson and Duclos 2000; Barrett and Donald 2003). Scaillet and Topaloglou (2010) extended stochastic dominance by allowing for full diversification. That is, they devised a method for testing a benchmark set of linearly weighted combination of variables against all other possible linear weighting combinations.

Stochastic dominance exploits all moments of the underlying distributions. Though SDE is rooted in the domain of finance, a strand of literature has recently emerged which apply the method to multidimensional well-being measures (e.g., Agliardi et al. 2012, 2014, 2015; Pinar et al. 2013, 2015, 2017a, b). In the context of multidimensional well-being, an "efficient" weighting vector is one which gives rise to the best-case scenario where more units (e.g., countries) realize better measured outcomes based on the least variable combination of components. The weights arising from this type of analysis identify the components that are the greatest contributors to improvements or hindrances in overall measured outcomes over time, which may be of great interest to policy makers and researchers.

SDE has been used to assess the HDI in Pinar et al. (2013, 2015, 2017a). The approach was also utilized by Pinar (2015) in examining the World Bank's World Governance Indicators.[3] While Pinar et al. (2013, 2017a) examined human development worldwide, Pinar et al. (2015) restricted the analysis to the Middle East and North Africa (MENA) region. Agliardi et al. (2015) used the approach in developing an environmental degradation index. Agliardi et al. (2012, 2014) and Pinar et al. (2017b) are examples of how SDE can be applied in the domain of finance.

Following such recent works, we rely on consistent tests of SDE to derive the most optimistic and most pessimistic scenarios for measured child development across countries. Our results suggest that more countries realize higher levels of measured child development when education receives more weight over time. On the other hand, shifting most of the weight towards health results in the worst-case scenario where more countries achieve worse measured outcomes. This reveals that more countries may have invested in their education system relative to health and nutrition over time, and thus weighting education more heavily allows more countries to achieve better measured child development. This also means that improvements in education systems may have outpaced health and nutrition. In terms of cross-country comparisons, re-weighting the indicators has little impact on countries that are well invested in all three dimensions while those invested in one or two dimensions experience substantial deviations in rankings.

The remainder of this article is organized as follows. We commence with Section 2 which introduces the concept of SDE in the context of measuring child development outcomes. In Section 3, we describe the data. We employ SDE and offer a discussion of our empirical results in Section 4. Section 5 concludes with reflections on the broad findings and implications.

[3] Pinar et al. (2013) found that weighting education relatively more results in the most optimistic scenario for measured human development. This study was later followed up by Pinar et al. (2017a) who found that under the updated methodology of the UN, the health dimension gets weighted increasingly more over time in the most optimistic scenario.

2 Stochastic Dominance Efficient Child Development Index

Consider $\mathbf{Y}$, an $N \times 3$ matrix of achievements taking values in $\mathbb{R}^3$, where the elements in the matrix consist of realization of achievements in the 3 CDI indicators for N countries. Let $F(\mathbf{Y})$ be the continuous cumulative distribution function (CDF) of $\mathbf{Y} = (Y_1, Y_2, Y_3)'$ at point $\mathbf{y} = (y_1, y_2, y_3)'$.

Consider some weighting vector $\lambda \in \mathbb{L}$ where $\mathbb{L} := \{\lambda \in \mathbb{R}^3_+ \mid e'\lambda = 1\}$, and e is a vector of ones of length 3 that imply the (non-negative) weights sum to unity. We denote the CDF of the composite index $\lambda'\mathbf{Y}$ at some measured child development level z as

$$G(z, \lambda; F) = \int_{\mathbb{R}^3} I(\lambda'\mathbf{u} \leq z) dF(\mathbf{u}), \tag{1}$$

where $I(\cdot)$ is the indicator function that equals 1 if its argument $(\cdot)$ is true, and 0 otherwise.

Denote the benchmark weighting vector by $\tau = (1/3, 1/3, 1/3)$, the vector of equal weights which Save the Children UK used to report its composite index of child development. With relatively little guidance regarding specifications of preferences and distributional assumptions regarding the components of child development, stochastic dominance offers a generalized approach that allows us to relax several assumptions typically synonymous with economic theory.

Distribution of the composite index constructed from λ stochastically dominates the one constructed from τ at first-order if $G(z, \lambda; F) \leq G(z, \tau; F)$ for all $z \in \mathbb{R}$. In other words, the proportion of countries under weighting vector, λ, with measured child development below some given level z is less than (or at most equal to) the proportion of countries under weighting vector, τ, as shown in Fig. 1.

In this article, our focus is strictly on first-order stochastic dominance which is the strongest form, but in general, let

$$\mathcal{J}_j(z, \lambda; F) = \frac{1}{(j-1)!} \int_{\mathbb{R}^3} (z - \lambda'\mathbf{u})^{j-1} I(\lambda'\mathbf{u} \leq z) dF(\mathbf{u}). \tag{2}$$

Distribution of the composite index constructed from λ dominates that of the composite index constructed from τ at order j if $\mathcal{J}_j(z, \lambda; F) \leq \mathcal{J}_j(z, \tau; F)$ for all $z \in \mathbb{R}$. It is straightforward to check that for $j = 1$, $\mathcal{J}_1(z, \cdot; F) = G(z, \cdot; F)$. Also note that higher order dominance is implied by lower order dominance.

Testing for SDE of τ against all other linear weighting combinations at order j can be formally stated as,

$$\begin{aligned} H_0^j &: \mathcal{J}_j(z, \tau; F) \leq \mathcal{J}_j(z, \lambda; F) \text{ for all } z \in \mathbb{R} \text{ and for all } \lambda \in \mathbb{L}, \\ H_1^j &: \mathcal{J}_j(z, \tau; F) > \mathcal{J}_j(z, \lambda; F) \text{ for some } z \in \mathbb{R} \text{ or for some } \lambda \in \mathbb{L}. \end{aligned}$$

In practice, F will be unknown so expression Eq. 2 will need to be replaced by its empirical counterpart,

$$\mathcal{J}_j(z, \lambda; \hat{F}) = \frac{1}{N(j-1)!} \sum_{i=1}^{N} (z - \lambda'\mathbf{Y}_i)^{j-1} I(\lambda'\mathbf{Y}_i \leq z). \tag{3}$$

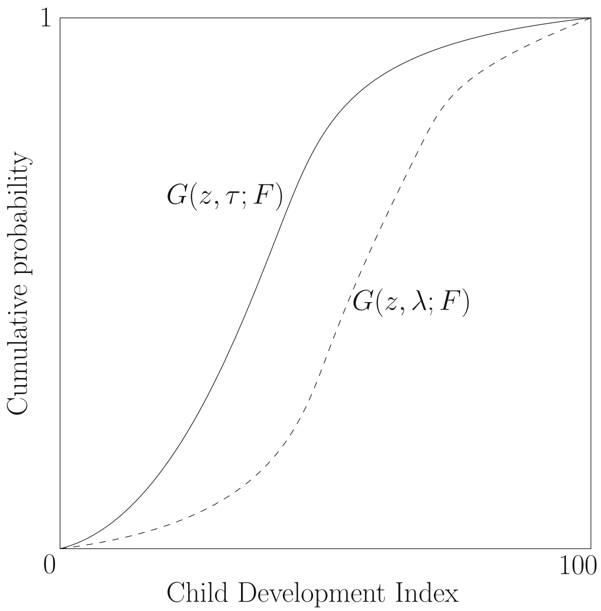

Fig. 1 First order stochastic dominance of λ weighted index over τ weighted index

Under the null hypothesis, τ is efficient at order j, in that no composite index can be constructed from λ that dominates (at order j) the index based on τ.

To test for first-order ($j = 1$) SDE of τ which is tantamount to testing whether τ offers the best-case scenario for measured child development, we use the one-sided weighted Kolmogorov-Smirnov type test statistic proposed by Scaillet and Topaloglou (2010),

$$\hat{S} = \sqrt{N}\frac{1}{N} \sup_{z,\lambda} [G(z, \tau; \hat{F}) - G(z, \lambda; \hat{F})]. \tag{4}$$

Naturally, one may be interested in whether τ actually gives rise to the *worst-case* scenario for measured child development. The test statistic in that case simply involves reversing (4).

The test statistic has a non-standard limiting distribution and thus, p-values need to be obtained using bootstrap methods. We use the block bootstrap procedure which is shown to be quite reliable even in small samples (see, e.g., Pinar et al. 2013; Scaillet and Topaloglou 2010). The computation of the test statistic which involves mixed integer linear programming, is enumerated in detail in Pinar et al. (2013) and alternatively, Scaillet and Topaloglou (2010). We use R's Rcplex package (Bravo and Theussl 2016; R Core Team 2018) which is an implementation of IBM's CPLEX optimizer.

If the null is rejected in the most optimistic scenario, then the benchmark weighting vector, τ, does not generate the most optimistic scenario for measured child development across countries. There exists some other weighting vector, λ, that leads to more countries achieving better measured outcomes. Such weights can be obtained by solving,

$$\max_{\lambda} \left[G(z, \tau; \hat{F}) - G(z, \lambda; \hat{F}) \right] \text{ for a given } z. \tag{5}$$

The above problem is solved for each observed Child Development Index level, z, under the equal-weighting scheme. The weights arising from Eq. 5 give us an indication as to the implicit preferences of countries. That is, they tell us which components are the greatest contributors to improvements in measured child development.

Analogously, the weights for the worst-case scenario can be obtained by reversing the maximization problem (5). Such weights provide insight on the indicators *preventing* improvement in measured outcomes.

3 Data

Our data spans from 1995 to 2016 and is primarily drawn from the UN's MDG indicators. There exists a fairly consistent series of data for mortality rates for children under 5, but data on net primary-school non-enrollment and underweight children under 5 is quite sparse. For this reason, we follow Save the Children UK and combine our data with World Bank data to fill in data gaps where possible. We take the average of each indicator across the same time frames as Save the Children UK: 1995 - 1999, 2000 - 2004, 2005 - 2010, and additionally 2011 - 2016. This results in a sample of 80, 76, 95, and 70 countries, respectively.[4]

As mentioned in Section 1, to place this in the SDE framework, we subtract the dimensional rates from 100% to transform the indicators into *survival* rates for children 5 (per 1,000 live births which we bound to 340 following Save the Children UK), net primary-school *enrollment* rates, and percentage of *healthy weight* children under 5. This ensures that a *higher* CDI corresponds with *better* measured child development outcomes.

4 Results

We now present our findings and offer a discussion. The means and standard deviations of the CDI components are reported in Table 1. Generally, it is evident that the means for education and nutrition are higher than the health dimension for all time periods. The null hypothesis of first-order dominance of the equally weighted CDI (i.e., that equal-weighting offers the most optimistic scenario), is rejected. It is also rejected for the worst-case scenario (i.e., the equally weighted CDI does not give rise to the most pessimistic scenario either). There exist several other weighting schemes

[4]Save the Children UK 2008 and 2012 were able to publish CDIs for 141 countries consistently across three time periods due to the fact that they were able to supplement their data with additional sources at the time. However, well-being indicators such as mortality, school enrollment, and malnutrition often get revised so we draw from the most recent database.

Table 1 Means and standard deviations of the Child Development Index components

	Mean			Standard deviation		
	Health[1]	Education[2]	Nutrition[3]	Health[1]	Education[2]	Nutrition[3]
1995 - 1999	76.0	78.6	84.0	19.3	19.9	12.2
2000 - 2004	77.5	81.6	82.9	15.8	18.0	12.3
2005 - 2010	83.2	86.6	86.8	13.1	14.0	10.9
2011 - 2016	84.7	88.4	86.5	10.6	11.9	10.2

[1]Survival rate for children under 5

[2]Net primary-school enrollment rate

[3]Percentage of healthy weight children under 5

(indices) that dominate equal-weighting. Our tests were based on 999 block bootstrap replications.

The average of the dominating weights for the most optimistic (best-case) and most pessimistic (worst-case) scenarios are presented in Table 2 while the country rankings are provided in Tables 3, 4, 5, and 6. It is immediately clear that the best-case scenario where more countries realize better measured child development, occurs when education receives more weight over time. That is, education is the main driving force towards higher measured child development as its weight more than doubles from 0.361 in 1995 - 1999 to 0.739 in 2011 - 2016 under the most optimistic

Table 2 Mean stochastic dominance efficient weights for best-case and worst-case scenarios for the Child Development Index

	N	Dominating indices	Health[1]	Education[2]	Nutrition[3]
Best-case scenario					
1995 - 1999	80	78	0.047	0.361	0.592
2000 - 2004	76	75	0.070	0.542	0.388
2005 - 2010	95	92	0.039	0.680	0.281
2011 - 2016	70	66	0.074	0.739	0.187
Worst-case scenario					
1995 - 1999	80	72	0.681	0.222	0.097
2000 - 2004	76	73	0.795	0.086	0.119
2005 - 2010	95	95	0.692	0.151	0.157
2011 - 2016	70	70	0.633	0.136	0.231

[1]Measured by survival rate for children under 5

[2]Measured by net primary-school enrollment rate

[3]Measured by percentage of healthy weight children under 5

Table 3 Child Development Index rankings under different weighting schemes, 1995 - 1999

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Kuwait	1	1	2	0	–1	98.51	97.78	96.86
Australia	2	2	1	0	1	97.91	97.50	97.44
Argentina	3	3	7	0	–4	96.84	96.20	95.08
Tunisia	4	10	21	6	–11	96.43	94.29	91.67
Jordan	5	8	14	3	–6	96.43	94.94	93.04
Qatar	6	5	4	–1	1	95.95	96.09	96.10
Republic of Macedonia	7	9	10	2	–1	95.75	94.75	94.02
Croatia	8	7	6	–1	1	95.70	95.34	95.71
Brazil	9	15	24	6	–9	95.66	93.44	90.65
United States	10	4	3	–6	1	95.62	96.20	96.84
Bahrain	11	6	5	–5	1	95.10	95.78	96.04
Jamaica	12	14	15	2	–1	94.75	93.80	92.92
Lebanon	13	12	13	–1	–1	94.72	93.87	93.20
Costa Rica	14	11	9	–3	2	94.01	94.03	94.52
Uruguay	15	13	11	–2	2	93.87	93.81	93.95
Romania	16	18	23	2	–5	93.82	92.43	91.30
Peru	17	26	31	9	–5	93.50	90.92	87.88
Mexico	18	16	18	–2	–2	93.46	93.06	92.20
Bolivia	19	38	48	19	–10	93.31	87.31	79.99
West Bank & Gaza	20	23	25	3	–2	93.21	91.65	90.40
Panama	21	17	17	–4	0	93.15	92.68	92.22
Albania	22	20	19	–2	1	92.89	92.36	91.71
South Africa	23	32	41	9	–9	92.25	89.19	85.13
Colombia	24	22	22	–2	0	92.20	91.76	91.53
Venezuela	25	24	20	–1	4	92.15	91.62	91.71
Georgia	26	28	30	2	–2	92.13	89.79	87.98
Turkey	27	29	33	2	–4	91.70	89.77	87.43
Syrian Arab Republic	28	19	16	–9	3	91.64	92.40	92.72
Egypt	29	31	38	2	–7	91.28	89.32	86.59
Guyana	30	30	34	0	–4	91.26	89.70	87.42
Ecuador	31	27	27	–4	0	90.64	90.64	89.97
Dominican Republic	32	36	37	4	–1	90.45	88.29	86.62
Mauritius	33	21	12	–12	9	90.43	92.08	93.39
Kyrgyzstan	34	35	39	1	–4	90.34	88.49	86.03
Azerbaijan	35	40	49	5	–9	89.83	85.02	79.31

Table 3 (continued)

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
China	36	34	32	–2	2	89.56	88.52	87.59
Iran	37	33	28	–4	5	89.02	88.85	88.50
Malaysia	38	25	8	–13	17	87.93	91.48	94.81
Mongolia	39	42	46	3	–4	87.51	84.20	80.43
Algeria	40	39	35	–1	4	87.46	87.31	87.38
Zimbabwe	41	47	51	6	–4	85.78	80.81	75.21
Oman	42	37	26	–5	11	85.77	87.49	90.27
El Salvador	43	41	40	–2	1	84.93	84.77	85.74
Nicaragua	44	45	43	1	2	84.37	83.71	84.03
Morocco	45	49	47	4	2	82.45	80.46	80.08
Botswana	46	48	50	2	–2	82.26	80.60	79.11
Indonesia	47	46	42	–1	4	81.54	83.42	84.08
Malawi	48	54	68	6	–14	81.29	71.27	57.14
Philippines	49	44	36	–5	8	79.84	83.74	86.87
Guatemala	50	50	45	0	5	79.20	80.34	81.72
Vietnam	51	43	29	–8	14	78.73	84.19	88.39
Togo	52	52	58	0	–6	78.50	73.30	67.10
Rwanda	53	63	72	10	–9	77.17	65.14	50.20
Zambia	54	61	70	7	–9	76.28	66.77	55.78
Kenya	55	56	57	1	–1	74.70	70.75	67.50
Comoros	56	53	52	–3	1	74.55	72.06	69.90
Bhutan	57	55	53	–2	2	74.31	71.23	69.88
Maldives	58	51	44	–7	7	74.29	79.77	83.50
Mauritania	59	58	59	–1	–1	73.31	69.23	66.30
Ghana	60	57	56	–3	1	72.79	69.85	67.67
Gambia	61	60	61	–1	–1	72.18	67.94	63.65
Cote d'Ivoire	62	64	66	2	–2	71.73	64.95	58.65
Senegal	63	65	65	2	0	69.65	63.90	59.02
Nigeria	64	71	73	7	–2	67.67	58.77	48.75
Cambodia	65	59	55	–6	4	67.56	68.76	68.08
Benin	66	66	67	0	–1	67.50	62.50	57.62
Tanzania	67	70	69	3	1	64.02	59.48	55.93
Mozambique	68	73	74	5	–1	63.66	55.15	46.95
Guinea	69	74	75	5	–1	63.43	54.40	46.43
Madagascar	70	67	62	–3	5	63.16	62.30	61.79

Table 3 (continued)

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Djibouti	71	69	63	–2	6	62.49	59.69	60.80
Nepal	72	62	54	–10	8	62.32	65.95	69.41
Congo Dem. Republic	73	72	71	–1	1	61.89	56.35	51.01
Yemen	74	68	60	–6	8	59.01	61.84	65.39
Angola	75	77	78	2	–1	58.29	50.42	41.63
Chad	76	76	76	0	0	56.98	50.68	44.69
Mali	77	79	79	2	0	55.39	46.85	37.67
Burkina Faso	78	78	77	0	1	53.55	47.73	43.19
Eritrea	79	75	64	–4	11	50.16	53.45	59.75
Niger	80	80	80	0	0	41.98	34.56	27.70
Mean	–	–	–	–	–	81.66	79.52	77.34

view of measured child development. This indicates that more countries find it easier to realize better measured child development if more weight is shifted towards education over time, a finding consistent with Pinar et al. (2013) who also found education to be the dominant dimension for the HDI. The dominance of education is underscored by Table 1 which shows that the mean net primary-school enrollment rate increased by almost 10 percentage points from 1995 to 2016, while its standard deviation decreased by almost 10 percentage points in that same period. As seen in Tables 3–6, the best-case scenario always results in a higher mean composite index. This is to be expected as one of the properties of first-order stochastic dominance is that the dominating distribution is guaranteed to have the higher mean (i.e., since more countries achieve a better measured outcome, it stands to reason that the overall mean will also rise).

On the other hand, the worst-case scenario occurs when most of the weight gets shifted towards health regardless of the time period. Under the worst-case scenario, the weight attached to the nutrition dimension increases gradually over time. This echoes our finding that improvements in survival rate outcomes for children under 5 and percentage of healthy weight children under 5 may be lagging behind compared to net primary-school enrollment rate outcomes. In other words, more countries find it harder to achieve better measured child development outcomes when nutrition gets weighted heavily.

The country rankings for the 1995 - 1999 period are presented in Table 3. For this period, the top 2 countries (Kuwait, and Australia) do not experience any major deviations in their ranking regardless of the weighting schemes. This suggests that such countries have balanced achievements across the child development indicators and

Table 4 Child Development Index rankings under different weighting schemes, 2000 - 2004

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Kuwait	1	1	3	0	–2	98.56	97.84	96.80
Jordan	2	4	11	2	–7	97.49	95.88	93.35
Peru	3	8	22	5	–14	97.19	94.84	91.30
Tunisia	4	7	14	3	–7	96.80	95.35	92.96
Cuba	5	2	1	–3	1	96.47	96.85	97.47
Ecuador	6	12	19	6	–7	96.45	94.51	91.74
Lebanon	7	5	5	–2	0	96.35	95.85	95.10
Turkey	8	17	23	9	–6	96.09	94.09	90.62
United States	9	3	2	–6	1	95.96	96.50	97.27
Trinidad & Tobago	10	10	15	0	–5	95.88	94.65	92.61
Brazil	11	14	16	3	–2	95.45	94.41	92.48
Panama	12	11	13	–1	–2	95.37	94.56	93.28
Romania	13	9	12	–4	–3	95.37	94.69	93.34
Syrian Arab Republic	14	15	7	1	8	95.08	94.34	93.79
Republic of Macedonia	15	6	4	–9	2	95.06	95.58	95.74
Colombia	16	13	10	–3	3	94.97	94.42	93.47
Egypt	17	20	25	3	–5	94.47	92.47	89.70
Jamaica	18	18	8	0	10	94.00	94.05	93.62
Venezuela	19	16	6	–3	10	93.95	94.15	94.14
Iran	20	19	20	–1	–1	93.72	93.05	91.60
Sao Tome & Principe	21	33	41	12	–8	93.10	87.84	79.52
Guyana	22	24	33	2	–9	93.06	90.61	87.42
Bolivia	23	31	38	8	–7	92.58	88.98	82.56
Algeria	24	22	27	–2	–5	91.97	91.07	89.59
Mongolia	25	32	37	7	–5	91.20	88.96	85.02
Armenia	26	21	17	–5	4	91.12	91.96	92.15
South Africa	27	35	40	8	–5	90.96	86.65	79.80
Iraq	28	30	31	2	–1	90.46	89.48	87.96
Nicaragua	29	26	24	–3	2	90.39	90.39	89.70
Honduras	30	28	26	–2	2	90.31	89.86	89.67
Sri Lanka	31	25	9	–6	16	90.01	90.41	93.62
Dominican Republic	32	27	28	–5	–1	89.83	90.07	89.23
El Salvador	33	23	21	–10	2	89.82	90.71	91.42
Albania	34	29	18	–5	11	88.99	89.60	91.82
Maldives	35	34	30	–1	4	88.48	87.55	88.76

Table 4 (continued)

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Indonesia	36	36	36	0	0	88.35	86.61	85.86
Malawi	37	45	64	8	–19	88.25	77.93	61.98
Azerbaijan	38	41	39	3	2	86.82	85.46	82.51
Vietnam	39	37	29	–2	8	86.54	86.58	88.85
Philippines	40	38	32	–2	6	85.93	86.13	87.77
Guatemala	41	40	35	–1	5	85.83	85.53	85.89
Morocco	42	39	34	–3	5	85.29	86.03	86.33
Namibia	43	43	42	0	1	85.12	82.39	78.81
Botswana	44	42	44	–2	–2	84.71	82.57	78.04
Rwanda	45	50	65	5	–15	82.47	74.40	61.19
Myanmar	46	44	43	–2	1	81.44	79.42	78.20
Lesotho	47	46	55	–1	–9	80.70	76.47	68.59
Benin	48	51	63	3	–12	80.37	74.29	64.30
Cambodia	49	47	46	–2	1	80.04	76.30	73.79
Swaziland	50	48	60	–2	–12	79.35	75.59	66.89
Gambia	51	49	52	–2	–3	77.58	75.32	70.09
Equatorial Guinea	52	55	62	3	–7	77.36	73.31	64.55
Comoros	53	52	50	–1	2	75.44	73.95	71.51
Tanzania	54	54	54	0	0	75.04	73.32	68.75
Zambia	55	58	66	3	–8	74.78	69.30	59.59
Kenya	56	53	51	–3	2	74.39	73.57	70.47
Nepal	57	56	45	–1	11	72.05	72.54	76.34
Lao	58	60	57	2	3	70.51	69.11	68.10
Ghana	59	57	47	–2	10	70.31	71.82	72.23
Senegal	60	59	61	–1	–2	69.90	69.12	65.85
Nigeria	61	66	72	5	-6	67.82	62.57	52.66
Madagascar	62	61	53	–1	8	67.02	67.80	69.88
Mozambique	63	65	69	2	–4	66.88	64.43	58.10
Mauritania	64	63	59	–1	4	66.76	67.08	67.06
Yemen	65	62	49	–3	13	65.92	67.52	72.01
Pakistan	66	64	58	–2	6	63.51	65.38	67.54
Guinea	67	68	70	1	–2	60.89	59.86	56.12
Guinea-Bissau	68	69	71	1	–2	60.56	59.03	52.96
Mali	69	73	75	4	–2	60.29	55.50	45.46
Chad	70	72	74	2	–2	59.37	56.18	48.98

Table 4 (continued)

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Burundi	71	71	68	0	3	54.55	56.62	58.85
Eritrea	72	67	48	–5	19	54.13	61.60	72.06
Ethiopia	73	74	67	1	7	51.69	55.02	59.45
Djibouti	74	70	56	–4	14	49.94	58.61	68.48
Burkina Faso	75	75	73	0	2	49.33	50.36	49.06
Niger	76	76	76	0	0	43.84	43.71	41.00
Mean	–	–	–	–	–	81.82	80.67	78.48

re-weighting the composite index has little to no effect on their measured outcomes. On the other hand, Bolivia improved its ranking by 19 positions under the best-case scenario going from 38th to 19th (out of 80 countries). In the 1995 - 1999 period, the SDE method assigned most of the weight towards nutrition which allowed Bolivia to move up. This is because Bolivia's score in the nutrition dimension is relatively high. However, under the worst-case scenario, Bolivia ranked 48th, dropping 10 spots from its original equal-weight ranking. Its composite index was 93.31 under the best-case scenario while it decreased to 79.99 under the worst-case scenario, a fairly significant composite index variation. This is indicative that countries such as Bolivia may excel is one dimension but may be lagging in others.

Table 4 reports the rankings for the 2000 - 2004 period when education became the dominant dimension (receiving more than half the weight under the best-case scenario). This suggests that more countries started finding it easier to improve childrens' education outcomes as opposed to outcomes in health or nutrition, thus showing an implicit preference for improving education systems perhaps by investing relatively more in education. Under the best-case scenario, Sao Tome & Principe (+12) and El Salvador (-10) experience the largest rank deviations from their benchmark (equal-weight) as a result of education being weighted heavily. Countries that improved their ranking were generally the ones that excelled in net primary-school enrollment rates while those that dropped in rankings tended to be the ones that were at a relative disadvantage when it came to education. Under the worst-case scenario, Malawi (-19) and Eritrea (+19) exhibited the largest deviations from their benchmark ranking. Recall that the vast majority of the weight is shifted to health in the worst-case scenario (see Table 2). Although relatively more countries (compared to the benchmark case) achieve worse measured outcomes in this scenario, those that did achieve good survival rate outcomes for children under 5, moved up. Once again, SDE demonstrates how significantly composite indices can vary by re-weighting. Malawi achieved 88.25 under the most optimistic scenario but just 61.98 under the

Table 5 Child Development Index rankings under different weighting schemes, 2005 - 2010

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Germany	1	1	1	0	0	99.31	99.04	98.85
Japan	2	2	2	0	0	99.01	98.53	98.76
Argentina	3	5	10	2	–5	98.56	97.39	96.30
Kuwait	4	4	6	0	–2	98.36	97.69	97.14
Jordan	5	9	16	4	–7	98.36	96.78	95.01
Tunisia	6	10	13	4	–3	98.17	96.61	95.25
Belize	7	12	14	5	–2	97.99	96.22	95.09
Uruguay	8	8	8	0	0	97.85	96.80	96.42
Turkey	9	13	27	4	–14	97.75	96.09	93.95
Kazakhstan	10	19	30	9	–11	97.39	95.38	93.58
Peru	11	17	24	6	–7	97.37	95.64	94.28
Australia	12	3	3	–9	0	97.09	98.05	98.27
Egypt	13	24	33	11	–9	96.44	94.61	93.30
Mongolia	14	32	43	18	–11	96.35	93.49	90.58
Syrian Arab Republic	15	23	20	8	3	96.35	94.68	94.84
United States	16	7	4	–9	3	96.00	96.93	97.37
Georgia	17	14	19	–3	–5	95.99	95.83	94.89
Chile	18	6	5	–12	1	95.98	97.10	97.26
Brazil	19	16	18	–3	–2	95.68	95.65	94.93
Algeria	20	26	35	6	–9	95.64	94.34	92.60
Mexico	21	18	15	–3	3	95.54	95.47	95.04
Panama	22	20	22	–2	–2	95.51	95.07	94.37
Republic of Macedonia	23	11	9	–12	2	95.38	96.27	96.39
Kyrgyzstan	24	31	41	7	–10	95.18	93.66	91.50
Sao Tome & Principe	25	47	54	22	–7	95.07	89.56	85.17
Oman	26	22	12	–4	10	95.07	94.78	95.66
Venezuela	27	21	17	–6	4	94.33	94.95	94.98
Colombia	28	25	23	–3	2	94.27	94.50	94.29
Malaysia	29	30	11	1	19	94.24	93.89	95.85
Thailand	30	27	21	–3	6	94.16	94.29	94.82
Tajikistan	31	48	52	17	–4	94.08	89.22	85.76
El Salvador	32	29	26	–3	3	94.02	93.90	93.96
Guatemala	33	43	45	10	–2	94.01	91.03	89.85
Honduras	34	37	37	3	0	93.80	92.72	92.25
Nicaragua	35	34	38	–1	–4	93.38	92.98	92.21

Table 5 (continued)

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Belarus	36	15	7	–21	8	93.26	95.74	96.84
South Africa	37	50	57	13	–7	93.09	88.51	83.73
Bolivia	38	44	49	6	–5	92.69	90.90	87.86
Malawi	39	59	67	20	–8	92.62	84.39	76.85
Moldova	40	28	25	–12	3	92.60	93.93	94.11
Maldives	41	41	34	0	7	92.24	91.11	93.11
Iraq	42	42	46	0	–4	91.98	91.09	89.75
Indonesia	43	51	47	8	4	91.80	88.46	88.73
Armenia	44	33	31	–11	2	91.74	93.11	93.56
Paraguay	45	36	36	–9	0	91.53	92.75	92.51
Albania	46	35	28	–11	7	91.33	92.79	93.77
Vietnam	47	49	44	2	5	90.93	88.86	90.39
Saudi Arabia	48	38	29	–10	9	90.68	92.70	93.73
Tanzania	49	55	61	6	–6	90.59	85.32	81.34
Zambia	50	64	74	14	–10	90.20	81.84	73.71
Sri Lanka	51	46	32	–5	14	90.18	89.89	93.37
Cambodia	52	58	55	6	3	90.14	84.70	84.56
Suriname	53	39	40	–14	–1	90.11	91.34	91.71
Dominican Republic	54	40	42	–14	–2	89.93	91.23	90.74
Togo	55	65	70	10	–5	87.95	80.99	75.37
Uganda	56	63	66	7	–3	87.40	82.07	77.08
Rwanda	57	61	63	4	–2	87.13	83.18	79.20
West Bank & Gaza	58	45	39	–13	6	86.93	90.75	91.88
India	59	71	62	12	9	86.50	78.65	79.37
Azerbaijan	60	52	51	–8	1	86.41	87.56	87.18
Guyana	61	53	50	–8	3	85.85	87.02	87.34
Philippines	62	54	48	–8	6	85.84	85.87	88.08
Namibia	63	60	58	–3	2	85.72	83.69	82.38
Bangladesh	64	67	60	3	7	85.24	79.83	81.45
Benin	65	69	75	4	–6	85.22	78.99	73.47
Myanmar	66	62	59	–4	3	85.20	82.67	82.27
Botswana	67	57	56	–10	1	84.96	85.10	84.00
Swaziland	68	66	76	–2	–10	84.10	80.37	72.84
Bhutan	69	56	53	–13	3	83.91	85.21	85.25
Angola	70	82	95	12	–13	83.78	71.06	55.91

Table 5 (continued)

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Mozambique	71	74	80	3	–6	82.88	77.19	70.89
Cameroon	72	72	78	0	–6	82.80	77.69	71.48
Kenya	73	68	68	–5	0	80.81	79.00	76.29
Lao	74	78	73	4	5	79.82	75.78	74.82
Lesotho	75	77	82	2	–5	79.28	76.35	70.50
Senegal	76	70	64	–6	6	77.87	78.81	77.39
Gambia	77	75	71	–2	4	76.90	77.14	75.27
Burundi	78	80	81	2	–1	76.80	72.57	70.57
Ghana	79	73	69	–6	4	76.18	77.65	76.16
Mauritania	80	79	79	–1	0	74.22	73.89	71.44
Guinea-Bissau	81	85	89	4	–4	73.58	70.12	63.06
Congo	82	76	65	–6	11	72.36	77.07	77.35
Timor-Leste	83	83	72	0	11	71.30	70.26	75.09
Guinea	84	84	87	0	–3	71.26	70.21	66.62
Mali	85	88	92	3	–4	69.28	66.20	60.16
Nigeria	86	89	90	3	–1	68.75	66.18	61.49
Equatorial Guinea	87	81	85	–6	–4	67.37	72.33	69.21
Cote d'Ivoire	88	86	88	–2	–2	66.24	67.93	66.35
Ethiopia	89	87	83	–2	4	64.54	67.41	70.48
Central African Republic	90	92	93	2	–1	64.34	62.54	57.15
Burkina Faso	91	94	91	3	3	60.09	62.01	60.96
Djibouti	92	91	84	–1	7	53.38	63.78	70.36
Liberia	93	90	86	–3	4	52.63	64.30	68.16
Niger	94	95	94	1	1	52.04	55.09	55.99
Eritrea	95	93	77	–2	16	50.62	62.45	72.83
Mean	–	–	–	–	–	86.51	85.52	84.26

most pessimistic scenario. This suggests that for countries such as Malawi, there exists quite a lot of disparity between their achievements in the different dimensions (they may excel in some but not others).

The rankings for the 2005 - 2010 period are provided in Table 5. Because of their balanced achievements across health, education, and nutrition outcomes for children, Germany and Japan consistently rank as the first and second best, respectively, regardless of the weighting scenario. Sao Tome & Principe (+22), Malawai (+20) and Belarus (-21) exhibited the largest deviations in rankings under the most optimistic

Table 6 Child Development Index rankings under different weighting schemes, 2011 - 2016

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Tunisia	1	1	5	0	–4	99.10	97.63	96.50
Iran	2	3	8	1	–5	98.46	96.74	95.72
Algeria	3	7	17	4	–10	98.21	96.16	94.37
Belize	4	10	10	6	0	98.12	96.09	95.30
Mexico	5	2	7	–3	–5	97.44	96.80	96.20
Kyrgyzstan	6	13	19	7	–6	97.29	95.67	94.10
Morocco	7	15	23	8	–8	97.23	95.20	93.14
Jordan	8	6	9	–2	–3	97.06	96.22	95.35
Mongolia	9	14	24	5	–10	97.01	95.35	93.13
Uruguay	10	5	6	–5	–1	96.48	96.27	96.36
Ecuador	11	17	20	6	–3	96.48	94.77	93.87
Egypt	12	20	21	8	–1	96.21	94.58	93.86
Vietnam	13	24	26	11	–2	96.16	93.12	92.47
Oman	14	18	11	4	7	95.92	94.73	95.25
Turkey	15	11	13	–4	–2	95.87	95.86	95.14
Tajikistan	16	30	35	14	–5	94.61	89.91	87.41
Peru	17	16	14	–1	2	94.49	95.01	95.03
Chile	18	4	1	–14	3	94.25	96.54	97.31
El Salvador	19	19	15	0	4	94.23	94.71	94.98
United States	20	9	3	–11	6	94.03	96.12	97.17
Kuwait	21	12	4	–9	8	93.67	95.73	96.60
Republic of Macedonia	22	8	2	–14	6	93.61	96.13	97.19
Honduras	23	23	22	0	1	93.59	93.31	93.23
Sao Tome & Principe	24	29	34	5	–5	93.39	90.17	87.44
Cambodia	25	35	36	10	–1	93.39	87.49	86.79
Togo	26	43	49	17	–6	93.21	85.19	79.63
Indonesia	27	32	31	5	1	92.70	89.06	89.19
India	28	44	41	16	3	92.38	84.38	82.79
Moldova	29	21	12	–8	9	92.27	94.55	95.22
Nepal	30	42	39	12	3	91.99	85.49	85.27
Sierra Leone	31	61	69	30	–8	91.90	77.25	64.80
Thailand	32	25	18	–7	7	91.75	92.79	94.21
Zimbabwe	33	41	48	8	–7	91.72	85.51	79.82
West Bank & Gaza	34	22	16	–12	6	91.62	94.03	94.37
Dominican Republic	35	26	27	–9	–1	91.23	92.49	92.31
Cameroon	36	47	57	11	–10	90.98	83.57	77.60

Table 6 (continued)

	Rank			Deviation from equal-weight		Index		
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case
Zambia	37	45	53	8	–8	90.86	84.16	78.82
Guatemala	38	31	30	–7	1	90.71	89.84	89.97
Benin	39	48	55	9	–7	90.63	83.47	78.59
Lao	40	51	51	11	0	90.60	82.63	79.58
Sri Lanka	41	33	28	–8	5	90.52	88.29	91.28
Congo	42	34	37	–8	–3	90.45	88.14	86.40
Azerbaijan	43	27	29	–16	–2	89.66	90.97	90.63
Uganda	44	39	42	–5	–3	89.41	85.57	82.59
Philippines	45	36	32	–9	4	88.86	87.28	88.43
Namibia	46	37	38	–9	–1	87.92	86.77	85.89
Tanzania	47	40	40	–7	0	85.98	85.53	84.87
Paraguay	48	28	25	–20	3	85.64	90.88	92.71
Timor-Leste	49	56	50	7	6	85.59	79.10	79.60
Mozambique	50	53	58	3	–5	85.35	81.42	77.54
Kenya	51	46	44	–5	2	85.19	84.12	81.77
Ghana	52	50	47	–2	3	85.16	82.91	80.04
Yemen	53	57	46	4	11	83.15	78.82	80.15
Comoros	54	54	52	0	2	82.82	80.96	78.89
Swaziland	55	49	45	–6	4	82.75	83.47	80.58
Lesotho	56	55	59	–1	–4	81.80	80.41	76.41
Guyana	57	38	33	–19	5	80.78	85.89	87.93
Ethiopia	58	58	54	0	4	80.13	78.28	78.65
Senegal	59	52	43	–7	9	79.65	81.47	82.29
Cote d'Ivoire	60	60	62	0	–2	78.85	77.46	74.24
Guinea	61	62	64	1	–2	76.74	75.94	73.26
Gambia	62	59	56	–3	3	74.23	77.55	78.11
Mauritania	63	63	63	0	0	72.98	74.16	73.79
Chad	64	68	70	4	–2	72.75	67.15	61.67
Pakistan	65	64	65	–1	–1	71.66	71.69	72.66
Burkina Faso	66	65	67	–1	–2	68.59	70.71	70.73
Djibouti	67	66	60	–1	6	63.88	69.96	74.43
Niger	68	70	68	2	2	63.20	64.26	65.74
Sudan	69	69	66	0	3	57.39	65.60	71.31
Liberia	70	67	61	–3	6	50.67	67.37	74.35
Mean	–	–	–	–	–	87.75	86.53	85.62

scenario for measured child development. The former two countries experienced upward movements thanks to education being weighted heavily. However, although Belarus moved down in rankings, it was the only one out of the three that actually improved (+8) its ranking under the most pessimistic scenario. Since the pessimistic scenario places more relative importance to health outcomes of children, it benefits countries such as Belarus which excels in this dimension. Malaysia also excels in this dimension which is why it moved up 19 positions under the most pessimistic scenario for measured child development.

Table 6 presents the rankings for the 2011 - 2016 period. Under the best-case scenario, Sierra Leone (+30), Paraguay (-20), and Guyana (-19) exhibited the largest movements in rankings. As a result of education receiving almost three-quarters of the weight, countries like Sierra Leone were able to improve their relative ranking. While it achieved a composite index of 64.80 under the worst-case scenario, it was able to achieve 91.90 under the best-case scenario, a difference of more than 27 points. Paraguay and Guyana experienced a deterioration in their rankings as a result of relatively poor achievements in net primary-school enrollment rates. Interesting though, under the worst-case scenario, Paraguay (+3) and Guyana (+5) improved their ranking as a result of childrens' health outcomes being weighted more, while Sierra Leone (-8) moved downwards.

5 Conclusion

Children's well-being in the era of SDGs is of salient importance for governments worldwide. The results presented in this paper have multifaceted implications. Much like the works of Pinar et al. (2013, 2017a) who focused on the HDI, we employed SDE in deriving the most optimistic scenario for measured child development where more countries achieve better measured outcomes based on the least variable combination of CDI components. We also derived the most pessimistic scenario where more countries achieve worse measured child development outcomes. Our analysis sheds some light and provides partial guidance for researchers and policy makers interested in measuring child development outcomes. The analysis presented here goes some ways towards informing public discourse surrounding the well-being of children.

Our findings revealed that the most optimistic scenario occurs when education receives more weight over time. While nutrition was the dominant dimension and main driver for improved measured child development outcomes during 1995 to 1999, more countries now find it easier to achieve better education outcomes for children as compared to nutrition or health outcomes. This is a result of more countries having improved net primary-school enrollment rates over time. Between 1995 and 2016, the average net primary-school enrollment improved by almost 10 percentage points. The most pessimistic scenario entails shifting majority of the weight towards the health component which is measured by survival rate for children under 5. Under this scenario, the nutrition component (measured by percentage of healthy weight children under 5) receives an increasing fraction of the weight over time, but not as much as health. This goes some ways towards indicating that improvements in children's education systems across countries have outpaced improvements in health care systems and nutrition.

The cross-country rankings identified countries that are trailing when it comes to education (e.g., countries which dropped in rankings as a result of the best-case scenario weighting). Those that have achieved good outcomes across the 3 dimensions of the CDI did not experience any major shifts in rankings as a result of re-weighting. However, those countries that have imbalanced achievement outcomes (i.e., those that excel in certain dimensions and not others) exhibited large deviations in rankings.

References

Agliardi, E., Agliardi, R., Pinar, M., Stengos, T., Topaloglou, N. (2012). A new country risk index for emerging markets: a stochastic dominance approach. *Journal of Empirical Finance*, *19*, 741–761.

Agliardi, E., Pinar, M., Stengos, T. (2014). A sovereign risk index for the eurozone based on stochastic dominance. *Finance Research Letters*, *11*, 375–384.

Agliardi, E., Pinar, M., Stengos, T. (2015). An environmental degradation index based on stochastic dominance. *Empirical Economics*, *48*, 439–459.

Barrett, G.F., & Donald, S.G. (2003). Consistent tests for stochastic dominance. *Econometrica*, *71*, 71–104.

Bravo, H.C., & Theussl, S. (2016). Rcplex: R Interface to CPLEX. https://CRAN.R-project.org/package=Rcplex.

Chzhen, Y., & Ferrone, L. (2017). Multidimensional child deprivation and poverty measurement: Case study of Bosnia and Herzegovina. *Social Indicators Research*, *131*, 999–1014.

Chzhen, Y., Gordon, D., Handa, S. (2017). Measuring multidimensional child poverty in the era of the sustainable development goals. *Child Indicators Research*, *12*, 1–3.

Davidson, R., & Duclos, J.Y. (2000). Statistical inference for stochastic dominance and the measurement of poverty and inequality. *Econometrica*, *68*, 1435–1464.

Fleurbaey, M. (2012). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, *47*, 1029–1075.

Grantham-McGregor, S., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L., Strupp, B. (2007). Development potential in the first 5 years for children in developing countries. *The Lancet*, *369*, 60–70.

Hagerty, M.R., & Land, K.C. (2007). Constructing summary indices of quality of life: a model for the effect of heterogeneous importance weights. *Sociological Methods and Research*, *35*, 455–496.

List, C. (2004). Multidimensional welfare aggregation. *Public Choice*, *119*, 119–142.

Maasoumi, E. (1999). Multidimensional approaches to welfare analysis. In Silber, J. (Ed.) *Handbook of income inequality measurement*. New York: Klewer Academic Publishers.

Main, G., & Bradshaw, J. (2016). Child poverty in the UK: measures, prevalence and intra-household sharing. *Critical Social Policy*, *36*, 38–61.

Neuman, M.J., & Devercelli, A.E. (2013). *What matters most for early childhood development: a framework paper. Systems Approach for Better Education Results (SABER) Working Paper Series, No. 5*. Washington DC: World Bank Group.

Pinar, M. (2015). Measuring world governance: revisiting the institutions hypothesis. *Empirical Economics*, *48*, 747–778.

Pinar, M., Stengos, T., Topaloglou, N. (2013). Measuring human development: a stochastic dominance approach. *Journal of Economic Growth*, *18*, 69–108.

Pinar, M., Stengos, T., Yazgan, M.E. (2015). Measuring human development in the mena region. *Emerging Markets Finance and Trade*, *51*, 1179–1192.

Pinar, M., Stengos, T., Topaloglou, N. (2017a). Testing for the implicit weights of the dimensions of the human development index using stochastic dominance. *Economics Letters*, *161*, 38–42.

Pinar, M., Stengos, T., Yazgan, M.E. (2017b). Quantile forecast combination using stochastic dominance. *Empirical Economics*, 1–39.

R Core Team (2018). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Roelen, K. (2017). Monetary and multidimensional child poverty: a contradiction in terms? *Development and Change*, *48*, 502–533.

Roelen, K., & Gassmann, F. (2008). Measuring child poverty and well-being: a literature review. Maastricht Graduate School of Governance Working Paper Series, No. 2008/WP001, Maastricht University.

Save the Children UK (2008). The Child Development Index: holding governments to account for children's wellbeing.

Save the Children UK (2012). The Child Development Index 2012: progress, challenges, and inequality.

Scaillet, O., & Topaloglou, N. (2010). Testing for stochastic dominance efficiency. *Journal of Business & Economic Statistics*, 169–180.

Trani, J.F., Biggeri, M., Mauro, V. (2013). The multidimensionality of child poverty: evidence from Afghanistan. *Social Indicators Research*, *112*, 391–416.

United Nations (2015a). The Millennium Development Goals report 2015.

United Nations (2015b). Transforming our world: The 2030 agenda for sustainable development.

World Health Organization (2007). Early child development: A powerful equalizer.

Paper 2B

Social Indicators Research (2019) 143:917–954
https://doi.org/10.1007/s11205-018-2014-0

Stochastic Dominance Approach to OECD's Better Life Index

Tahsin Mehdi[1]

Accepted: 8 October 2018 / Published online: 13 October 2018

Abstract

Ever since the inception of OECD's Better Life Index in 2011, a string of literature have emerged offering different aggregation procedures for the 11 dimensions of the index encompassing the broad categories of material living standards and quality of life. What is the most optimistic weighting scheme that allows more countries to achieve better measured well-being outcomes? What is the most pessimistic weighting scheme that worsens outcomes for more countries? Stochastic dominance efficiency is a data driven aggregation method that allows us to answer such questions which may be beneficial to policy makers and researchers. We offer rankings of countries across dimensions as well as rankings based on a single composite index aggregating all dimensional indicators. This type of analysis not only presents an opportunity to examine the sensitivity associated with re-weighting indicators, but this approach also reveals which indicators are driving overall improvement in measured well-being and which ones are hindering it. We find that the worst-case scenario rankings are generally more correlated with the equal-weighting scheme. And the best-case scenario weights offer a far more equal distribution of achievements across countries.

Keywords Better Life Index · Stochastic dominance · Multidimensional welfare

1 Introduction

The year 2015 marked the end of United Nation's (UN) Millennium Development Goals (MDGs), and ushered in the era of Sustainable Development Goals (SDGs), an overarching framework for the world to *"end poverty, protect the planet and ensure prosperity for all"*. It is increasingly being recognized that social welfare measures based on a single attribute (e.g., income per capita) are inadequate for measuring well-being [see, e.g., Boarini and D'Ercole (2013) or Fleurbaey (2012)]. Traditional measures such as gross domestic product per capita only provide a partial portrait of a nation's social well-being as it relies solely on income and is insensitive to other aspects individuals may care about in assessing quality of life (e.g., education, environment, safety, work-life balance, etc.).

✉ Tahsin Mehdi
tahsin.mehdi@ryerson.ca

[1] Department of Economics, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada

In recent decades, there has been a proliferation of multidimensional welfare measures such as the UN's Multidimensional Poverty Index (MPI) or Human Development Index (HDI), both of which are composite indices comprising health, education, and standard of living as its dimensions.[1] The ability of single indices to synthesize complex multi-faceted phenomena for benchmarking policies or ranking distributions is indeed appealing but the inherent subjectivity involved in weighting indicators has long been a point of contention among researchers, policy makers, and the public.[2]

In light of the recommendations put forth by the Commission on the Measurement of Economic Performance and Social Progress (Stiglitz et al. 2009), the Organisation for Economic Co-operation and Development (OECD) under its Better Life Initiative, introduced the Better Life Index (BLI) in 2011, a multidimensional measure for gauging social well-being of its member as well as some non-member countries (see, e.g., Organisation for Economic Co-operation and Development 2017). It comprises 11 dimensions which can be broadly categorized as material living standards (housing, income, and jobs), and quality of life (community, education, environment, governance, health, life satisfaction, safety, and work-life balance). Each dimension consists of one to four indicators which are normalized to lie between 0 and 10 prior to taking their arithmetic average in order to produce composite indices for each dimension (this is tantamount to weighting indicators equally within dimensions). A value of 0 is the worst possible outcome while a value of 10 is the best possible outcome. The dimensions and their indicators are provided in Table 1.

The BLI differs from the HDI in several ways. The BLI is published only for OECD countries and some non-OECD countries and it comprises more dimensions than the HDI. Also, unlike the HDI, the BLI does not weight any of its dimensions but rather it allows users to construct their own composite index based on their value judgments regarding the 11 dimensions by visiting www.oecdbetterlifeindex.org. That is, the OECD does not provide an "overall" index that encompasses all dimensions. Instead, they leave it up to users to assign relative importance to each dimension. What type of weighting scheme does it take for more countries to achieve better measured outcomes? Conversely, what type of weighting scheme does it take for more countries to achieve worse measured outcomes? Our aim in this paper will be to answer these questions. A preference for dimension *A* over dimension *B* does not necessarily imply that they also prefer all indicators in dimension *A* over those of dimension *B*. So rather than weighting the 11 dimensions, we go further and actually weight all of the indicators embedded within the dimensions. Prior to constructing this aggregate index, we also examine each of the dimensions separately.

The introduction of the BLI in 2011 has already been followed up by a number of studies surrounding the issue of indicator aggregation (e.g., Boarini et al. 2012; Decancq 2017; Kasparian 2012; Lorenz et al. 2017; Markovic et al. 2016; Mizobuchi 2014; Monika 2018; Nikolaev 2014, or von Reumont et al. 2017). A wide range of methods have been proposed in the literature from the *I*-distance method of Markovic et al. (2016) to the "Benefit of Doubt" (BOD) of Mizobuchi (2014). The former approach iteratively narrows down a given set of indicators based on their correlations while the BOD approach (closely related to Data Envelopment Analysis) maximizes composite indices for individual countries by assigning the most favourable weights to each country. The latter approach recognizes

[1] For a survey of multidimensional approaches to measuring well-being, see Maasoumi (1999) or Greco et al. (2018)

[2] See Decancq and Lugo (2013) for an exposition of weighting well-being indicators in a multidimensional setting.

Table 1 Organisation for economic co-operation and development (OECD) Better Life Index indicators

Dimension	Indicator	Unit	Years
Housing	Dwellings without basic facilities (−)	Percentage	2011–2017
	Housing expenditure (−)	Percentage	2012–2017
	Rooms per person (+)	Ratio	2011–2017
Income	Household net adjusted disposable income (+)	US Dollar	2011–2017
	Household net financial wealth (+)	US Dollar	2011–2017
Jobs	Labour market insecurity (−)	Percentage	2012–2017
	Employment rate (+)	Percentage	2011–2017
	Long-term unemployment rate (−)	Percentage	2012–2017
	Personal earnings (+)	US Dollar	2012–2017
Community	Quality of support network (+)	Percentage	2011–2017
Education	Educational attainment (+)	Percentage	2011–2017
	Student skills (+)	Average score	2011–2017
	Years in education (+)	Years	2012–2017
Environment	Air pollution (−)	$\mu g/m^3$	2011–2017
	Water quality (+)	Percentage	2012–2017
Civic engagement	Stakeholder engagement for developing regulations (+)	Average score	2011–2017
	Voter turnout (+)	Percentage	2011–2017
Health	Life expectancy (+)	Years	2011–2017
	Self-reported health (+)	Percentage	2011–2017
Life satisfaction	Life satisfaction (+)	Average score	2011–2017
Safety	Assault rate (−)	Percentage	2011–2015
	Feeling safe walking alone at night (+)	Percentage	2016–2017
	Homicide rate (−)	Ratio	2011–2017
Work-life balance	Employees working very long hours (−)	Percentage	2011–2017
	Time devoted to leisure and personal care (+)	Hours	2011–2017
	Employment rate of women with children (+)	Percentage	2011

The (+/−) in parentheses indicates whether the indicator measures a positive (+) or negative (−) component

comparative advantages of each country and assigns weights accordingly. However, this may not be appropriate since re-weighting indicators differently for each country would render the rankings incomparable. A fixed weighting scheme ensures that cross-country performance is measured fairly but it does not, however, mean that indicators should be weighted equally. Furthermore, much of the existing aggregation procedures are quite reliant on correlations or variability of welfare attributes (i.e., the second moment of the distribution) rather than focusing on the entire distribution.

A fairly common approach for reducing dimensionality is principal component analysis (PCA) which attempts to construct a set of linearly uncorrelated variables from a set of (possibly) correlated variables through orthogonal transformation. The first principal component explains the highest proportion of the variance in the original variables, and each subsequent component accounts for as much of the remaining variability as possible. Ogwang and Abdou (2003) and Biswas and Caliendo (2002) used the technique in examining the HDI. One criticism leveled against PCA is that it relies solely on the second moment of the distribution (after standardizing for common mean). Quiet often, welfare attributes are not characterized by just a single moment of a distribution.

In this paper, we take into account all moments of the underlying welfare distributions by relying on consistent tests of stochastic dominance efficiency (SDE) proposed by Scaillet and Topaloglou (2010). We obtain weighting schemes along with their associated country rankings for two extreme scenarios: best-case (optimistic) and worst-case (pessimistic). The former results in more countries achieving better outcomes while the latter results in the converse. This gives us a chance to study the sensitivity of rankings to the choice of weights, which may be of great interest to policy makers and researchers. The SDE approach has several advantages relative to alternative methods, aside from the fact that it takes into account all moments of a distribution. The best-case scenario results in more countries achieving better measured outcomes based on the least variable components of a composite index. Another benefit of the SDE approach is that, under the best-case scenario, the SDE weighting scheme provides insight on which indicators are driving improvements in measured well-being over time. On the other hand, the SDE weighting scheme under the worst-case scenario gives us a glimpse into the indicators hindering improvements in measured well-being. We find that in general, the worst-case scenario rankings are more correlated with the equal-weighting scheme. Re-weighting has little effect on countries with balanced achievements across indicators and thus they experience very little fluctuations in rankings. However, countries that perform well in a few indicators but underperform in others, exhibit large shifts in rankings.

SDE is an extension of stochastic dominance which is a nonparametric statistical method of ranking distributions in a robust manner taking all moments of a distribution into account. Stochastic dominance has gained significant interest across multiple disciplines and the importance of robust welfare comparisons have been highlighted in works such as Davidson and Duclos (2000), Duclos et al. (2006), Bennett and Mitra (2013), List (2004), and Mehdi (2017). The SDE approach has been used to assess the HDI in Pinar et al. (2013, 2015), and Pinar et al. (2017a). The approach was also utilized by Pinar (2015) in examining the World Governance Indicators published by the World Bank.[3] While Pinar et al. (2013) and Pinar et al. (2017a) examined human development worldwide, Pinar et al. (2015) restricted the analysis to the Middle East and North Africa (MENA) region and showed how the choice set of countries can alter rankings. Agliardi et al. (2015) used the approach in developing an environmental degradation index. Agliardi et al. (2012, 2014), and Pinar et al. (2017b) are examples of how SDE can be applied in the domain of finance.

The reason stochastic dominance was not applied in deriving weights for multidimensional welfare indices prior to Pinar et al. (2013) was that until Scaillet and Topaloglou (2010), only pairwise testing of dominance was possible (e.g., Barrett and Donald 2003, or Davidson and Duclos 2000). A typical application of pairwise stochastic dominance, in economics, is poverty comparisons between two distributions (e.g., two countries). Rather than having to specify a single poverty line, stochastic dominance allows us to evaluate a continuum of poverty lines or even the entire income distribution (see, e.g., Davidson and Duclos 2000 or Duclos et al. 2006). Stochastic dominance is nonparametric and does not require the specification of an explicit social welfare functional form or probability distribution. First-order stochastic dominance (see Sect. 2) satisfies all types of non-decreasing utility functions. Stochastic dominance simply requires economic agents to be rational in the sense that more of a social “good” increases utility level while more of a social “bad”

[3] Pinar et al. (2013) found that weighting education relatively more results in the most optimistic scenario for measured human development. This study was later followed up by Pinar et al. (2017a) who found that under the updated methodology of the UN, the health dimension gets weighted increasingly more over time in the most optimistic scenario.

decreases utility level (monotonicity axiom). Scaillet and Topaloglou (2010) extended stochastic dominance by allowing for complete diversification. That is, they devised a method for testing a "benchmark" (e.g., equally weighted) set of linearly weighted combination of variables against all other possible linear weighting combinations.

The remainder of this paper is organized as follows. Section 2 provides an overview of the SDE methodology. In Sect. 3, we describe the data. We employ SDE and offer a discussion of our empirical results in Sect. 4. Section 5 concludes and reflects on the findings and their implications.

2 Stochastic Dominance Efficient Composite Indices

Consider a strictly stationary process $\{\mathbf{Y}_t; t \in \mathbb{Z}\}$ taking values in $\mathbb{R}^K$. The observations consist of a realization of $\{\mathbf{Y}_t; t = 1, \ldots, T\}$ where T is the total number of observations in the panel dataset of countries. These correspond to the observed values of the K indicators. Let $F(\mathbf{Y})$ be the continuous cumulative distribution function (CDF) of $\mathbf{Y} = (Y_1, \ldots, Y_K)'$ at point $\mathbf{y} = (y_1, \ldots, y_K)'$. Since different indicators are measured in different units (see Table 1), we follow OECD's normalization procedure and convert them into composite scores between 0 and 10. For positive components, this is done through the formula, $10 \times [y_{k,t} - \min(Y_k)]/[\max(Y_k) - \min(Y_k)]$ while negative components are re-cast as positive ones via $10 \times [1 - [y_{k,t} - \min(Y_k)]/[\max(Y_k) - \min(Y_k)]]$ for $k = 1, \ldots, K$, and $t = 1, \ldots, T$, where $y_{k,t}$ is the kth indicator for observation t.

Consider some weighting vector $\lambda \in \mathbb{L}$ where $\mathbb{L} := \{\lambda \in \mathbb{R}_+^K \mid e'\lambda = 1\}$, and e is a K-vector of ones that imply the (non-negative) weights sum to unity. We denote the CDF of the composite index $\lambda'\mathbf{Y}$ at some well-being level z as

$$G(z, \lambda; F) = \int_{\mathbb{R}^K} I(\lambda'\mathbf{u} \le z) dF(\mathbf{u}), \tag{1}$$

where $I(\cdot)$ is the indicator function that equals 1 if its argument $(\cdot)$ is true, and 0 otherwise. This CDF simply measures the proportion of composite indices that are no higher than some given index level, z.

We denote by $\tau = \{1/K\}_{k=1}^K$, the vector of equal weights which the OECD uses to report composite scores for its dimensions. We will consider this as our benchmark weighting vector. SDE will test whether equal weighting results in the best-case scenario in terms of producing the maximum measured well-being and lower variability across countries, given the K indicators, or whether it is possible to come up with a different set of weights, λ, that offers the best-case scenario. The end-result are indices based on the least variable combination of components that maximize measured well-being. With relatively little guidance regarding specifications of preferences and distributional assumptions regarding the components of well-being, stochastic dominance offers a generalized approach that allows us to relax several assumptions typically synonymous with economic theory.

Distribution of the composite index constructed from λ dominates the distribution of the composite index constructed from τ at first-order if $G(z, \lambda; F) \le G(z, \tau; F)$ for all $z \in \mathbb{R}$. This is illustrated in Fig. 1. The implication is that the proportion of countries under weighting vector, λ, with measured well-being below some given level z is less than (or at most equal to) the proportion of countries under (equal) weighting vector, τ.

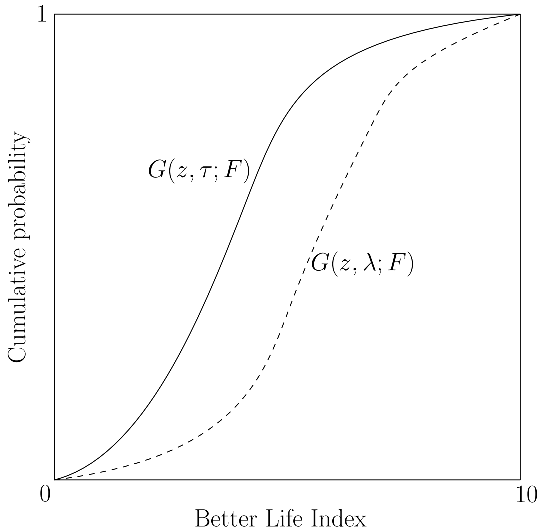

Fig. 1 First order stochastic dominance of λ weighted index over τ weighted index

In this paper, our focus is strictly on first-order stochastic dominance but generally, following Davidson and Duclos (2000), let

$$\mathcal{J}_j(z, \lambda; F) = \frac{1}{(j-1)!} \int_{\mathbb{R}^K} (z - \lambda' \mathbf{u})^{j-1} I(\lambda' \mathbf{u} \leq z) dF(\mathbf{u}). \tag{2}$$

Distribution of the composite index constructed from λ dominates that of the composite index constructed from τ at order j if $\mathcal{J}_j(z, \lambda; F) \leq \mathcal{J}_j(z, \tau; F)$ for all $z \in \mathbb{R}$. It is straightforward to check that for $j = 1$, $\mathcal{J}_1(z, \cdot; F) = G(z, \cdot; F)$. Also note that higher order dominance is implied by lower order dominance.

Testing for SDE of τ against all other linear weighting combinations at order j can be formally stated as,

$$H_0^j : \mathcal{J}_j(z, \tau; F) \leq \mathcal{J}_j(z, \lambda; F) \text{ for all } z \in \mathbb{R} \text{ and for all } \lambda \in \mathbb{L},$$
$$H_1^j : \mathcal{J}_j(z, \tau; F) > \mathcal{J}_j(z, \lambda; F) \text{ for some } z \in \mathbb{R} \text{ or for some } \lambda \in \mathbb{L}.$$

In practice, F will be unknown so expression (2) will need to be replaced by its empirical counterpart,

$$\mathcal{J}_j(z, \lambda; \hat{F}) = \frac{1}{T(j-1)!} \sum_{t=1}^{T} (z - \lambda' \mathbf{Y}_t)^{j-1} I(\lambda' \mathbf{Y}_t \leq z). \tag{3}$$

Under the null hypothesis, τ is SDE at order j in that no index can be constructed from λ that dominates (at order j) the index based on τ. Failure to reject the null implies that it is possible that some other alternative weighting combination leads to the nullification of dominance by τ over λ.

To test for (first-order) SDE of τ which is tantamount to testing whether τ offers the most optimistic scenario for measured well-being, we use the one-sided weighted Kolmogorov-Smirnov type test statistic proposed by Scaillet and Topaloglou (2010),

$$\hat{S} = \sqrt{T}\frac{1}{T} \sup_{z,\lambda} [G(z, \tau;\hat{F}) - G(z, \lambda;\hat{F})].$$

Naturally, one may be interested in whether τ actually gives rise to the most pessimistic (worst-case) scenario for measured well-being. The test statistic for the worst-case scenario is,

$$\hat{S} = \sqrt{T}\frac{1}{T} \sup_{z,\lambda} [G(z, \lambda;\hat{F}) - G(z, \tau;\hat{F})].$$

The theoretical underpinnings of these test statistics can be traced back to Barrett and Donald (2003), but this formulation allows for full diversification. That is, it allows us to test the benchmark weighting vector, τ, against all possible weighting combinations, λ. The test statistic has a non-standard limiting distribution and thus, p-values need to be obtained using bootstrap methods. We use the block bootstrap procedure which is shown to be quite reliable even in small samples (see, e.g., Pinar et al. 2013, or Scaillet and Topaloglou 2010). The other issue here is the practical computation of $\hat{S}$ which involves mixed integer linear programming. The computation of the test statistic is enumerated in detail in Pinar et al. (2013) and alternatively, Scaillet and Topaloglou (2010). We use R's Rcplex package (Bravo and Theussl 2016, and R Core Team 2018) which is an implementation of IBM's CPLEX optimizer.

If the null is rejected in the best-case scenario, then the benchmark weighting vector, τ, is not efficient which means it does not offer the most optimistic scenario for measured well-being across countries. There exists some other weighting vector, λ, that results in more countries achieving better outcomes. For the best-case scenario, we solve the following maximization problem,

$$\max_{\lambda} [G(z, \tau;\hat{F}) - G(z, \lambda;\hat{F})] \text{ for a given } z. \tag{4}$$

The weights arising from (4) give us an indication as to which indicators drive improvements in measured well-being.

Analogously, if the null is rejected in the worst-case scenario, then there exists some weighting vector, λ, that lowers outcomes for more countries than τ. For the worst-case scenario, we simply reverse the above maximization problem and solve,

$$\max_{\lambda} [G(z, \lambda;\hat{F}) - G(z, \tau;\hat{F})] \text{ for a given } z. \tag{5}$$

The resulting weights from (5) shed light on the indicators preventing improvement in measured well-being.

3 Data

We obtain the data from OECD's BLI website: www.oecdbetterlifeindex.org. They collect information on several aspects of objective and subjective well-being from its member countries as well as non-member countries such as Brazil, Russia, and South Africa in more recent years. As stated in Sect. 1, they consider 11 dimensions of well-being: housing, income, jobs, community, education, environment, civic engagement, health, life satisfaction, safety, and work-life balance. The indicators, currently available up to 2017, are

Table 2 Mean normalized indicators

Dimension	Indicator	2011	2012	2013	2014	2015	2016	2017
Housing	Dwellings without basic facilities	8.35	8.09	8.21	8.30	8.32	8.46	9.07
	Housing expenditure	–	3.98	3.89	3.70	3.44	3.28	4.64
	Rooms per person	5.18	5.42	4.36	4.74	4.76	5.07	5.23
Income	Household net adjusted disposable income	4.68	4.77	4.38	4.39	4.20	4.65	4.29
	Household net financial wealth	3.58	3.52	2.80	2.66	2.60	2.52	2.71
Jobs	Labour market insecurity	–	7.57	7.21	8.04	7.88	8.03	8.41
	Employment rate	6.00	6.00	5.84	5.61	5.27	5.99	5.70
	Long-term unemployment rate	6.66	6.70	6.55	7.69	8.08	8.10	8.13
	Personal earnings	–	5.53	5.28	5.65	4.95	5.44	5.07
Community	Quality of support network	6.66	7.61	6.63	7.58	7.35	6.10	6.28
Education	Educational attainment	7.07	7.03	7.05	6.92	6.89	6.99	6.96
	Student skills	5.98	5.70	6.50	6.65	6.58	6.78	6.94
	Years in education	–	6.50	5.41	5.91	5.73	5.59	4.03
Environment	Air pollution	7.75	7.88	7.33	7.06	7.06	6.59	5.82
	Water quality	–	7.26	7.09	7.35	6.51	6.98	6.33
Civic engagement	Stakeholder engagement for developing regulations	5.56	5.47	5.34	5.34	5.34	4.54	4.63
	Voter turnout	5.12	5.35	5.42	5.15	4.81	4.91	5.05
Health	Life expectancy	6.15	7.34	7.54	7.67	7.34	8.45	8.35
	Self-reported health	6.45	7.31	6.29	6.36	6.32	5.93	6.29
Life satisfaction	Life satisfaction	6.35	5.98	6.10	6.22	6.60	5.95	6.40
Safety	Assault rate	7.95	7.12	7.64	7.61	7.61	–	–
	Feeling safe walking alone at night	–	–	–	–	–	5.83	6.31
	Homicide rate	8.16	8.82	8.86	8.89	9.00	8.92	9.00
Work-life balance	Employees working very long hours	8.25	8.24	7.88	7.81	7.79	7.71	7.45
	Time devoted to leisure and personal care	6.18	4.22	6.72	5.54	5.54	6.25	5.95
	Employment rate of women with children	6.75	–	–	–	–	–	–

All indicators measuring negative achievements were re-cast into positive achievements. An achievement score of 0 is the worst possible outcome while a score of 10 is the best possible outcome

enumerated in Table 1 which also shows yearly availability and whether an indicator measures a positive or negative component. Since indicators are measured in different units, they are normalized to lie between 0 and 10, and the indicators measuring negative components were re-cast into positive components (see Sect. 2). The means and standard deviations of the normalized indicators are presented in Tables 2 and 3, respectively. The coverage of countries vary by year. The most amount of countries included in the sample was 38 in 2016 and 2017 while 2011 saw 34 countries in the sample, the fewest amount. Latvia and South Africa were excluded prior to 2016 as were Brazil and Russia in 2011.

For all but the environment and civic engagement dimensions, we were able to obtain stochastic dominance efficient weights. Since 2011 was the initial year for the BLI, some indicators were changed in subsequent years. Only the income and health dimensions maintained a consistent set of indicators which is why we were able to make use of all available

Table 3 Standard deviations of the normalized indicators

Dimension	Indicator	2011	2012	2013	2014	2015	2016	2017
Housing	Dwellings without basic facilities	2.48	2.76	2.49	2.49	2.48	2.34	1.79
	Housing expenditure	–	1.95	1.88	1.78	1.90	1.96	2.19
	Rooms per person	2.51	3.15	2.59	2.74	2.70	2.59	2.62
Income	Household net adjusted disposable income	2.20	2.36	2.50	2.39	2.45	2.49	2.42
	Household net financial wealth	2.73	2.66	2.49	2.34	2.28	2.21	2.23
Jobs	Labour market insecurity	–	2.31	2.26	1.87	1.80	2.14	2.05
	Employment rate	2.26	2.18	2.31	2.38	2.30	2.10	1.90
	Long-term unemployment rate	2.71	2.65	2.84	2.27	2.12	2.18	2.20
	Personal earnings	–	2.84	2.78	2.72	3.25	2.64	2.79
Community	Quality of support network	2.63	1.90	2.29	2.25	2.12	2.23	2.05
Education	Educational attainment	2.66	2.79	2.79	2.69	2.76	2.63	2.76
	Student skills	2.01	2.39	2.13	2.24	2.17	2.18	2.43
	Years in education	–	2.13	2.54	2.26	2.40	2.40	2.18
Environment	Air pollution	2.06	2.02	2.12	2.22	2.22	2.16	2.34
	Water quality	–	2.54	2.27	2.27	2.59	2.41	2.45
Civic engagement	Stakeholder engagement for developing regulations	2.70	2.73	2.82	2.82	2.82	2.62	2.58
	Voter turnout	2.62	2.61	2.59	2.56	2.77	2.63	2.75
Health	Life expectancy	2.78	2.24	2.39	2.31	2.36	1.80	1.77
	Self-reported health	2.53	2.35	2.42	2.40	2.40	2.52	2.51
Life satisfaction	Life satisfaction	2.69	2.78	2.76	2.91	2.92	2.84	2.88
Safety	Assault rate	1.93	2.35	1.98	1.95	1.95	–	–
	Feeling safe walking alone at night	–	–	–	–	–	2.61	2.56
	Homicide rate	2.03	2.16	2.19	2.27	2.30	2.19	2.00
Work-life balance	Employees working very long hours	1.95	1.96	2.14	2.10	2.05	2.11	2.32
	Time devoted to leisure and personal care	2.10	2.87	1.92	2.05	2.05	1.90	2.00
	Employment rate of women with children	1.91	–	–	–	–	–	–

All indicators measuring negative achievements were re-cast into positive achievements. An achievement score of 0 is the worst possible outcome while a score of 10 is the best possible outcome

years (2011–2017) for those two dimensions. For dimensions other than safety, we used data from 2012 to 2017. From 2016 onwards, one of the indicators from the safety dimension, assault rate, was replaced by an indicator measuring whether one feels safe walking alone at night which has more variability as reported in Table 3. To maximize the number of observations in our panel for the safety dimension, we use data from 2011 to 2015 which consistently used assault rate and homicide rate as its indicators. For the overall composite index based on all the dimensional indicators, we use data from 2012 to 2015 to maximize the number of observations in the panel. Rankings for the best-case scenario, worst-case scenario, and equal-weight are obtained for the latest year of the sub-period.

Table 4 Mean stochastic dominance efficient dimensional indicator weights for best-case and worst-case scenario for the Better Life Index

Dimension	Indicator	Period	Obs.	Best-case scenario		Worst-case scenario	
				Dominating indices	Weight	Dominating indices	Weight
Housing	Dwellings without basic facilities	2012–2017	211	210	0.864	110	0.014
	Housing expenditure	2012–2017	211	210	0.126	110	0.655
	Rooms per person	2012–2017	211	210	0.010	110	0.331
Income	Household net adjusted disposable income	2011–2017	242	229	0.972	49	0.302
	Household net financial wealth	2011–2017	242	229	0.028	49	0.698
Jobs	Labour market insecurity	2012–2017	215	212	0.700	158	0.022
	Employment rate	2012–2017	215	212	0.007	158	0.611
	Long-term unemployment rate	2012–2017	215	212	0.264	158	0.038
	Personal earnings	2012–2017	215	212	0.029	158	0.329
Community[a]	Quality of support network	2011–2017	254	–	1.000	–	1.000
Education	Educational attainment	2012–2017	219	209	0.731	158	0.099
	Student skills	2012–2017	219	209	0.195	158	0.176
	Years in education	2012–2017	219	209	0.074	158	0.725
Environment[b]	Air pollution	2012–2017	220	–	0.500	–	0.500
	Water quality	2012–2017	220	–	0.500	–	0.500
Civic engagement[b]	Stakeholder engagement for developing regulations	2011–2017	253	–	0.500	–	0.500
	Voter turnout	2011–2017	253	–	0.500	–	0.500
Health	Life expectancy	2011–2017	254	239	0.908	202	0.150
	Self-reported health	2011–2017	254	239	0.092	202	0.850
Life satisfaction	Life satisfaction	2011–2017	254	–	1.000	–	1.000
Safety	Assault rate	2011–2015	178	171	0.079	166	0.871
	Homicide rate	2011–2015	178	171	0.921	166	0.129
Work-life balance	Employees working very long hours	2011–2017	205	198	0.948	163	0.066
	Time devoted to leisure and personal care	2011–2017	205	198	0.052	163	0.934

[a]Dimension has only one indicator which always gets full weight

[b]No dominating indices could be found for either scenario

4 Results

We now present our findings and offer a discussion on the implications of our results. Since the community and life satisfaction dimensions consist of a single indicator, that indicator always gets a weight of unity and thus they were not considered for the dimensional analyses. They were, however, considered for the aggregate composite index which uses all dimensional indicators (see Sect. 4.8). Other than the environment and civic engagement dimensions, the null hypotheses for the remaining seven dimensions were rejected at the 1% or 5% level suggesting that there exist other weighting schemes that dominate equal-weighting. Our tests were based on 999 block bootstrap replications.

Upon rejecting the null, we proceed with maximizing the objective functions (4) and (5) in order to determine the best-case and worst-case scenario weighting schemes, respectively. Recall that the two scenarios represent two extremes and will allow us to examine the sensitivity of BLI. Under the best-case scenario, more countries achieve better measured outcomes while under the worst-case scenario, more countries achieve worse measured outcomes. We consider each dimension separately, before proceeding with constructing an aggregate composite index comprising all indicators from every dimension in Sect. 4.8. The SDE weights, when examining dimensions separately, are reported in Table 4 (weights sum to unity within dimensions) while weights for the aggregate composite index can be found in Table 12 (weights sum to unity across all indicators). The country rankings along with their composite scores are reported in Tables 5, 6, 7, 8, 9, 10, 11 and 13. In addition to the best-case and worst-case scenario, rankings and composite scores are also provided for the benchmark equal-weight scenario case. The last column of the tables report the index ranges, calculated as a country's maximum observed composite score less the minimum observed composite score.

The best-case scenario always produces the highest mean achievement scores while the worst-case scenario produces the lowest mean scores. This is to be expected as one of the properties of first-order stochastic dominance is that the dominating distribution is guaranteed to have the higher mean. The tables also present Spearman rank correlations and Gini coeffients for the indices.[4] The Spearman rank correlation is a metric that lies in the unit interval. Given two rankings, a number closer to 0 is an indication of relatively weak rank correlation while a number closer to unity is an indication of relatively strong rank correlation. An overarching theme we found is that generally the worst-case scenario rankings are more correlated with the equal-weight rankings. The Gini coeffient, a well-known metric for measuring the degree of dispersion in data, is almost always lower for the best-case scenario indices suggesting that not only do the best-case scenario weights result in more countries achieving better outcomes, but the distribution of achievements between countries is more stable than the equal-weight and worst-case scenarios.

The remainder of this section is organized into eight Sects. 4.1 (housing), 4.2 (income), 4.3 (jobs), 4.4 (education), 4.5 (health), 4.6 (safety), 4.7, and 4.8 (aggregate index). The first seven subsections offer a focused discussion on the individual dimensions we analyzed while the last subsection discusses the aggregate composite index comprising all indicators.

[4] The Gini coeffcient for some vector of composite indices, $x_1, \ldots, x_n$, is computed as $\frac{2}{\bar{x}n^2}\sum_{i=1}^{n} x_{(i)}(i-0.5)-1$, where $\bar{x}$ is the mean and $x_{(i)}$ is the ith order statistic of the vector (see, e.g., Davidson 2009).

Table 5 Rankings for housing dimension, 2017

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Norway	1	2	1	1	1	9.75	8.65	8.07	1.68
United States	2	1	2	− 1	− 1	9.63	8.90	8.03	1.60
Slovenia	3	10	7	7	3	9.53	7.21	6.38	3.15
Luxembourg	4	6	8	2	− 2	9.40	7.56	6.11	3.29
Netherlands	5	9	9	4	0	9.40	7.37	5.92	3.48
Sweden	6	11	13	5	− 2	9.39	7.19	5.74	3.65
Germany	7	12	14	5	− 2	9.37	7.18	5.74	3.63
Switzerland	8	14	16	6	− 2	9.28	7.07	5.33	3.95
Ireland	9	7	15	− 2	− 8	9.27	7.43	5.69	3.58
Australia	10	3	5	− 7	− 2	9.16	8.02	6.65	2.51
France	11	15	17	4	− 2	9.16	6.84	5.14	4.02
Canada	12	4	11	− 8	− 7	9.15	7.86	5.83	3.32
Spain	13	17	22	4	− 5	9.14	6.76	4.73	4.41
Portugal	14	18	18	4	0	9.04	6.61	4.95	4.09
Austria	15	20	20	5	0	9.03	6.43	4.77	4.26
Korea	16	5	3	− 11	2	8.96	7.58	7.96	1.38
Finland	17	21	26	4	− 5	8.93	6.42	4.13	4.80
Iceland	18	27	33	9	− 6	8.92	5.61	2.99	5.93
Italy	19	28	32	9	− 4	8.88	5.48	3.21	5.67
United Kingdom	20	22	28	2	− 6	8.85	6.31	3.72	5.13
Denmark	21	25	30	4	− 5	8.80	6.11	3.54	5.26
Greece	22	32	37	10	− 5	8.78	4.82	2.25	6.53
Czech Republic	23	29	36	6	− 7	8.77	5.18	2.62	6.15
Belgium	24	8	10	− 16	− 2	8.76	7.42	5.87	2.89
New Zealand	25	19	31	− 6	− 12	8.66	6.45	3.27	5.39

Table 5 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Hungary	26	23	12	− 3	11	8.58	6.30	5.81	2.77
Slovak Republic	27	36	38	9	− 2	8.57	4.55	2.06	6.51
Poland	28	34	35	6	− 1	8.38	4.74	2.66	5.72
Israel	29	26	24	− 3	2	8.33	5.68	4.62	3.71
Mexico	30	31	29	1	2	8.25	5.03	3.66	4.59
Estonia	31	16	6	− 15	10	8.00	6.80	6.53	1.47
Turkey	32	30	25	− 2	5	7.83	5.12	4.24	3.59
Brazil	33	35	27	2	8	7.77	4.73	3.87	3.90
Japan	34	24	23	− 10	1	7.67	6.19	4.71	2.96
Chile	35	13	4	− 22	9	7.43	7.13	7.07	0.36
Russia	36	33	19	− 3	14	6.24	4.77	4.81	1.47
Latvia	37	37	34	0	3	6.00	4.01	2.80	3.20
South Africa	38	38	21	0	17	0.92	2.42	4.76	3.84
Mean	–	–	–	–	–	8.47	6.31	4.90	–
Gini[a]	–	–	–	–	–	0.07	0.12	0.18	–
Spearman[b]	0.78	–	0.84	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 4

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

Table 6 Rankings for income dimension, 2017

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
United States	1	1	1	0	0	10.00	10.00	10.00	0.00
Luxembourg	2	3	4	1	− 1	9.03	6.66	5.66	3.37
Switzerland	3	2	2	-1	0	7.68	7.47	7.39	0.29
Norway	4	16	19	12	− 3	7.31	4.27	2.99	4.32
Germany	5	9	11	4	− 2	6.76	5.02	4.29	2.47
Australia	6	10	14	4	− 4	6.69	4.99	4.27	2.42
Austria	7	12	13	5	− 1	6.44	4.91	4.28	2.16
France	8	15	15	7	0	6.03	4.70	4.14	1.89
Sweden	9	5	6	− 4	− 1	5.91	5.51	5.34	0.57
Iceland	10	14	12	4	2	5.84	4.74	4.28	1.56
Belgium	11	4	3	− 7	1	5.76	5.81	5.83	0.07
Canada	12	7	8	− 5	− 1	5.69	5.26	5.08	0.61
Finland	13	19	21	6	− 2	5.46	3.53	2.72	2.74
Denmark	14	13	10	− 1	3	5.41	4.78	4.51	0.90
Netherlands	15	8	7	− 7	1	5.39	5.22	5.15	0.24
Japan	16	6	5	− 10	1	5.36	5.42	5.45	0.09
United Kingdom	17	11	9	− 6	2	5.27	4.98	4.85	0.42
Italy	18	17	16	− 1	1	4.55	4.07	3.86	0.69
Ireland	19	21	20	2	1	4.33	3.38	2.98	1.35
New Zealand	20	20	18	0	2	4.03	3.49	3.25	0.78
Israel	21	18	17	− 3	1	3.95	3.70	3.59	0.36
Spain	22	22	22	0	0	3.64	2.80	2.45	1.19
Korea	23	23	23	0	0	3.23	2.53	2.24	0.99
Czech Republic	24	25	25	1	0	3.03	2.17	1.82	1.21
Portugal	25	24	24	− 1	0	2.87	2.31	2.07	0.80

Table 6 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Slovenia	26	26	26	0	0	2.85	1.96	1.59	1.26
Slovak Republic	27	27	31	0	− 4	2.76	1.66	1.20	1.56
Poland	28	29	30	1	− 1	2.37	1.58	1.24	1.13
Estonia	29	28	28	− 1	0	2.31	1.60	1.30	1.01
Greece	30	32	32	2	0	1.82	1.38	1.19	0.63
Turkey	31	34	34	3	0	1.82	1.00	0.65	1.17
Hungary	32	30	27	− 2	3	1.78	1.50	1.39	0.39
Chile	33	31	29	− 2	2	1.71	1.41	1.29	0.42
Russia	34	35	36	1	− 1	1.69	0.87	0.53	1.16
Latvia	35	33	33	− 2	0	1.31	1.09	1.00	0.31
Mexico	36	36	37	0	− 1	0.89	0.53	0.37	0.52
Brazil	37	38	38	1	0	0.40	0.34	0.32	0.08
South Africa	38	37	35	− 1	2	0.02	0.43	0.59	0.57
Mean	–	–	–	–	–	4.25	3.50	3.19	–
Gini[a]	–	–	–	–	–	0.32	0.34	0.37	–
Spearman[b]	0.93	–	0.99	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 4

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

Table 7 Rankings for jobs dimension, 2017

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Japan	1	13	15	12	− 2	9.66	7.98	6.77	2.89
Iceland	2	1	1	− 1	0	9.62	9.53	9.56	0.09
Switzerland	3	2	2	− 1	0	9.59	9.18	8.85	0.74
Korea	4	19	24	15	− 5	9.55	7.26	5.20	4.35
Norway	5	3	3	− 2	0	9.51	8.67	7.72	1.79
Denmark	6	5	4	− 1	1	9.50	8.57	7.71	1.79
Germany	7	8	8	1	0	9.49	8.24	7.28	2.21
Israel	8	18	19	10	− 1	9.44	7.40	5.64	3.80
Czech Republic	9	21	21	12	0	9.41	6.99	5.44	3.97
Netherlands	10	6	5	− 4	1	9.36	8.42	7.69	1.67
United Kingdom	11	12	12	1	0	9.36	8.04	6.99	2.37
Austria	12	11	14	− 1	− 3	9.26	8.04	6.94	2.32
United States	13	4	6	− 9	− 2	9.22	8.58	7.42	1.80
Luxembourg	14	7	11	− 7	− 4	9.20	8.36	7.02	2.18
Finland	15	16	16	1	0	9.16	7.55	6.21	2.95
Canada	16	10	10	− 6	0	9.10	8.16	7.11	1.99
Ireland	17	17	17	0	0	9.00	7.46	6.14	2.86
Australia	18	9	7	− 9	2	8.93	8.20	7.30	1.63
Russia	19	24	25	5	− 1	8.91	6.63	5.04	3.87
Mexico	20	30	35	10	− 5	8.82	5.91	3.36	5.46
New Zealand	21	15	13	− 6	2	8.79	7.80	6.95	1.84
Estonia	22	23	22	1	1	8.73	6.70	5.40	3.33
Poland	23	26	28	3	− 2	8.64	6.35	4.49	4.15
Sweden	24	14	9	− 10	5	8.52	7.86	7.24	1.28
Slovenia	25	25	23	0	2	8.45	6.59	5.21	3.24

Table 7 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Hungary	26	27	27	1	0	8.44	6.17	4.49	3.95
Belgium	27	20	18	− 7	2	8.35	7.06	5.66	2.69
Brazil	28	32	34	4	− 2	8.22	5.52	3.68	4.54
France	29	22	20	− 7	2	8.20	6.80	5.55	2.65
Latvia	30	29	26	− 1	3	7.68	5.94	4.80	2.88
Chile	31	28	31	− 3	− 3	7.61	5.98	4.30	3.31
Slovak Republic	32	31	32	− 1	− 1	7.44	5.53	4.30	3.14
Portugal	33	33	29	0	4	7.40	5.51	4.39	3.01
Italy	34	34	33	0	1	6.89	5.32	3.93	2.96
Turkey	35	35	36	0	− 1	6.15	4.52	2.25	3.90
Spain	36	36	30	0	6	3.91	4.29	4.38	0.47
Greece	37	37	37	0	0	2.64	2.09	2.23	0.55
South Africa	38	38	38	0	0	0.22	0.20	0.03	0.19
Mean	–	–	–	–	–	8.22	6.83	5.65	–
Gini[a]	–	–	–	–	–	0.10	0.14	0.19	–
Spearman[b]	0.86	–	0.98	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 4

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

Table 8 Rankings for education dimension, 2017

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Japan	1	5	22	4	− 17	9.38	7.47	4.55	4.83
Russia	2	22	30	20	− 8	8.89	6.45	3.75	5.14
Finland	3	1	2	− 2	− 1	8.88	8.73	8.22	0.66
Canada	4	11	19	7	− 8	8.85	7.27	4.75	4.10
Czech Republic	5	15	18	10	− 3	8.82	6.97	5.07	3.75
Poland	6	8	12	2	− 4	8.79	7.37	5.65	3.14
Estonia	7	18	31	11	− 13	8.55	6.72	3.71	4.84
Korea	8	10	15	2	− 5	8.41	7.32	5.43	2.98
Slovenia	9	6	9	− 3	− 3	8.40	7.47	6.10	2.30
United States	10	21	20	11	1	8.34	6.60	4.75	3.59
Germany	11	4	8	− 7	− 4	8.31	7.50	6.30	2.01
Switzerland	12	13	16	1	− 3	8.31	7.09	5.39	2.92
Latvia	13	16	13	3	3	8.25	6.91	5.62	2.63
Slovak Republic	14	27	33	13	− 6	8.05	5.46	3.10	4.95
Sweden	15	7	4	− 8	3	7.77	7.46	7.11	0.66
Australia	16	2	1	− 14	1	7.74	8.49	9.40	1.66
Austria	17	23	21	6	2	7.70	6.38	4.71	2.99
Norway	18	14	10	− 4	4	7.69	7.06	5.95	1.74
Denmark	19	3	3	− 16	0	7.69	7.80	7.74	0.11
Israel	20	28	34	8	− 6	7.55	5.35	3.02	4.53
Ireland	21	9	6	− 12	3	7.53	7.35	6.66	0.87
United Kingdom	22	24	23	2	1	7.30	6.20	4.40	2.90
Hungary	23	26	28	3	− 2	7.25	5.62	3.89	3.36

Table 8 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Netherlands	24	12	7	– 12	5	7.18	7.17	6.60	0.58
New Zealand	25	20	14	– 5	6	6.99	6.63	5.54	1.45
Iceland	26	17	5	– 9	12	6.98	6.88	6.95	0.10
France	27	25	27	– 2	– 2	6.88	5.79	3.97	2.91
Belgium	28	19	11	– 9	8	6.80	6.68	5.93	0.87
Luxembourg	29	30	35	1	– 5	6.62	4.79	2.23	4.39
Greece	30	31	29	1	2	5.59	4.72	3.83	1.76
Chile	31	34	26	3	8	4.57	4.18	3.98	0.59
Spain	32	29	17	– 3	12	4.49	5.27	5.16	0.78
Italy	33	32	32	– 1	0	4.46	4.45	3.41	1.05
Portugal	34	33	24	– 1	9	3.06	4.35	4.13	1.29
Brazil	35	36	36	1	0	1.73	1.37	1.51	0.36
Turkey	36	35	25	– 1	10	1.08	2.54	3.98	2.90
South Africa	37	37	37	0	0	0.86	0.63	0.68	0.23
Mexico	38	38	38	0	0	0.35	0.60	0.32	0.28
Mean	–	–	–	–	–	6.74	5.98	4.83	–
Gini[a]	–	–	–	–	–	0.18	0.17	0.22	–
Spearman[b]	0.76	–	0.84	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 4

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

Table 9 Rankings for health dimension, 2017

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Switzerland	1	5	7	4	− 2	9.55	9.07	8.65	0.90
Australia	2	3	4	1	− 1	9.47	9.48	9.48	0.01
Spain	3	12	14	9	− 2	9.43	8.42	7.55	1.88
Iceland	4	10	10	6	0	9.32	8.67	8.12	1.20
Norway	5	9	9	4	0	9.32	8.84	8.42	0.90
Israel	6	4	5	− 2	− 1	9.31	9.28	9.26	0.05
Sweden	7	8	8	1	0	9.31	8.94	8.63	0.68
New Zealand	8	1	2	− 7	− 1	9.24	9.56	9.83	0.59
Luxembourg	9	15	16	6	− 1	9.19	8.13	7.21	1.98
Italy	10	21	22	11	− 1	9.18	7.73	6.49	2.69
Canada	11	2	1	− 9	1	9.18	9.55	9.86	0.68
France	12	18	20	6	− 2	9.15	7.89	6.81	2.34
Japan	13	34	37	21	− 3	9.13	5.26	1.94	7.19
Ireland	14	7	6	− 7	1	9.08	9.03	8.98	0.10
Netherlands	15	11	11	− 4	0	9.01	8.50	8.05	0.96
Finland	16	17	17	1	0	8.91	7.92	7.07	1.84
Belgium	17	13	12	− 4	1	8.82	8.26	7.78	1.04
Greece	18	14	13	− 4	1	8.81	8.24	7.75	1.06
Austria	19	19	18	0	1	8.81	7.86	7.05	1.76
United Kingdom	20	20	19	0	1	8.70	7.81	7.04	1.66
Denmark	21	16	15	− 5	1	8.66	7.93	7.30	1.36
Slovenia	22	22	24	0	− 2	8.59	7.34	6.27	2.32
Germany	23	23	25	0	− 2	8.51	7.27	6.21	2.30
Korea	24	35	38	11	− 3	8.46	4.66	1.40	7.06
Portugal	25	31	34	6	− 3	8.38	5.74	3.47	4.91

Table 9 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
United States	26	6	3	− 20	3	8.25	9.04	9.71	1.46
Chile	27	28	30	1	− 2	7.85	6.33	5.03	2.82
Czech Republic	28	27	28	− 1	− 1	7.77	6.60	5.59	2.18
Turkey	29	24	23	− 5	1	7.62	6.94	6.35	1.27
Poland	30	30	31	0	− 1	7.34	6.09	5.01	2.33
Estonia	31	33	33	2	0	7.27	5.53	4.04	3.23
Slovak Republic	32	25	26	− 7	− 1	7.17	6.65	6.20	0.97
Hungary	33	32	32	− 1	0	6.66	5.59	4.67	1.99
Mexico	34	29	27	− 5	2	6.58	6.29	6.04	0.54
Brazil	35	26	21	− 9	5	6.54	6.60	6.65	0.11
Latvia	36	36	35	0	1	6.12	4.48	3.07	3.05
Russia	37	37	36	0	1	4.94	3.57	2.39	2.55
South Africa	38	38	29	0	9	0.57	3.11	5.29	4.72
Mean	–	–	–	–	–	8.16	7.32	6.60	–
Gini[a]	–	–	–	–	–	0.09	0.13	0.18	–
Spearman[b]	0.80	–	0.98	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 4

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

Table 10 Rankings for safety dimension, 2015

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Japan	1	1	2	0	– 1	9.99	9.96	9.92	0.07
United Kingdom	2	4	5	2	– 1	9.96	9.74	9.55	0.41
Iceland	3	8	11	5	– 3	9.90	9.39	8.94	0.96
Denmark	4	17	19	13	– 2	9.82	8.87	8.03	1.79
Austria	5	14	15	9	– 1	9.82	9.07	8.40	1.42
Slovenia	6	18	20	12	– 2	9.79	8.85	8.03	1.76
Poland	7	2	3	– 5	– 1	9.77	9.84	9.89	0.12
Germany	8	15	16	7	– 1	9.77	8.96	8.25	1.52
Australia	9	5	6	– 4	– 1	9.76	9.55	9.37	0.39
Luxembourg	10	23	23	13	0	9.76	8.68	7.72	2.04
Norway	11	13	14	2	– 1	9.75	9.07	8.47	1.28
Switzerland	12	20	21	8	– 1	9.73	8.70	7.79	1.94
Ireland	13	9	10	– 4	– 1	9.73	9.34	8.99	0.74
Czech Republic	14	11	12	– 3	– 1	9.71	9.25	8.84	0.87
Spain	15	22	22	7	0	9.69	8.68	7.79	1.90
Korea	16	6	7	– 10	– 1	9.65	9.49	9.35	0.30
France	17	25	27	8	– 2	9.64	8.33	7.18	2.46
Italy	18	24	25	6	– 1	9.62	8.44	7.40	2.22
New Zealand	19	7	8	– 12	– 1	9.61	9.43	9.27	0.34
Sweden	20	27	29	7	– 2	9.59	8.27	7.10	2.49
Canada	21	3	1	– 18	2	9.56	9.76	9.94	0.38
Slovak Republic	22	12	13	– 10	– 1	9.55	9.08	8.67	0.88
Netherlands	23	26	26	3	0	9.53	8.32	7.24	2.29
Finland	24	10	9	– 14	1	9.52	9.30	9.11	0.41
Hungary	25	19	17	– 6	2	9.48	8.80	8.21	1.27

Table 10 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Turkey	26	28	28	2	0	9.42	8.21	7.15	2.27
Portugal	27	29	30	2	− 1	9.41	7.93	6.63	2.78
Greece	28	21	18	− 7	3	9.36	8.70	8.12	1.24
Belgium	29	30	33	1	− 3	9.34	7.54	5.94	3.40
Israel	30	31	32	1	− 1	8.92	7.39	6.03	2.89
United States	31	16	4	− 15	12	8.19	8.94	9.60	1.41
Chile	32	33	34	1	− 1	8.12	6.75	5.55	2.57
Estonia	33	32	31	− 1	1	8.07	7.28	6.59	1.48
Russia	34	34	24	0	10	5.26	6.43	7.47	2.21
Mexico	35	36	36	1	0	0.77	0.42	0.11	0.66
Brazil	36	35	35	− 1	0	0.34	2.13	3.71	3.37
Mean	–	–	–	–	–	8.89	8.30	7.79	–
Gini[a]	–	–	–	–	–	0.08	0.10	0.12	–
Spearman[b]	0.72	–	0.96	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 4

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

Table 11 Rankings for work-life balance dimension, 2017

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Netherlands	1	1	2	0	− 1	9.86	9.35	8.85	1.01
Russia	2	9	16	7	− 7	9.80	8.06	6.38	3.42
Sweden	3	7	9	4	− 2	9.57	8.29	7.06	2.51
Denmark	4	2	4	− 2	− 2	9.36	9.05	8.75	0.61
Latvia	5	27	35	22	− 8	9.11	6.36	3.69	5.42
Estonia	6	14	17	8	− 3	9.09	7.69	6.33	2.76
Norway	7	6	6	− 1	0	9.04	8.49	7.96	1.08
Hungary	8	13	12	5	1	9.01	7.85	6.72	2.29
Luxembourg	9	12	11	3	1	8.82	7.86	6.93	1.89
Finland	10	11	10	1	1	8.78	7.86	6.98	1.80
Belgium	11	5	5	− 6	0	8.75	8.60	8.46	0.29
Italy	12	15	18	3	− 3	8.74	7.49	6.28	2.46
Canada	13	20	30	7	− 10	8.73	6.88	5.10	3.63
Spain	14	4	3	− 10	1	8.70	8.78	8.85	0.15
Germany	15	8	7	− 7	1	8.64	8.27	7.91	0.73
Ireland	16	10	8	− 6	2	8.58	7.90	7.24	1.34
Slovenia	17	18	23	1	− 5	8.57	7.23	5.93	2.64
Slovak Republic	18	16	14	− 2	2	8.44	7.48	6.56	1.88
Czech Republic	19	17	13	− 2	4	8.24	7.44	6.67	1.57
Switzerland	20	19	15	− 1	4	7.91	7.22	6.55	1.36
Poland	21	25	31	4	− 6	7.89	6.46	5.06	2.83
Austria	22	23	28	1	− 5	7.88	6.61	5.38	2.50
France	23	3	1	− 20	2	7.86	8.87	9.85	1.99
Brazil	24	26	29	2	− 3	7.77	6.43	5.13	2.64
Greece	25	22	25	− 3	− 3	7.75	6.70	5.67	2.08

Table 11 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Portugal	26	21	19	− 5	2	7.53	6.85	6.20	1.33
Chile	27	24	20	− 3	4	7.01	6.59	6.19	0.82
United States	28	30	32	2	− 2	6.55	5.77	5.02	1.53
United Kingdom	29	28	21	− 1	7	6.27	6.23	6.19	0.08
Australia	30	31	33	1	− 2	6.05	5.39	4.76	1.29
New Zealand	31	29	22	− 2	7	5.60	5.81	6.02	0.42
Iceland	32	33	34	1	− 1	5.49	4.85	4.23	1.26
Israel	33	36	36	3	0	5.47	4.56	3.69	1.78
South Africa	34	32	26	− 2	6	4.55	5.08	5.60	1.05
Korea	35	35	27	0	8	3.94	4.72	5.48	1.54
Japan	36	34	24	− 2	10	3.68	4.78	5.83	2.15
Mexico	37	37	37	0	0	1.23	0.84	0.46	0.77
Turkey	38	38	38	0	0	0.00	0.00	0.00	0.00
Mean	–	–	–	–	–	7.38	6.70	6.05	–
Gini[a]	–	–	–	–	–	0.15	0.15	0.17	–
Spearman[b]	0.84	–	0.92	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 4

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

Table 12 Mean stochastic dominance efficient weights for best-case and worst-case scenario for the composite Better Life Index

Indicator	Period	Obs.	Best-case scenario		Worst-case scenario	
			Dominating indices	Weight	Dominating indices	Weight
Dwellings without basic facilities	2012–2015	124	122	0.049	73	0.000
Housing expenditure	2012–2015	124	122	0.006	73	0.066
Rooms per person	2012–2015	124	122	0.000	73	0.000
Household net adjusted disposable income	2012–2015	124	122	0.000	73	0.000
Household net financial wealth	2012–2015	124	122	0.000	73	0.337
Labour market insecurity	2012–2015	124	122	0.029	73	0.000
Employment rate	2012–2015	124	122	0.000	73	0.003
Long-term unemployment rate	2012–2015	124	122	0.133	73	0.000
Personal earnings	2012–2015	124	122	0.000	73	0.000
Quality of support network	2012–2015	124	122	0.000	73	0.000
Educational attainment	2012–2015	124	122	0.012	73	0.000
Student skills	2012–2015	124	122	0.000	73	0.000
Years in education	2012–2015	124	122	0.000	73	0.000
Air pollution	2012–2015	124	122	0.049	73	0.000
Water quality	2012–2015	124	122	0.000	73	0.000
Stakeholder engagement for developing regulations	2012–2015	124	122	0.000	73	0.361
Voter turnout	2012–2015	124	122	0.056	73	0.229
Life expectancy	2012–2015	124	122	0.000	73	0.000
Self-reported health	2012–2015	124	122	0.022	73	0.000
Life satisfaction	2012–2015	124	122	0.000	73	0.000
Assault rate	2012–2015	124	122	0.052	73	0.000
Homicide rate	2012–2015	124	122	0.489	73	0.000
Employees working very long hours	2012–2015	124	122	0.103	73	0.000
Time devoted to leisure and personal care	2012–2015	124	122	0.000	73	0.004

4.1 Housing

The housing dimension comprises three indicators: *dwellings without basic facilities*, *rooms per person*, and *housing expenditure*. *Dwellings without without basic facilities* measures the proportion of the population living in a dwelling without indoor flushing toilet for the sole use of their households. *Rooms per person* excludes kitchens, scullery/utility rooms, bathrooms, garages, consulting rooms, offices, and shops. *Housing expenditure* measures the proportion of household gross adjusted disposable income devoted towards housing and maintenance of the house.

Under the best(worst)-case scenario, the weights were 0.864 (0.014) for *dwellings without basic facilities*, 0.126 (0.655) for *housing expenditure*, and 0.010 (0.331) for *rooms per person*. This reveals that the *dwellings without basic facilities* indicator is the dominant component or main driving force behind improved measured well-being in the housing

Table 13 Country rankings for overall composite index comprising all indicators, 2015

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Luxembourg	1	13	8	12	5	9.39	7.26	5.35	4.04
Sweden	2	3	2	1	1	9.37	7.81	6.97	2.40
Norway	3	2	16	− 1	− 14	9.35	7.85	4.41	4.94
Denmark	4	7	11	3	− 4	9.32	7.66	5.06	4.26
Australia	5	1	3	− 4	− 2	9.21	7.96	6.87	2.34
Iceland	6	10	20	4	− 10	9.08	7.42	3.92	5.16
Germany	7	12	25	5	− 13	9.03	7.32	3.53	5.50
New Zealand	8	8	7	0	1	9.01	7.45	5.38	3.63
Canada	9	5	6	− 4	− 1	8.98	7.69	5.64	3.34
Austria	10	17	14	7	3	8.94	6.88	4.65	4.29
Finland	11	11	17	0	− 6	8.88	7.36	4.28	4.60
Netherlands	12	9	12	− 3	− 3	8.86	7.43	5.04	3.82
United Kingdom	13	16	4	3	12	8.86	7.02	6.03	2.83
France	14	18	26	4	− 8	8.82	6.44	3.52	5.30
Switzerland	15	4	10	− 11	− 6	8.75	7.77	5.16	3.59
Belgium	16	14	9	− 2	5	8.75	7.20	5.21	3.54
Czech Republic	17	21	30	4	− 9	8.67	5.89	2.72	5.95
Ireland	18	15	13	− 3	2	8.62	7.09	4.77	3.85
Italy	19	24	21	5	3	8.47	5.66	3.84	4.63
Korea	20	22	5	2	17	8.44	5.88	5.68	2.76
Slovenia	21	20	19	− 1	1	8.38	6.05	3.96	4.42
Hungary	22	29	27	7	2	8.38	5.15	3.45	4.93
Poland	23	26	18	3	8	8.26	5.50	4.08	4.18
Israel	24	25	32	1	− 7	8.18	5.63	2.61	5.57
Slovak Republic	25	28	33	3	− 5	8.17	5.25	2.43	5.74

Table 13 (continued)

	Rank			Deviation from equal-weight		Index			Index range
	Best-case	Equal-weight	Worst-case	Best-case	Worst-case	Best-case	Equal-weight	Worst-case	
Japan	26	19	15	− 7	4	8.13	6.24	4.42	3.71
United States	27	6	1	− 21	5	8.07	7.67	7.13	0.94
Spain	28	23	22	− 5	1	7.74	5.73	3.79	3.95
Portugal	29	30	28	1	2	7.73	5.04	3.18	4.55
Estonia	30	27	34	− 3	− 7	7.72	5.40	1.73	5.99
Turkey	31	35	24	4	11	7.29	3.43	3.58	3.86
Greece	32	32	29	0	3	7.17	4.47	2.83	4.34
Chile	33	34	36	1	− 2	6.84	4.10	0.72	6.12
Russia	34	31	35	− 3	− 4	6.19	4.62	1.73	4.46
Brazil	35	33	31	− 2	2	3.63	4.37	2.67	1.70
Mexico	36	36	23	0	13	3.18	3.22	3.76	0.58
Mean	–	–	–	–	–	8.16	6.25	4.17	–
Gini[a]	–	–	–	–	–	0.08	0.12	0.20	–
Spearman[b]	0.86	–	0.77	–	–	–	–	–	–

The weights for the best-case and worst-case scenarios can be found in Table 12

[a]A value closer to 0 implies a more equal distribution while a value closer to 1 implies a more unequal distribution

[b]A value closer to 0 implies a weaker correlation with equal-weight rank while a value closer to 1 implies stronger

dimension, evidenced by the fact that the most optimistic view has it being weighted considerably higher than the other two indicators. In other words, more countries find it easier to achieve better measured outcomes when this component receives relatively more weight because its upper bound is more achievable. Tables 2 and 3 sheds some light on this as well by showing that the *dwelling without basic facilities* indicator has the highest mean and additionally also happens to be the least variant among the housing indicators. The most pessimistic view weights *housing expenditure* relatively more suggesting that more countries find it harder to achieve better measured outcomes when this receives relatively more weight because its upper bound is harder to achieve.

Table 5 reports the rankings for 2017. There are several drastic rank deviations between scenarios. Under the best-case scenario weighting scheme, Korea, Belgium, Estonia, Japan, and Chile dropped at least 10 positions from the equal-weight ranking, with Chile dropping 22 positions going from being ranked 13th under equal-weighting to 35th (out of 38 countries). On the other hand, under the worst-case scenario weighting scheme, Chile moved to 4th place as a result of weighting *housing expenditure* more heavily. Another case where ranking changes quite dramatically is Estonia which ranked 16th under equal-weighting, 31st under best-case scenario, and 6th under worst-case scenario.

As discussed before, SDE gives us a chance to study the sensitivity of rankings associated with re-weighting indices. The last column of Table 5 reports the index ranges which shows the composite score variation achieved by each country. They are calculated as the maximum observed composite score less the minimum observed composite score. Unsurprisingly, the top 2 countries (Norway and the United States) exhibited little variation in achievement scores (1.68 for Norway and 1.60 for the United States) and they remained in the top 2 regardless of our weighting scenarios. This goes some ways towards indicating that such countries are less subject to normative judgments (i.e., re-weighting has little effect) as they have achieved well-balanced outcomes across the housing dimension components. The highest composite score variations were observed for Greece (6.53), Slovak Republic (6.51), and Czech Republic (6.15), all of which ranked among the bottom half of countries regardless of the three scenarios. Although in 2017, Greece, Slovak Republic, and Czech Republic achieved relatively good outcomes for the *dwellings without basic facilities* indicator (relatively more important to induce the best-case scenario), they ranked near the bottom when it came to *housing expenditure* (relatively more important for the worst-case scenario).

The Gini coefficients for the best-case, equal-weight, and worst-case scenarios were 0.07, 012, and 0.18, respectively, suggesting that the best-case scenario not only produces higher mean achievement scores but it also offers the most stable distribution of measured outcomes. The Spearman rank correlation with the equal-weight scenario was 0.78 for the best-case scenario while it was 0.84 for the worst-case scenario which indicates that the latter ranking is a little more correlated with the benchmark case.

4.2 Income

Income has traditionally been the "yard stick" used to judge welfare distributions. For the BLI, this dimension is captured by *household net adjusted disposable income*, and *household net financial wealth*. Both are measured in current US dollars at current purchasing power parities. *Household net adjusted disposable income* is the maximum amount a household can afford to consume without having to reduce its assets or increase liabilities. *Household net financial wealth* includes currency and deposits, securities other than share, loans, shares and other equity, insurance technical reserves, and other accounts receivable or payable, net of household financial liabilities.

SDE weights for the best(worst)-case scenario were 0.972 (0.302) for *household net adjusted disposable income*, and 0.028 (0.698) for *household net financial wealth*. Table 6 reports the rankings for 2017. Spearman rank correlations with equal-weight ranking for best-case and worst-case scenarios were 0.93 and 0.99, respectively, suggesting both scenarios are highly correlated with the benchmark ranking (i.e., equal-weight). United States consistently ranked at the top in all three scenarios (with a perfect composite score of 10) which is not surprising given that it had the highest *household net adjusted disposable income* as well as *household net financial wealth* of all the countries in the sample. Under the best-case scenario which favours *household net adjusted disposable income*, Japan dropped 10 positions going from 6th (equal-weight) to 16th (best-case), while Norway improved its standing by 12 spots going from 16th (equal-weight) to 4th (best-case).

Majority of the countries experienced very little variation in composite scores under the three scenarios suggesting that there weren't much variation between achievements in the two indicators for most countries. This is re-enforced by the Gini coefficients (0.32 for the best-case scenario, 0.34 for equal-weight, and 0.37 for worst-case scenario) which suggest the distribution of measured outcomes are about the same across the three different scenarios.

4.3 Jobs

A nation's prosperity depends on a productive citizenry which goes hand-in-hand with the ability to create meaningful employment opportunities. The jobs dimension includes four indicators, more than any other dimension. They are *employment rate*, *long-term unemployment rate*, *labour market insecurity*, and *personal earnings*.[5] *Employment rate* measures the proportion of employed persons between the ages of 15 and 64. *Long-term unemployment rate* is the proportion of persons in the labour force who have been unemployed for at least one year. *Labour market insecurity* (an indicator that is used after 2015) is a measure of expected earnings loss, expressed as the proportion of previous earnings associated with unemployment. *Personal earnings* is reported in US dollars at current purchasing power parities. It measures the average annual wages per full-time equivalent employee.

The best(worst)-case scenario weights were 0.700 (0.022) for *labour market insecurity*, 0.264 (0.038) for *long-term unemployment rate*, 0.029 (0.329) for *personal earnings*, and 0.007 (0.611) for *employment rate*. As seen in Tables 2 and 3, although the *personal earnings* and *employment rate* indicators are less variable, the other two indicators had higher means in 2017. *Labour market insecurity* with a weight of 0.700 under the best-case scenario is the main indicator driving improvement in the jobs dimension as more countries achieve better measured outcomes if it gets weighted more. More than a quarter of the weight goes towards *long-term unemployment rate* so in the most optimistic scenario, this indicator also plays a role. The 2017 rankings are reported in Table 7 and from the Spearman rank correlation, we can see that the worst-case scenario is substantially more correlated with the equal-weight rank.

[5] *Labour market insecurity* was introduced after 2015. Prior to that a similar but slightly different metric, *job security*, was used in its place that measured duration of employment tenure or probability of becoming unemployed.

The best-case scenario has Japan, Korea, Israel, Czech Republic, and Mexico improving their ranks by at least 10 positions. This is because in 2017, these countries were either near the top of the rankings in terms of job security, and/or long-term employment. Interestingly, the bottom 6 countries (Portugal, Italy, Turkey, Spain, Greece, and South Africa) did not deviate from their equal-weight rank at all in the best-case scenario (minor deviations under the worst-case scenario). These countries were among the bottom 10 in terms of *labour market insecurity* in 2017 and with the exception of Turkey, they ranked at the very bottom in terms of *long-term unemployment rate* as well. Regardless of the scenario, Iceland and Switzerland ranked in the top 3 suggesting their rank is invariant to the weighting schemes because they have achieved good outcomes across the jobs dimension.

Mexico exhibited the largest index variation with an index range of 5.46. This is due to an imbalance of achievements across the jobs dimension. When it came to *labour market insecurity* and *long-term unemployment rate*, the dominant indicators under the most optimistic scenario, Mexico ranked relatively high. However, in terms of the *employment rate* and *personal earnings* indicators, which receive relatively greater importance in the most pessimistic scenario, Mexico ranked near the bottom.

4.4 Education

Education is arguably one of the top priorities for many parts of the world. According to Irwin et al. (2007), *"economists now argue on the basis of the available evidence that investment in early childhood is the most powerful investment a country can make, with returns over the lifecourse many times the size of the original investment"*. The components of this dimension are *educational attainment*, *student skills*, and *years in education*. *Educational attainment* measures the proportion of 25–64 years olds with at least an upper secondary degree. *Student skills* are average scores in reading, mathematics, and science as assessed by the OECD's Programme for International Student Assessment (PISA). *Years in education* are the average years of education in which a 5 years old child can expect to enroll during their lifetime until they turn 39.

Under the best(worst)-case scenario, the weights were 0.731 (0.099) for *educational attainment*, 0.195 (0.176) for *student skills*, and 0.074 (0.725) for *years in education*. Clearly, the fact that the most optimistic view of measured education outcomes has *educational attainment* receiving nearly three-quarters of the weight is indicative of the fact that OECD countries, on average, are becoming more educated over time. In 2017, the proportion of of 25–64 years olds with at least an upper secondary degree was 74% among OECD countries.

Table 8 reports the 2017 rankings. Some drastic deviations in rankings are apparent. Russia, Czech Republic, Estonia, United States, and Slovak Republic moved up at least 10 positions in the most optimistic scenario, with Russia gaining 20 positions going from 22nd (equal-weight) to 2nd (best-case). The fact that 94.9% of 25–64 years olds in Russia had at least an upper secondary degree in 2017, the highest observed rate in the sample, goes some ways towards explaining its drastic rank improvement. However, when weight was shifted towards *years in education*, Russia ranked 30th out of the 38 countries, quite a stark contrast. Japan's *educational attainment* in 2017 at 94.4% was not too far behind Russia's. Japan had the highest average PISA score at 529 in 2017. This combined with its *educational attainment* resulted in Japan ranking as the top country in education under the best-case scenario. However, shifting weight heavily towards *years in education* (worst-case scenario) results in Japan being ranked 22nd and Australia being ranked at the top.

Finland ranked in the top 3 regardless of the scenario and furthermore due to its relatively good achievements across the components, its composite index range was only 0.66. Russia's composite index was 8.89 (2nd) under the best-case scenario but 3.75 (30th) under the worst-case scenario, a difference of 5.14 which was the highest observed in the education dimension. This has to do with the fact that Russia performs well when it comes to *educational attainment* but ranks near the bottom in terms of *years in education*.

4.5 Health

Health is an essential component of the well-being of a nation. Good health and longevity have long been a universal goal. The health dimension consists of *life expectancy*, and *self-reported health* (proportion of the population aged 15 or older who report "good" or better health).

The best(worst)-case scenario weights were 0.908 (0.150) for *life expectancy* and 0.092 (0.850) for *self-reported health*. Interestingly, the most optimistic scenario for measuring well-being in the health dimension has the objective indicator (*life expectancy*) getting substantially more of the weight than the subjective indicator (*self-reported health*). This is consistent with the fact that longevity has steadily risen in OECD countries over time (around 80 years in 2017). So it's not surprising that weighting *life expectancy* more results in more countries with better measured health outcomes. The 2017 rankings are in Table 9 which shows that the worst-case scenario rankings are highly correlated with the equal-weight ranking (Spearman rank correlation of 0.98).

Under the best-case scenario, Japan and the United States exhibited large rank deviations. Japan moved up by 21 positions going from 34th under equal-weight to 13th under best-case scenario weighting. At 83.9 years, Japan ranked at the top in terms of *life expectancy*. The United States dropped 20 positions going from 6th (equal-weight) to 26th (best-case) due to a *life expectancy* of 78.8 years (although quite high, it is one of the 13 countries out of the 38 which had a *life expectancy* below 80 years). On the other hand, for the most pessimistic view of measured health outcomes with *self-reported health* getting most of the weight, Japan ranked second last (after Korea) since just 35.4% of Japan's population aged 15 or older reported "good" or better health (second lowest after Korea which reported 32.5%). The United States, under the pessimistic scenario, moved to 3rd place thanks to 88.1% of its population aged 15 or older reporting "good" or better health. (tied with Canada which ranked at the top).

Japan and Korea also happen to exhibit the highest variations in achievement scores with index ranges of 7.19 and 7.06, respectively. Although both countries enjoy relatively high *life expectancy*, their *self-reported health* is relatively low compared to the other countries due to a smaller share of their populations reporting "good" or better health. This could reflect demographic shifts such as an increasing elderly population who are more likely to report poor health conditions (see, e.g., Organisation for Economic Co-operation and Development 2013). According to World Bank (2017), Japan's share of the population aged 65 or older has increased from 15% in 1997 to 27% in 2017. Likewise, Korea's share has increased from 6% to 14%.

4.6 Safety

How safe an individual feels living in their country has a significant impact on how they perceive their own well-being. Safety is measured by *assault rate* (prior to 2016), and *homicide rate*. We focus exclusively on the period from 2012 to 2015 which consistently used the two indicators. *Assault rate* is the proportion of the population that self-reported being assaulted or "mugged" within the past year (World Gallup Poll). *Homicide rate* measures unlawful homicides per 100,000 persons.

The best(worst)-case scenario SDE weights were 0.921(0.129) for *homicide rate*, and 0.079(0.871) for *assault rate*. In 2015, 19 out of 36 countries had a *homicide rate* of of less than 1 per 100,000 persons and little variability was observed among them. Furthermore, looking at Table 2, the mean normalized *homicide rate* indicator remains consistently over 8 (out of 10) across time. The rankings for 2015 are reported in Table 10. It is evident that considerably more countries achieve better measured safety outcomes in the best-case scenario which weights the *homicide rate* indicator heavily. In fact, 29 out of the 36 countries actually achieved an index greater than 9. Japan ranked in the top with an index of 9.99 while Belgium ranked at 29th with an index of 9.34, a separation of just 0.65, in the most optimistic scenario. Only Russia, Mexico, and Brazil had an index below 8 in the best-case scenario. Mexico and Brazil ranked among the bottom 2 countries regardless of the scenario. On the other hand, Japan ranked in the top 2 regardless of the scenario with an index range of 0.07, suggesting that it's a country that has achieved very good outcomes when it comes to safety. Re-weighting seemed to have a big impact on the United States which ranked at 31st under the best-case scenario, 16th under equal-weighting, and 4th under the worst-case scenario. The United States is one of the few countries in the sample which achieved a composite index less than 9 when *homicide rate* was weighted more under the best-case scenario, but its index improves to 9.6 under the worst-case scenario which weights assault relatively rate more (its *homicide rate* is higher than the OECD average but its *assault rate* is lower).

An even distribution of achievements is reflected by fairly low Gini coefficients of 0.08 (best-case), 0.10 (equal-weight), and 0.12 (worst-case) for the composite indices. However, the worst-case scenario ranking is substantially more consistent with the equal-weight ranking (Spearman rank correlation of 0.96).

4.7 Work-Life Balance

Balancing work and personal time is an aspect of well-being that is receiving increasingly more attention. The dimension of work-life balance includes two components: *employees working very long hours* (exceeding 50 h per week), and *time devoted to leisure and personal care*.

Under the best(worst)-case scenario, the weights were 0.948 (0.066) for *employees working very long hours*, and 0.052 (0.934) for time devoted to leisure and personal care. Table 11 reports the 2017 rankings which are demonstrably different for a few countries. In the most optimistic view of work-life balance, Latvia improved its ranking by 22 positions shifting from 27th (equal-weight) to 5th (best-case) while France dropped 20 positions going from 3rd (equal-weight) to 23rd (best-case). In the worst-case scenario which weights *time devoted to leisure and personal care*, Latvia ranked 35th out of the 38 countries while France moved right to the top of the rankings. Latvia also had the highest composite score variation at 5.42 (an index of 9.11 under the best-case scenario but 3.69 under the worst-case). Netherlands ranked as the best or second-best country regardless of the scenario and its composite score only varied by 1.01 thanks to its balanced achievements across the dimension.

Mexico and Turkey ranked at the bottom regardless of which weighting scheme we used. Re-weighting had no effect on these countries because they ranked at the bottom in terms of both indicators in the work-life balance dimension. For Mexico, 29.48% of employees reported working more than 50 h per week while for Turkey it was 33.77%. Number of hours devoted to leisure and personal care was 12.74 and 12.59 for Mexico and Turkey, respectively.

4.8 Aggregate Index

We now focus on constructing a composite index synthesizing every single indicator, something not done by the OECD. Instead, they leave it up to its website visitors to assign relative importance to each dimension. However, our concern here is not weighting the dimensions but rather all of the indicators embedded within the dimensions. von Reumont et al. (2017) provides a discussion of the perils involved in weighting dimensions rather than individual indicators contained in the dimensions. For example, one may have a preference for education over the jobs dimension, yet they may prioritize *personal earnings* (an indicator belonging to the jobs dimension) over *student skills* (an indicator belonging to the education dimension).

We consider all 24 indicators for the 2012–2015 period which has the most consistent set of indicators, in constructing an overall composite index. The weights were provided in Table 12. A total of 11 indicators are identified in the best-case scenario, that are responsible for driving improvement in composite well-being. The weights were 0.489 for *homicide rate*, 0.133 for *long-term unemployment rate*, 0.103 for *employees working very long hours*, 0.056 for *voter turnout*, 0.052 for *assault rate*, 0.049 for both *air pollution* as well as *dwellings without basic facilities*, 0.029 for *labour market insecurity*, 0.022 for *self-reported health*, 0.012 for *educational attainment*, and 0.006 for *housing expenditure*, with the remaining indicators receiving weights of zero. The fact that *homicide rate* receives nearly half of the weight can be explained by Table 2 which shows the indicator has consistently had the highest mean in most years and from Table 3 we also know that it has been fairly stable.

For the worst-case scenario, 6 indicators are identified as holding back improvements in composite measured well-being. The weights were 0.361 for *stakeholder engagement for developing regulations*, 0.337 for *household net financial wealth*, 0.229 for *voter turnout*, 0.066 for *housing expenditure*, 0.004 for *time devoted to leisure and personal care*, and 0.003 for *employment rate*, with the remaining indicators receiving weights of zero. From Table 2, we know that the *stakeholder engagement for developing regulations* along with *household net financial wealth* indicators have relatively low means. Thus, when such indicators receive relatively more importance, more countries find it harder to achieve better measured outcomes.

The 2015 rankings are reported in Table 13. In this instance, the best-case scenario rankings are more correlated with the equal-weight ranking as opposed to prior cases where the worst-case scenario was more correlated. The United States exhibited large fluctuations as it ranked 27th in the best-case scenario, 6th under equal-weight, and first in the worst-case scenario. This can be explained by the fact that the United States had the fourth highest *homicide rate* at 5.2 homicides per 100,000 persons among the 36 countries in 2015 and we know that the best-case scenario entails shifting almost half the weight towards *homicide rate*. This is in contrast with Luxembourg which had a *homicide rate* of 0.3 per 100,000 persons, moving from 13th (equal-weight) to first. However, under the worst-case scenario which weights *household net financial wealth* and *stakeholder engagement for developing regulations* relative more, the United States moved to the top position. The United States generally leads all countries in the income dimension so this goes some ways towards explaining this. Weighting stakeholder engagement for developing regulations more heavily in the worst-case scenario also resulted in Korea moving from 22nd (equal-weight) to 5th place.

Under the best-case scenario, 27 out of the 36 countries achieved an index greater than 8 (including the United States and Korea). Under the worst-case scenario, all countries, except for the United States, achieved composite indices lower than 7. Australia and Sweden rank among the top 5 countries regardless of the scenario reflecting well-balanced achievements across most aspects of the BLI.

Due to a large number of indicators, there exists a great deal of variation in composite scores between scenarios. The highest index ranges are observed for Chile (6.12), Estonia (5.99), Czech Republic (5.95), Slovak Republic (5.74), Israel (5.57), France (5.30), Germany (5.50), and Iceland (5.16). Chile and Estonia consistently ranked among the bottom 10 countries despite which weighting scheme was used, reflecting the fact that both countries had relatively low achievements across many of the indicators.

The distribution of achievements were most equal under the best-case scenario (Gini coefficient of 0.08) while the worst-case scenario showed the most disparity (Gini coefficient of 0.20).

5 Conclusion

It is increasingly being understood that traditional measures of well-being such as income per capita are insufficient measures of social welfare. In light of such criticism, there has been a growing body of research surrounding multidimensional social welfare measures. One such metric is OECD's Better Life Index (BLI) which measures well-being in 11 dimensions which can be broadly categorized as material living standards (housing, income, and jobs), and quality of life (community, education, environment, governance, health, life satisfaction, safety, and work-life balance). Components within a given dimension are aggregated into a single composite index to summarize a country's composite achievement in that dimension. The aggregation procedure uses an equal-weighting scheme (i.e., the composite index is simply taken as the arithmetic average of the dimensional indicators).

In this paper, we utilized consistent tests of stochastic dominance efficiency (SDE) to derive the best-case and worst-case scenarios for measured well-being using the BLI. The former scenario weights indicators under the most optimistic conditions such that more countries achieve better measured outcomes based on the least variable combination of components over time, while the latter scenario results in more countries achieving worse measured outcomes. This type of analysis not only allowed us to examine the sensitivity of the BLI, but it also revealed which indicators are implicitly favoured by more countries (i.e., the indicators for which the upper bounds are relatively more achievable).

We found that other than the environment and civic engagement dimensions, weighting indicators equally does not offer the most optimistic scenario, nor does it offer the most pessimistic scenario. There exist several other weighting combinations that dominate equal-weighting. The weights arising from SDE revealed which indicators are most responsible for improvements in measured well-being over time and which ones hinder improvements. We derived weighting schemes for the best-case as well as the worst-case scenario, and this allowed us to study which countries are more/less susceptible to normative judgments. That is, countries with relatively good achievements across different aspects of well-being were less sensitive to re-weighting while those with poor achievements or good achievements in some indicators but not others, were more prone to experience large fluctuations in rankings. The worst-case scenario rankings were generally found to be more correlated with the equal-weight ranking. Not only do more countries achieve better outcomes in the best-case scenario but the distribution of outcomes were generally more equal as well.

For the housing dimension, we found that attaching relatively more weight to the *dwellings without basic facilities* indicator induced the best-case scenario (i.e., more countries would find it easier to achieve better measured outcomes in the housing dimension). Shifting weight towards the other two indicators, *housing expenditure* and *rooms per person*, induced the worst-case scenario where more countries find it harder to achieve better measured outcomes. This revealed that the upper bound of the *dwellings without basic facilities* indicator is relatively more achievable for most countries as compared to the other housing indicators.

In the income dimension, the most optimistic scenario for better measured outcomes across countries entails shifting almost all the weight towards *household net adjusted disposable income*. Weighting the other indicator, *household net financial wealth*, heavily results in the most pessimistic scenario for measured outcomes in the income dimension.

The jobs dimension consists of four indicators: *labour market insecurity*, *long-term unemployment rate*, *personal earnings*, and *employment rate*. The best-case scenario for better measured job outcomes occurs when the first indicator is given majority of the weight, while the worst-case scenario occurs when the fourth indicator is given majority of the weight.

The education dimension comprises *educational attainment*, *student skills*, and *years in education*. More countries find it easier to achieve better measured outcomes in this dimension when nearly three-quarters of the weight gets attached to the first indicator. On the other hand, if we were to shift almost three-quarters of the weight towards the last indicator, more countries find it harder to achieve better measured outcomes.

In the health dimension, the best-case scenario is obtained when the vast majority of the weight is given to *life expectancy*. The worst-case scenario occurs when most of the weight is shifted to *self-reported health*. Given that longevity among OECD countries has generally been steadily rising over time, it is not surprising that more countries do well in the health dimension when *life expectancy* gets weighted more.

The safety dimension consists of *homicide rate* and *assault rate*. Relative to other indicators across the different dimensions, *homicide rate* has been reliably stable across OECD countries. The most optimistic scenario is induced when majority of the weight gets shifted to *homicide rate*. Conversely, if *assault rate* gets most of the weight, it results in the most pessimistic scenario.

The best-case scenario for the work-life balance dimension occurs when majority of the weight is given to the *employees working very long hours* indicator while the worst-case scenario occurs when most of the weight is shifted to the *time devoted to leisure and personal care* indicator.

Although the OECD does not produce an overall index that aggregates all the well-being aspects, it does let users weight the 11 dimensions which is quite different than what we did in constructing the overall index. A preference for dimension *A* over dimension *B* does not necessarily imply that they also prefer all indicators in dimension *A* over those of dimension *B*. For this reason, it might be more beneficial to allow users to weight the individual indicators rather than the dimensions. In constructing the aggregate composite indicator, we actually weighted all 24 indicators as opposed to just weighting the dimensions due to embedding effects.

We identified 11 out of 24 indicators as driving overall improvement in measured well-being with *homicide rate* receiving almost half the weight in the best-case scenario. This is an indication that the upper bound of the *homicide rate* indicator is relatively more achievable as compared to other indicators. To induce the worst-case scenario, 6 indicators were identified, with *household net financial wealth* and *stakeholder engagement for developing regulations* receiving relatively more weight. This suggests that the upper bounds of such indicators are harder to achieve for more countries.

References

Agliardi, E., Agliardi, R., Pinar, M., Stengos, T., & Topaloglou, N. (2012). A new country risk index for emerging markets: A stochastic dominance approach. *Journal of Empirical Finance*, *19*, 741–761.

Agliardi, E., Pinar, M., & Stengos, T. (2014). A sovereign risk index for the eurozone based on stochastic dominance. *Finance Research Letters*, *11*, 375–384.

Agliardi, E., Pinar, M., & Stengos, T. (2015). An environmental degradation index based on stochastic dominance. *Empirical Economics*, *48*, 439–459.

Barrett, G. F., & Donald, S. G. (2003). Consistent tests for stochastic dominance. *Econometrica*, *71*, 71–104.

Bennett, C. J., & Mitra, S. (2013). Multidimensional poverty: Measurement, estimation, and inference. *Econometric Reviews*, *32*, 57–83.

Biswas, B., & Caliendo, F. (2002). A multivariate analysis of the Human Development Index. *Indian Economic Journal*, *49*, 96–100.

Boarini, R., Comola, M., Smith, C., Manchin, R., & de Keulenaer, F. (2012). *What makes for a better life? The determinants of subjective well-being in OECD countries—Evidence from the Gallup World Poll*. OECD statistics working papers no. 1815-2031, Paris

Boarini, R., & D'Ercole, M. M. (2013). Going beyond GDP: An OECD perspective. *Fiscal Studies*, *34*, 289–314.

Bravo, H. C., & Theussl, S. (2016). *Rcplex: R interface to CPLEX*. https://CRAN.R-project.org/package=Rcplex.

Davidson, R. (2009). Reliable inference for the gini index. *Journal of Econometrics*, *50*, 30–40.

Davidson, R., & Duclos, J. Y. (2000). Statistical inference for stochastic dominance and the measurement of poverty and inequality. *Econometrica*, *68*, 1435–1464.

Decancq, K. (2017). Measuring multidimensional inequality in the OECD member countries with a distribution-sensitive Better Life Index. *Social Indicators Research*, *131*, 1057–1086.

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, *32*, 7–34.

Duclos, J. Y., Sahn, D. E., & Younger, S. D. (2006). Robust multidimensional poverty comparisons. *The Economic Journal*, *116*, 943–968.

Fleurbaey, M. (2012). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic Literature*, *47*, 1029–1075.

Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2018). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*. https://doi.org/10.1007/s11205-017-1832-9.

Irwin, L. G., Siddiqi, A., & Hertzman, C. (2007). *Early child development: A powerful equalizer*. World Health Organization. http://www.who.int/maternal_child_adolescent/documents/ecd_final_m30/en/.

Kasparian, J. (2012). OECD's 'better life index': Can any country be well ranked? *Journal of Applied Statistics*, *39*, 2223–2230.

List, C. (2004). Multidimensional welfare aggregation. *Public Choice*, *119*, 119–142.
Lorenz, J., Brauer, C., & Lorenz, D. (2017). Rank-optimal weighting or “How to be best in the OECD Better Life Index?”. *Social Indicators Research*, *134*, 75–92.
Maasoumi, E. (1999). Multidimensional approaches to welfare analysis. In J. Silber (Ed.), *Handbook of income inequality measurement*. Dordrecht: Klewer.
Markovic, M., Zdravkovic, S., Mitrovic, M., & Radojicic, A. (2016). An iterative multivariate post hoc i-distance approach in evaluating OECD Better Life Index. *Social Indicators Research*, *126*, 1–19.
Mehdi, T. (2017). Poverty comparisons with common relative poverty lines. *Communications in Statistics: Theory and Methods*, *46*, 2029–2036.
Mizobuchi, H. (2014). Measuring world better life frontier: A composite indicator for OECD Better Life Index. *Social Indicators Research*, *118*, 987–1007.
Monika, G. T. (2018). The weight of weighting: An empirical study based on the OECD Better Life Index. *The Business and Management Review*, *9*, 443–450.
Nikolaev, B. (2014). Economic freedom and quality of life: Evidence from the OECD’s Your Better Life Index. *The Journal of Private Enterprise*, *29*, 61–96.
Ogwang, T., & Abdou, A. (2003). The choice of principal variables for computing some measures of human child development. *Social Indicators Research*, *64*, 139–152.
Organisation for Economic Co-operation and Development. (2013). *How’s life? 2013: Measuring well-being*. Paris: OECD Publishing.
Organisation for Economic Co-operation and Development. (2017). *How’s life? 2017: Measuring well-being*. Paris: OECD Publishing.
Pinar, M. (2015). Measuring world governance: Revisiting the institutions hypothesis. *Empirical Economics*, *48*, 747–778.
Pinar, M., Stengos, T., & Topaloglou, N. (2013). Measuring human development: A stochastic dominance approach. *Journal of Economic Growth*, *18*, 69–108.
Pinar, M., Stengos, T., & Topaloglou, N. (2017a). Testing for the implicit weights of the dimensions of the human development index using stochastic dominance. *Economics Letters*, *161*, 38–42.
Pinar, M., Stengos, T., & Yazgan, M. E. (2015). Measuring human development in the MENA region. *Emerging Markets Finance and Trade*, *51*, 1179–1192.
Pinar, M., Stengos, T., & Yazgan, M. E. (2017b). Quantile forecast combination using stochastic dominance. *Empirical Economics*. https://doi.org/10.1007/s00181-017-1343-1.
R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Scaillet, O., & Topaloglou, N. (2010). Testing for stochastic dominance efficiency. *Journal of Business & Economic Statistics*, *28*, 169–180.
Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). *The measurement of economic performance and social progress revisited*. Reflections and overview Commission on the Measurement of Economic Performance and Social Progress, Paris
von Reumont, L., Schob, R., & Hetschko, C. (2017). *Embedding effects in the OECD Better Life Index*. Beitrage zur Jahrestagung des Vereins fur Socialpolitik 2017: Alternative Geld-und Finanzarchitekturen—Session: Political Economy II, No. E20-V1
World Bank. (2017). *World Bank open data*. Data retrieved from https://data.worldbank.org.

Paper 2C

Assessing Job Quality in Canada: A Multidimensional Approach

WEN-HAO CHEN AND TAHSIN MEHDI
Statistics Canada, Ottawa, Ontario

Dans l'examen auquel ils soumettent les aspects multidimensionnels de la qualité des emplois au Canada, les auteurs évaluent six grandes dimensions de cette qualité : les revenus et avantages sociaux, les perspectives de carrière, l'intensité du travail, la qualité de l'horaire de travail, les compétences et le degré de latitude, ainsi que le milieu social. Les résultats de l'analyse descriptive et de l'analyse à structure latente auxquelles ils procèdent révèlent l'existence d'une grande variation dans la qualité du travail selon les secteurs et les groupes sociodémographiques. Ils constatent notamment que certains des segments les plus importants du marché de l'emploi, comme celui de l'hébergement et des services personnels, sont associés à de multiples caractéristiques négatives de l'emploi. De plus, les travailleurs dont les contrats sont atypiques ou qui occupent des emplois à temps partiel subissent maints inconvénients rattachés au milieu du travail, outre celui d'être mal rémunérés.

Mots clés : bien-être, conditions de travail, qualité du travail, sécurité d'emploi

In this article, we examine multidimensional aspects of job quality in Canada. Six broad dimensions of job quality were assessed: income and benefits, career prospects, work intensity, working-time quality, skills and discretion, and social environment. Results from both descriptive and latent class analysis reveal a great deal of variation in job quality across sectors and socio-demographic groups. In particular, we found that some of the largest labour market segments, such as hospitality and personal services, are associated with many negative job features. Moreover, workers with atypical contracts or in part-time employment have many disadvantages in the workplace other than being low-paid.

Keywords: job quality, working conditions, well-being, employment security

Introduction

In the past decade, work surrounding statistical indicators for measuring individual and societal well-being has proliferated. It is now widely recognized that the measure of a country's progress encompasses more than just economic quantity or the monetary value of goods and services. Economic quality and social and environmental well-being also matter. Initiatives such as the United Nation's Happiness Index or the OECD's Better Life Index offer complementary statistics that can capture different aspects of life across populations. In this context, the focus of employment growth has also shifted from the number of jobs created to the types of jobs created.

The notion of job quality has traditionally been understood as being represented by wage level or type of employment, information about which is often accessible in most labour force surveys. However, job quality could also refer to physical working conditions, social environments, flexibility, or skills development, which affect or foster a worker's well-being. Although there is no commonly accepted definition of job quality, considerable effort has been made by international organizations to identify various dimensions of workplace quality in ways that cover its many facets (Cazes, Hijzen, and Saint-Martin 2016; OECD 2014).

The conceptual underpinnings of multidimensional job quality frameworks are closely related to the well-being discussion. There is growing international evidence that different job characteristics (such as excessive job demands, low workplace autonomy, or an unsafe work environment) have an independent influence on workers' well-being, including their physical and mental health (see OECD 2014 for a review) as well as their life satisfaction—a measure of overall quality of life (Drobnic, Beham, and Prag 2010; Garcia-Mainar, Montuenga, and Navarro-Paniagua 2015).

The focus on multidimensional job quality is pertinent to the discussion of equality and economic empowerment in the fast-changing world of work. The modern economy is becoming increasingly knowledge based, which tends to benefit those who are highly educated. As a result of an aging population, many new jobs have been created in

doi:10.3138/cpp.2018-030

personal care services. At the same time, so-called atypical employment (which includes self-employment and temporary and part-time jobs) has increased (ILO 2016). Whether this changing economy is inclusive or not can be better understood by looking at multiple aspects of job quality in addition to wages or compensation. These aspects also have important implications for gender equality; women are still more likely than men to work part time, often by choice, and to be overrepresented or underrepresented in particular industries and occupations (Moyser 2017). Variability in the quality of jobs offered in these sectors could have ramifications for an inclusive economy.

Moreover, job quality has implications for economic and labour market performance. Empirical evidence has shown that high-quality jobs improve employees' subjective well-being (Horowitz 2016; Salvatori 2010) and contribute to at-work productivity (Arends, Prinz, and Abma 2017). Better job quality also makes work more attractive and thus stimulates employment growth by encouraging inactive persons to enter the labour market and prevent early exits. However, a poor working environment is often associated with health risks, leading to quitting (Green 2010), labour market withdrawal (Park 2010; Turcotte and Schellenberg 2005), and more sickness absence (Catalina-Romero et al. 2015; Milner et al. 2015).

To date, the Canadian literature on job quality has been rather scarce, in part because of a lack of comprehensive data on workplace issues as well as a lack of consensus on which of the frameworks that have been developed to measure job quality are suitable. Some early studies paint a partial portrait of job quality, either by focussing solely on compensation-based quality using the Labour Force Survey (Tal 2016) or by combining a wide range of data sources or using small-scale surveys (Brisbois 2003; Jackson and Kumar 1998; Lowe 2007; Lowe and Schellenberg 2001; Shields 2006), but a comprehensive assessment of multiple job quality aspects using a unified data source has remained elusive. This study fills this gap by using the rich information on working conditions collected by the 2016 General Social Survey (GSS) to construct indicators of job quality situated in an international framework.

The remainder of this article is organized as follows. In the next section, we review the international framework on job quality and discuss how the 2016 GSS can be used to construct relevant indicators. We then present a portrait of multidimensional job quality in Canada across sectoral, occupational, and socio-economic groups. Finally, we use a latent class analysis (LCA) to categorize workers with similar-quality jobs and assess the relationship between the predicted job quality profiles and the observed characteristics.

Framework and Data

Various multidimensional frameworks have been proposed in the literature to assess job quality. Some focus on the attributes of the job itself, some include employment relationships, and others encompass broader labour market and social contextual information, such as provisions of social protection schemes (see Cazes et al. 2016 for a review). Depending on the framework, suggested indicators may be objective or subjective and measured at an individual or aggregate level, or both. In this study, we incorporate the framework proposed by Eurofound (2016).

One important feature of the Eurofound framework is that it is data driven: the organization used its own surveys (i.e., the European Working Conditions Survey [EWCS]) to construct the proposed job quality indicators. This ensures that all indicators are measured in a consistent manner. We use the Eurofound framework because the 2016 Canadian GSS also included several core job-quality-related questions based on the EWCS modules. The GSS data allow us to construct multiple dimensions of job quality indicators for Canada according to a well-developed international framework, with potential for comparability to European studies.

Figure 1 illustrates the seven dimensions of job quality developed by Eurofound, with minor modifications: (a) prospects, (b) work intensity, (c) working-time quality, (d) skills and discretion, (e) social environment; (f) income and benefits from employment, and (g) physical health risks. Each of these dimensions include one or more subtopics, which can be assessed through a set of questions in the survey. The first and the sixth dimensions relate to extrinsic job features, and the remaining five dimensions together measure the quality of the working environment. The latter grouping can further be divided into job demands and job resources (OECD 2014).

We constructed indicators of job quality dimensions following three guidelines (Eurofound 2016). First, all indicators were defined at the level of the job or worker on the basis of microdata, which allows us to examine job quality outcomes across socio-economic groups or sectors to address distributional issues. Second, each indicator can be categorized into either a positive or a negative job feature. Higher job demands, such as heavy workload, would be regarded negatively, and better job resources (e.g., paid training) are indicative of positive job features. Third, we did not consider purely subjective measures that involve an individual's feelings or perceptions (e.g., job satisfaction).

Job Quality Indicators: Data and Summary Statistics

Our analysis draws data primarily from the 2016 GSS. The GSS is an annual cross-sectional survey conducted by Statistics Canada since 1985, with each year covering one topic in-depth. It selects representative samples aged 15 years or older from Canada's 10 provinces. A weight was assigned to each respondent to reflect the probability of selection. The 2016 cycle focussed on the lifestyle behaviour

 doi:10.3138/cpp.2018-030

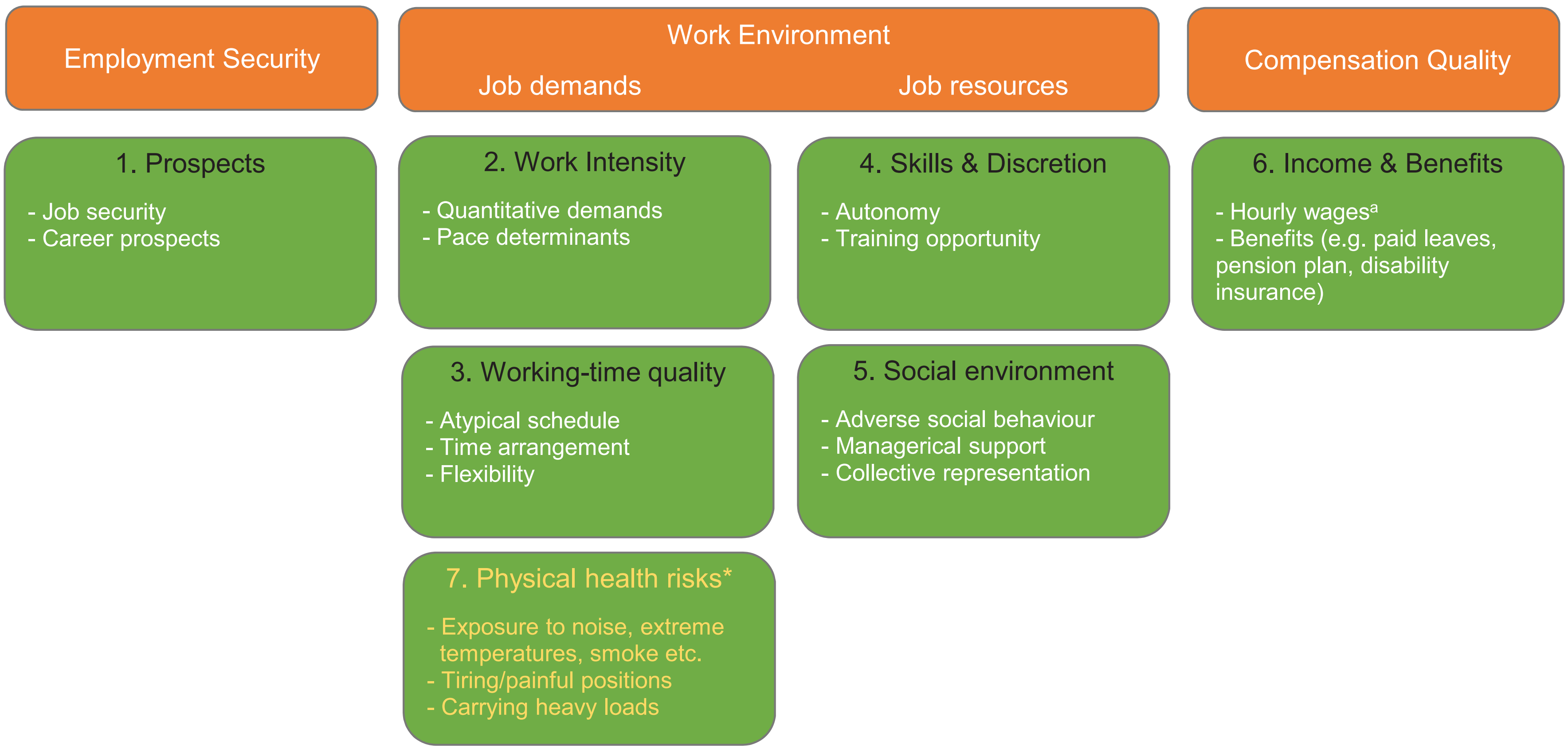

Figure 1: Job Quality Dimensions

[a] Information not available in the 2016 General Social Survey.

Source: Eurofound (2016, 37, Figure 26).

of Canadians that affects their health and well-being in the workplace and at home.

Following the described framework, we constructed 23 indicators to capture six of the seven job quality dimensions mentioned.[1] The omission is the physical environment dimension, for which information was not available. Table 1 lists a brief description of the 23 job quality indicators and their mean value. The sample was restricted to current workers (including those who were self-employed) aged 18 years and older. Survey weights were used in the entire analysis.

We used two indicators to capture the prospects dimension of the current job: the future continuity and the possibility of career progression. Only about 11 percent of workers indicated that they might lose their job in the next 6 months, and more than half said their job offered good prospects for career advancement. Note that we did not consider contract type as a prospects indicator, which is commonly seen in the European literature. This is because from a Canadian standpoint, it is difficult to associate some atypical forms of work with negative or positive prospects. Some self-employed professionals

Table 1: Job Quality Indicators

Job Quality Dimension and Sub-Topics	Indicator	P or N Indicator	Mean %[a] (weighted)
1. Prospects			
Job security	1. May lose job in the next 6 months	N	10.5
Career prospects	2. Job offers good prospects for career advancement[b]	P	51.7
2. Work intensity			
Quantitative demands	3. Workload not often manageable	N	26.2
	4. Can't often finish assigned work during regular working hours[b]	N	25.2
3. Working-time quality			
Atypical work schedule	5. Involuntary irregular (rotating, split, on-call) shift job	N	12.1
Time arrangement	6. Can choose start and end time of your work day[b]	P	41.5
Flexibility	7. Easy to take 1 or 2 hours for personal matters[b]	P	71.2
4. Skills and discretion			
Autonomy	8. Can choose the sequence of tasks	P	66.4
	9. Have opportunities to provide input into decision[b]	P	77.0
Training	10. Had formal training paid by employer[b]	P	41.5
	11. Had informal or on-the-job training[b]	P	56.1
5. Social environment			
Adverse social behaviour	12. Experienced verbal, sexual, or physical violence at work	N	15.1
Managerial support	13. Received support from managers[b]	P	63.8
	14. Had a formal job performance assessment[b]	P	58.5
Collective representation	15. Covered by a union contract or collective agreement[b]	P	32.1
6. Income and benefits			
Hourly wage (from LFS), $	16. Mean hourly earnings	P	26.8
Employment benefits	17. Workplace pension plan[b]	P	39.0
	18. Paid sick leave[b]	P	42.4
	19. Paid vacation leave[b]	P	56.1
	20. Disability insurance[b]	P	42.3
	21. Supplemental medical/dental care[b]	P	46.6
	22. Worker's compensation[b]	P	49.6
	23. Maternity, parental, or layoff benefits	P	42.4

Notes: Indicators were constructed from 2016 GSS questions, except for hourly wage, which was derived from the monthly LFS data (March 2016). The sample was restricted to workers (including those who were self-employed) aged 18 y and older. The number of overall observations is 10,680. Actual sample size, however, varies across indicators because not every person in the sample answered all the job quality questions. GSS = General Social Survey; LFS = Labour Force Survey; N = negative; P = positive.

[a] Unless otherwise indicated.

[b] Excluding the self-employed.

Source: Statistics Canada, 2016 GSS and 2016 LFS data.

and entrepreneurs indeed improve their job security and career prospects as they gain more experience. Similarly, fixed-term jobs are frequently renewed, whereas indefinite contracts can easily be terminated. This is echoed by OECD (2013), in which Canada ranked low among OECD countries in terms of the protection of permanent workers against dismissal.

Work intensity is captured by two negative indicators of job quality: workload not often manageable and often cannot finish work during regular hours. The former gives a broad sense of time pressure from work, and the latter reveals a certain degree of constant pressure on a regular basis. Using these two indicators, we found that about a quarter of workers reported high work intensity.

The assessment of working-time quality includes three indicators: involuntary atypical work schedule, time arrangement, and flexibility regarding working time. About 12 percent of workers reported an irregular work schedule. This is considered a negative job feature because of its involuntary nature. For control over time arrangement, nearly 42 percent of workers reported the ability to choose the start or end time of the work day, and 71 percent indicated they had the flexibility to take some hours off for personal matters.

The dimension of skills and discretion refers to the job aspect that allows workers to apply their skills with some degree of autonomy over their tasks and resources, as well as the training opportunity to develop the skills required for the job. Four indicators were used. For autonomy, we measured the ability to choose the sequence of tasks and the opportunity to provide input into decisions that affect work. Overall, about two-thirds to three-quarters of workers reported a high degree of autonomy in their job. For training opportunities, about 42 percent of workers had received formal training in the past 12 months and about 56 percent had received informal training.

We assessed social environment in the workplace by means of indicators covering three elements: adverse social behaviour, managerial support, and collective representation. About 15 percent of workers reported abusive experiences at work, such as verbal abuse, sexual harassment, threats, humiliation, or physical violence. The extent of managerial support was measured with two indicators: received support from manager and had a formal job performance assessment. The former refers to support from an immediate supervisor, and the latter refers to a management system that enables lines of communication, recognition, and identification of areas of improvement. Overall, about 60 percent of employees received some type of managerial support in their job. Another indicator of social environment was captured by collective representation. In 2016, fewer than one-third of workers were covered by a union contract or collective agreement.

The dimension of income and benefits includes hourly wages[2] and seven employment benefits (workplace pension plan, paid sick leave, paid vacation leave, disability insurance, supplemental medical or dental care, workers' compensation, and parental or layoff benefits). These seven employment benefits are employer-provided benefits in addition to provisions mandated by government laws (e.g., provincial regulations on vacation leave). It is noteworthy that the Eurofound framework did not include employment benefits because the EWCS surveys did not collect such information. However, they acknowledged that those benefits, such as a pension plan, may be considered "suspended" earnings and are therefore important in capturing the job's earnings quality (Eurofound 2016, 38). Overall, about 40 percent or more of paid workers stated that their job included at least one of the employment benefits mentioned.

Do the proposed positive and negative nature of various job quality indicators correspond to workers' subjective life experience in general? For example, one can argue that having flexible start and end times should not be considered a positive job feature if persons in such jobs tend to work very long hours because no one is telling them exactly what time their work starts and ends. In that case, those with flexible work hours may not feel better than those without such flexibility. To investigate this, Appendix Table A.1 relates all job quality indicators in the GSS to the life and job satisfaction questions and confirms that persons with positive job attributes do express greater life and job satisfaction than those without.

Finally, although we closely follow Eurofound's job quality framework, we acknowledge some limitations. In particular, the Eurofound framework excludes certain aspects of job quality and does not account for the interrelationships between the different quality measures at the level of individual workers (OECD 2014). For example, the role of institutions (e.g., social benefits, employment protection) in providing workers with better labour market security was not included. Also, their approach does not measure some combined effects, such as job strain, that capture the imbalance between job demands and resources.

Portrait of job quality in Canada

In this section, we present a portrait of job quality in Canada using the multidimensional indicators defined earlier. We examine the distribution of job quality by sector and occupation and by socio-demographic group. This makes it possible to assess whether jobs associated with the larger or fastest growing sectors and professions are of higher or lower quality and also to assess whether workers from different demographic backgrounds have equal representation in higher quality jobs.

How Do Sectors and Occupations Compare?

Table 2 presents the six job quality dimensions for nine industrial and eight occupational groups.[3] The top row

doi:10.3138/cpp.2018-030

Table 2: Job Quality Dimensions by Sector and Occupation

Sector and Occupation	GSS Weighted Share (%)	Employment Growth 2006–2016[a]	1. Prospects		2. Work Intensity		3. Working-Time Quality			4. Skills & Discretion				5. Social Environment				6. Income & Benefits							
			Might Lose Job in Next 6 Mo	Job Offers Good Advancement Prospects†	Workload not Often Manageable	Can't Often Finish Work in Regular Hours[b]	Involuntary Irregular Schedule	Flexible Start& End Hours[b]	Can Take Time Off for Personal Reasons[b]	Can Decide Sequence of Tasks	Have Opportunities to Provide Input[b]	Paid Formal Training[b]	Informal Training[b]	Verbal or Sexual, Threats, Humiliation, or Physical Abuse	Support from Managers[b]	Had Formal Job Performance Evaluation†	Have Formal Employee Representation Body[b]	Workplace Pension Plan†	Paid Sick Leave[b]	Paid Vacation Leave[b]	Disability Insurance[b]	Supplemental Medical & Dental Care[b]	Workers' Compensation[b]	Maternity, Parental, or Layoff Benefits[b]	Mean Hourly Earnings, %[a]
Overall mean	100.0	11.6	10.5	51.7	26.2	25.2	12.1	41.5	71.2	66.4	77.0	41.5	56.1	15.1	63.8	58.5	32.1	39.0	42.4	56.1	42.3	46.6	49.6	42.4	26.8
Deviations from the mean																									
Industry																									
Trade or transportation	18.7	8.1	−2.6	0.1	−2.8	−6.6	6.9	−1.3	−1.5	−2.1	−2.6	−5.9	−4.3	0.2	0.5	−3.9	−11.9	−3.6	−3.2	1.5	−2.2	−0.1	1.9	−1.9	−4.8
Finance or professional	17.6	15.8	0.3	8.1	1.8	6.1	−5.7	17.6	13.3	5.9	5.9	−0.3	6.9	−5.1	2.6	12.1	−22.5	−2.5	3.2	−1.1	1.0	1.2	−9.5	−1.1	5.0
Hospitality	15.9	18.5	0.9	−4.9	−1.6	−4.5	2.8	2.2	−6.5	−1.9	−4.0	−7.7	−12.0	−0.1	0.0	−15.5	−16.1	−17.6	−12.4	−4.8	−13.5	−14.9	−4.6	−9.2	−8.0
Health care	13.4	34.5	−2.5	−2.9	3.9	−1.1	5.1	−10.2	−12.8	−2.0	−2.7	7.2	2.9	9.9	−9.1	4.5	24.5	8.1	7.4	4.6	0.9	2.7	3.2	6.3	0.0
Manufacturing	8.8	−19.0	2.9	−7.6	−1.1	−4.0	−2.6	−7.6	8.6	−2.6	−2.5	−9.6	−7.9	−3.9	−1.6	−6.1	−10.0	1.8	−5.4	6.3	6.1	5.5	3.3	−1.5	−1.2
Education	7.9	9.9	−2.9	−2.1	4.7	17.8	−7.1	−13.3	−12.9	5.5	3.0	7.5	11.5	1.5	−2.0	10.1	41.0	12.3	12.3	−8.7	2.9	v2.0	−3.8	4.2	6.7
Public administration	6.5	9.3	−3.1	3.6	2.9	6.4	−3.3	9.4	8.7	−1.1	−3.0	18.0	15.3	3.3	5.1	24.0	38.0	17.3	13.3	0.0	9.5	4.9	0.9	7.9	8.6
Construction	6.4	31.2	8.2	5.7	−4.3	−5.4	−5.6	−6.8	6.5	−0.3	8.6	−1.8	−7.6	−5.0	7.5	−27.6	−6.3	−4.7	−15.0	−3.9	−0.5	2.7	12.8	−2.5	2.5
Primary	4.7	−0.1	5.4	2.8	−5.0	−3.9	−1.1	0.5	1.5	−3.9	6.3	9.5	4.7	−2.6	5.5	0.5	−2.2	8.2	6.5	7.2	14.2	10.6	14.5	9.4	6.5
White-collar occupation																									
Professionals	23.3	32.1	−1.4	2.9	5.1	12.2	−5.0	12.0	1.2	6.4	4.4	5.9	11.2	1.1	−2.1	14.3	12.5	4.3	8.2	−4.5	3.6	2.8	−5.1	3.9	8.7
Clerks	12.2	2.5	0.3	−3.9	−1.7	−1.3	−6.7	4.5	12.6	4.4	−2.1	−6.3	−2.1	−1.3	2.5	2.1	−0.6	1.2	2.1	0.2	0.6	0.7	−3.7	−0.5	−3.2
Technicians	11.0	21.5	0.5	−1.4	−0.2	0.5	3.8	−8.4	−4.7	−2.3	0.4	7.4	3.8	−0.7	0.5	5.8	4.8	3.1	5.5	0.5	0.3	−0.7	−1.4	−0.7	2.4
Managers	10.1	−10.1	−1.9	11.1	6.8	19.4	−5.3	27.6	15.9	10.0	12.7	3.7	2.5	−3.0	0.6	12.2	−25.2	−2.7	5.4	2.2	4.6	5.1	−0.9	2.7	14.0
Blue-collar occupation																									
Machine operators	12.3	6.0	3.8	3.4	−5.0	−8.1	−1.8	−15.7	0.0	−9.1	0.1	−0.3	−10.3	−2.5	3.5	−17.1	4.2	4.2	−7.1	6.2	8.4	7.5	16.5	1.9	−1.6
Elementary	6.1	−4.6	7.5	−6.1	−5.4	−12.5	3.6	−21.8	−2.4	−11.2	0.6	−6.3	−8.1	−0.3	1.6	−10.1	−7.3	−0.6	−10.8	4.1	1.4	−2.1	3.9	−0.5	−2.7
Sales	9.7	14.4	−1.8	−2.6	−2.5	−8.5	13.0	−0.6	−4.2	1.9	−4.6	−7.9	1.5	−0.2	−0.5	−6.1	−18.6	−10.5	−8.4	−2.3	−10.1	−7.5	−4.6	−6.2	−8.0
Personal services	15.2	19.4	−2.2	−4.2	−3.1	−10.1	5.5	−7.9	−13.9	−7.9	−8.8	−2.4	−7.6	4.2	−2.7	−11.3	3.5	−4.4	−5.0	−0.5	−9.2	−7.0	0.4	−3.5	−8.3

Job quality indicator (significance level of difference between group mean and overall mean)

- Much better than average, at 1% level
- Better than average, between 1% and 10% level
- Within the average, not significant at 10% level
- Worse than average, between 1% and 10% level
- Much worse than average, at 1% level

Note: The number of overall observations is 10,680. Actual sample size, however, varies across indicators. GSS = General Social Survey; LFS = Labour Force Survey.

[a] Data from LFS.

[b] Excluding self-employed individuals.

Source: Statistics Canada, 2016 GSS and 2006–2016 LFS.

doi:10.3138/cpp.2018-030

shows the overall mean for each of the 23 job quality indicators, and the remainder of the table reports deviations from the respective means. Cells are presented in five grey scales to indicate the significance level of their deviation values. A white (black) colour indicates significantly better (worse) than average job quality, a light (dark) colour indicates better (worse) than average job quality, and a medium colour indicates average job quality.[4]

Overall, job quality seems to be better than average in public administration, primary, and finance and professional service sectors; average in education, trade, transportation, and construction; and worse than average in health care, manufacturing, and hospitality. Public administration jobs scored relatively high in nearly all quality dimensions, particularly in working-time quality, training, social environment, and income and benefits. For example, nearly 60 percent of public administration workers had paid formal training over the previous year, compared with only 42 percent on average overall. The financial and professional sector also did well in the areas of prospects, flexibility, and autonomy, whereas workers in the primary sector (agriculture, fishing, and oil extraction) were slightly above average across the board.

However, some of the larger (and fastest growing) sectors, such as health care, hospitality (accommodation and food services), and construction did poorly in multiple job quality dimensions. Health care, which saw a 35 percent growth in employment between 2006 and 2016 and accounted for 13.4 percent of the workforce in 2016, exhibited low scores in working-time quality and high incidence of workplace violence. This is consistent with the literature on workplace aggression in health care professions (Chappell and Di Martino 2006; Shields and Wilkins 2009). Similarly, hospitality (16 percent of workforce) performed poorly in nearly all job quality dimensions, especially training opportunity, social environment, and income and benefits.

Cross-sectoral differences in job quality are more evident in the dimensions of working-time quality, training, social environment, and benefits. For example, 59 percent of finance and professional sector workers enjoyed flexible starting and end hours, compared with only 28 percent in education and 31 percent in health care. Nearly 56 percent of public administration employees had a retirement pension plan, whereas only 21 percent of hospitality workers did.

There is also marked heterogeneity across job quality dimensions within sectors. In education, for instance, some job features are very favourable (skills and discretion and social environment), and some are very disadvantageous (work intensity and inflexibility). Similar patterns are also evident in health care and, to a lesser extent, in construction.

Such diverse sectoral patterns may reflect very different job types within a broadly defined sector. Table 2 also reports how job quality is distributed across occupations. Overall, differences in job quality are more visible along the blue-collar–white-collar line. Office-based jobs generally exhibit more desirable job features in three or more of the quality dimensions, whereas jobs that require manual labour or customer interaction do poorly in nearly all areas (the only exception is work intensity).

It is of interest to relate these findings to the job polarization literature, which argues that employment growth has been polarized into both high-skill professional jobs and low-skill service-related jobs, with a hollowing out of the middle over the past few decades (Autor and Dorn 2013; Goos, Manning, and Salomons 2009), although the pattern seems to have stalled after 2000 in Canada (Green and Sand 2015). Our findings of marked job quality gaps between professional and low-end service jobs imply that if the pattern of job polarization continues, it could lead to an increasingly divided labour market in which jobs are associated with either many good quality features or many poor features, not somewhere in between.

How Do Workers Compare?

The population groups more likely to get lower quality jobs were youth, less educated persons, and persons with disabilities (Table 3). Although the literature has shown that these groups are often associated with low pay, their poor outcomes along multiple job quality dimensions make them even more vulnerable. Workers with a high school diploma or less, for instance, were more likely to work in jobs with less flexible work schedules, low autonomy, and lack of training opportunities and employment benefits, in addition to low pay. They were also less likely to be recognized for their work efforts, given the low access to performance evaluation: only 47 percent of less-educated workers reported having a formal job performance assessment over the past year, compared with nearly 70 percent of university-educated workers. Marked differences in nearly all job quality dimensions along the education line indicate a wider gap between skilled and less-skilled workers.

Table 3 also indicates that the concerns relating to the youth labour market pertain to more than just unemployment, as often emphasized. Younger workers are more likely to be in a job with an irregular work schedule, without formal performance assessments, and with limited or no employment benefits. This may reflect the fact that fewer employed youth today are in full-time jobs (Morissette 2016). Nevertheless, young workers still did relatively well in terms of career prospects, manageable workload, and access to informal training.

Visible minorities generally faced no significant disadvantages in job quality compared with non-visible minorities. By contrast, persons with disabilities fared much worse in all job quality dimensions. Their workplace disadvantages include not only lower income benefits

Table 3: Job Quality Dimensions by Selected Socio-Demographic Group

Socio-Demographic group	% Weighted	1. Prospects		2. Work Intensity		3. Working-Time Quality			4. Skills & Discretion				5. Social Environment				6. Income & Benefits							
		Might Lose Job in Next 6 Mo	Job Offers Good Advancement Prospects[a]	Workload Not Often Manageable	Can't Often Finish Work in Regular Hours[a]	Involuntary Irregular Schedule	Flexible Start & End Hours[a]	Can Take Time Off for Personal Reasons[a]	Can Decide Sequence of Tasks	Have Opportunities to Provide Input[a]	Paid Formal Training[a]	Informal Training[a]	Verbal or Sexual Threats, Humiliation, or Physical Abuse	Support from Managers[a]	Had Formal Job Performance Evaluation[a]	Have Formal Employee Representation Body	Workplace Pension Plan[a]	Paid Sick Leave[a]	Paid Vacation Leave[a]	Disability Insurance[a]	Supplemental Medical & Dental Care[a]	Workers' Compensation[a]	Maternity, Parental, or Layoff Benefits[a]	Mean Hourly Earnings, $[b]
Overall mean	100.0	10.5	51.7	26.2	25.2	12.1	41.5	71.2	66.4	77.0	41.5	56.1	15.1	63.8	58.5	32.1	39.0	42.4	56.1	42.3	46.6	49.6	42.4	26.8
		Deviations from the Mean[c]																						
Sex																								
Men	52.9	1.0	2.5	−1.2	−1.0	−0.3	1.4	2.4	−1.0	1.5	1.5	−0.6	−2.7	2.1	−0.6	−2.2	1.5	−0.4	1.8	3.1	3.5	3.9	0.2	1.9
Women	47.1	−1.2	−2.7	1.3	1.1	0.3	−1.5	−2.6	1.2	−1.7	−1.6	0.7	3.1	−2.3	0.6	2.3	−1.7	0.4	−1.9	−3.4	−3.8	−4.3	−0.3	−1.9
(Women–men gap)		−2.2	−5.2	2.5	2.1	0.6	−2.9	−5.0	2.2	−3.2	−3.1	1.3	5.8	−4.4	1.2	4.5	−3.2	0.8	−3.7	−6.4	−7.3	−8.2	−0.5	−3.8
Age group																								
18–29	22.5	1.4	5.6	−4.0	−7.7	6.3	−3.2	−7.0	−6.2	−1.5	0.9	5.7	2.2	2.0	−17.0	−8.8	−7.9	−7.2	−2.8	−10.5	−9.2	−4.0	−6.0	−6.0
30–44	33.6	−0.5	3.4	2.4	3.5	−0.8	1.3	0.8	0.8	2.4	3.4	1.7	1.5	−0.2	5.6	1.2	2.9	3.4	1.9	4.6	4.7	1.3	4.7	1.8
45–59	33.9	−0.6	−5.1	2.2	2.6	−2.3	0.9	2.9	2.0	−1.5	−1.6	−3.0	−1.2	−0.7	6.5	5.1	3.3	2.3	0.1	3.9	2.6	1.4	1.0	2.0
≥60	10.1	0.6	−10.8	−7.0	−2.8	−3.5	0.5	5.6	4.3	0.3	−10.8	−12.6	−6.1	−2.3	0.9	0.8	−2.4	−2.4	0.0	−4.1	−3.2	0.6	−6.5	−0.7
Education																								
High school or less	29.8	0.3	−1.2	−3.1	−6.7	2.7	−5.3	−2.8	−4.1	−3.4	−7.0	−7.2	−2.3	2.9	−11.5	−6.6	−5.6	−9.0	−1.9	−5.7	−5.7	0.9	−5.3	−5.8
Some post-secondary	36.2	0.2	−1.3	−0.1	−2.9	1.0	−4.7	−0.2	−1.5	−0.2	0.7	−2.7	0.4	−2.2	0.5	2.3	2.2	2.6	5.1	3.3	3.8	4.0	3.4	−1.4
University degree	33.9	−0.6	2.4	2.8	9.2	−3.5	9.8	2.6	5.0	2.9	5.4	9.5	1.6	−0.5	9.7	3.6	2.3	4.5	−4.2	1.1	0.6	−5.3	0.9	6.7
Visible minority																								
Yes	21.9	0.0	7.3	0.0	−3.4	2.0	5.2	−2.0	−0.3	−0.5	−2.8	−3.6	−1.4	−1.3	−1.5	−4.3	−1.3	0.9	−0.7	−4.3	−1.2	−6.1	−2.6	n.a.
No	78.1	−0.1	−2.2	−0.1	0.9	−0.6	−1.5	0.5	0.2	0.1	0.9	1.1	0.4	0.5	0.4	1.3	0.7	−0.1	0.5	1.4	0.6	2.0	0.9	n.a.
Persons with disability																								
Yes	7.5	2.6	−8.3	5.7	8.1	3.6	−2.4	−9.1	−5.3	−12.9	−4.9	4.5	7.3	−9.2	−1.1	−1.3	−5.7	−5.6	−4.8	−3.7	−3.9	−1.2	−6.0	n.a.
No	92.5	−0.2	0.6	−0.4	−0.6	−0.3	0.2	0.7	0.4	1.0	0.4	−0.4	−0.6	0.7	0.1	0.1	0.4	0.4	0.4	0.3	0.3	0.1	0.5	n.a.

Job quality indicator (significance level of difference between group mean and overall mean)

- Much better than average, at 1% level
- Better than average, between 1% and 10% level
- Within the average, not significant at 10% level
- Worse than average, between 1% and 10% level
- Much worse than average, at 1% level

Note: The number of overall observations is 10,680. Actual sample size, however, varies across indicators. LFS = Labour Force Survey.

[a] Excluding self-employed individuals.

[b] Data from LFS.

[c] Unless otherwise indicated.

Source: Statistics Canada, 2016 GSS and 2016 LFS.

doi:10.3138/cpp.2018-030

and higher incidence of workplace violence but also low advancement prospects, high time pressure, lack of control over tasks, and limited support from managers. These findings suggest that discussions aimed at extending working lives to cope with fiscal or macroeconomic pressures need to take into consideration the impacts of health status on access to quality jobs to ensure a healthy labour force.

Table 3 provides a new perspective for assessing gender equality in the workplace. Overall, we found significant gender differences—at the 1 percent level—in 10 of 23 job quality indicators. For example, female workers earned less (about $3.80, or 13 percent lower) than their male counterparts in hourly wages. This is consistent with the literature. Other noticeable disadvantages faced by female workers include higher incidence of workplace violence, poor career prospects, less working-time flexibility, and limited access to certain employee benefits (such as disability and supplemental medical insurance). However, the data show small or little gender gap in job features relating to work intensity, skills and discretion, performance evaluation, paid sick leave, and parental or layoff benefits.

Identifying groups of workers with similar-quality jobs: A Latent Class Analysis

In this section, we examine which workers are more likely to have jobs associated with multiple good or bad job features and what the potential drivers for these outcomes are. The descriptive analysis we have presented seems to suggest that workers can be assigned to a small number of groups, each with broadly similar job quality outcomes. Unfortunately, we do not have a single variable that identifies the groups. One common approach is to construct a composite index from an aggregation of all indicators and use it to classify workers into different job quality groups. However, there are often technical issues (weighting, in particular) involved in aggregating indicators. Without constructing an overall index, we identify the unobserved groups by applying a technique known as Latent Class Analysis (LCA) to cluster workers into similar classes according to their responses to job quality indicators.

LCA, or finite mixture modelling, is a statistical procedure for identifying unmeasured class membership probabilities among subjects, using their responses to a set of observed variables (Vermunt and Magidson 2002). In the context of this study, observed variables are the six job quality dimensions, each of which contains a number of indicators, as mentioned earlier. For simplicity, we created a categorical variable with three responses for each of the job quality dimensions; the item response was coded as "1" if workers experienced none or very few good job quality features in the dimension in question, as "2" if they experienced partial good job features, and as "3" if they experienced all or most good job features.[5]

Table 4 shows the distribution of the six job quality dimensions. The sample was restricted to individuals who provided valid responses to all job quality questions in the GSS. This excluded all self-employed workers because they were exempt from answering certain questions. The analysis included about 8,000 workers in the analysis. Overall, 80–90 percent of workers reported having at least partial good job features (Categories 2 or 3) in five of the six job quality dimensions. The only exception is benefits, for which 43 percent of workers indicated that their jobs provided none or only one of the seven employment benefits

Table 4: Distribution of Categorical Response of Job Features by Job Quality Dimension

Item-Response Category (N = 8,004)	Dimension					
	1. Prospects	2. Work Intensity	3. Working-Time Quality	4. Skills and Discretion	5. Social Environment	6. Benefits
No. of indicators used within the dimension	2	2	3	4	4	7
1 (none or very few good job quality features)	5.9	3.1	8.4	19.8	14.3	43.4
2 (some good job quality features)	47.6	8.7	60.3	32.0	37.7	22.2
3 (all or most good job quality features)	46.6	88.3	31.3	48.3	48.1	34.4
Total	100.0	100.0	100.0	100.0	100.0	100.0

Notes: For Dimensions 1 and 2, Categories 1, 2, and 3 refer to having zero, one, and two good features, respectively. For Dimension 3, Categories 1, 2, and 3 refer to having zero, one to two, and three good features, respectively. For Dimensions 4 and 5, Categories 1, 2, and 3 refer to having zero to one, two, and three to four good features, respectively. For Dimension 6, Categories 1, 2, and 3 refer to having zero to one, two to five, and six to seven good features, respectively.

Source: Statistics Canada, 2016 General Social Survey.

included (Category 1). The other stand-out dimension is work intensity, for which more than 88 percent of workers reported being in the best category (i.e., their jobs were not associated with the two negative indicators listed).

Using these categorical variables, which indicate whether workers experienced none, partial, or all good job features in each of the six job quality dimensions, we estimated a latent class model with *n* unobserved job quality profiles with and without the inclusion of covariates. The model was fitted using a STATA Plugin developed by Lanza et al. (2015). When no covariates were included, we estimated two sets of parameters: (a) probabilities of latent class membership and (b) item-response probabilities that express the correspondence between the observed six job quality dimensions and the latent classes. When covariates are included, the probabilities of latent class membership are predicted as functions of regression coefficients for covariates and the values of the covariates (Lanza et al. 2015).

The optimal number of latent classes, *n,* can be determined by selecting the model that results in the lowest Bayesian information criterion statistic. Table A.2 in the Appendix fitted a baseline model with each additional class up to the six-class model. An inspection of the Bayesian information criterion values indicates that the four-class model is optimal because the addition of classes beyond four results in no improvement. After selecting the number of latent classes, Table 5 reports the estimated class membership probabilities (top half) as well as the item-response probabilities of the six job quality dimensions (bottom half) for each class without covariates. More alternative model specifications and sensitivity analysis can be found in Chen and Mehdi (2018).

Four very distinct job quality profiles were identified from the LCA. About 37 percent of workers were predicted to be in the best job quality group (Class 1), judging by the probabilities of having many or all good job features in all six quality dimensions. Conversely, 23 percent of workers were predicted to be in the worst job quality group (Class 4), with very low probabilities of experiencing many good job features in most dimensions. The

Table 5: Four-Class Latent Class Analysis Model Estimates**:** Class Membership and Item-Response Probabilities

Job Quality Dimensions by Response Category	Jobs with			
	High Overall Quality: Class 1	Decent Work Benefits & Social Environment, Poor Working-Time and Skills: Class 2	Manageable Work Demands, Poor Benefits & Social Environment: Class 3	Poor Overall Quality: Class 4
Class membership probability, sample *M* (*SE*)	0.368 (0.044)	0.122 (0.053)	0.278 (0.048)	0.232 (0.033)
Item 1: Probability of having none or few good job features				
Prospects	0.013	0.044	0.039	0.167
Work intensity	0.020	0.078	0.002	0.056
Working-time quality	0.067	0.109	0.053	0.136
Skills & training	0.008	0.447	0.083	0.525
Social environment	0.016	0.081	0.135	0.388
Benefits	0.325	0.085	0.479	0.714
Item 2: Probability of having partial good job features				
Prospects	0.311	0.529	0.478	0.710
Work intensity	0.088	0.149	0.025	0.126
Working-time quality	0.509	0.751	0.588	0.703
Skills & training	0.198	0.413	0.493	0.268
Social environment	0.215	0.395	0.556	0.420
Benefits	0.097	0.113	0.451	0.212
Item 3: Probability of having many or all good job features				
Prospects	0.676	0.427	0.482	0.123
Work intensity	0.892	0.773	0.973	0.818
Working-time quality	0.424	0.140	0.359	0.161
Skills & training	0.793	0.140	0.424	0.207
Social environment	0.769	0.525	0.309	0.192
Benefits	0.578	0.801	0.070	0.074

Note: Sample includes 8,001 respondents. The latent class analysis model was estimated without the covariates.

Source: Statistics Canada, 2016 General Social Survey.

remaining two classes were considered to have average jobs but still differed substantially from each other in some aspects. Workers in Class 2 (12 percent) in general enjoyed good employment benefits and social environment but tended to have lower quality in terms of working time as well as skills and training. The probability of having a positive response to all three indicators in the working-time dimension was rather small, indicating a lack of work–life balance for this class. Finally, the remaining 28 percent (Class 3) did well in work intensity and flexibility but less so in benefits and social environment. Very few workers in this class had jobs that offered six or all seven employment benefits listed.

Latent Class Job Profiles by Socio-Economic Characteristic

Table 6 reports the proportions of workers in each job quality class by selected personal and job characteristics, based on each worker's posterior probabilities derived from the model in the preceding section. This would allow us to explore whether some variables of interests could affect the probability of class membership and, consequently, may be considered for inclusion in the LCA regression as predictors.

Youth, less-educated people, and those in blue-collar occupations were underrepresented in the best job quality group. This is consistent with the broad patterns in the "Portrait of Job Quality in Canada" section. For instance, only about 29 percent of workers with a high school diploma or less were in Class 1, whereas 46 percent of university graduates were in that class. At the same time, young or less-educated workers were more likely than their counterparts to be in the worst job class. Nevertheless, they were overrepresented in jobs that offered poor social environment and fewer employment benefits (Class 3). A great deal of variation exists in job quality profiles across sectors. Nearly 60 percent of workers in public administration, 49 percent of those in education, and 45 percent of those in finance and professional service sectors were clustered into high overall quality jobs, whereas at least one-quarter to one-third of workers in hospitality, manufacturing, and trade or transportation were associated with the worst job quality class.

Table 6 also reveals a strong association between job quality class and contract type, part-time and full-time status, and workplace size. Concerning the effects of employment contract on job quality, there has been much debate about whether the growing volume of non-standard contract jobs are associated with precarious work (Galarneau 2010; OECD 2015). Table 6 indicates that workers in atypical forms of employment—which includes seasonal, fixed-term, and casual workers—were overrepresented in the worst job quality class (36 percent) compared with 21 percent for regular contract holders. At the same time, the vulnerability of atypical workers was further reinforced by their limited representation in the best class (24 percent) versus regular contract workers (40 percent).

The other noteworthy contrast in job quality is along the line of hours of work. Part-time workers were much more likely than their full-time counterparts to be in poor job categories and less likely to be in good job categories. The findings add to the part-time penalty literature, which has focussed primarily on hourly wages (Bardasi and Gornick 2008; OECD 2015; O'Dorchai, Plasman, and Rycx 2007), by suggesting that part-time penalty can also be reflected in a wider range of job quality indicators, including prospects, working-time quality, skills, and social environment.

Marked differences are also seen for firm size: the number of employees at a workplace is positively associated with job quality. Among workers in large firms (with 500 or more employees), about 50 percent were clustered into the best-quality class and only 17 percent were clustered in the worst job class. The comparable figures for workers in small firms with fewer than 20 employees were 26 percent and 29 percent, respectively. Not surprisingly, small firm workers were overrepresented in jobs that offer fewer employment benefits and a less desirable social environment (Class 3). This suggests that the demand-side factor also plays a role in shaping the distributions of job quality profiles.

Job Quality Profiles for Groups at Risk of Non-Standard Employment

Table 6 shows that non-standard work (NSW) arrangements (i.e., part-time and atypical contracts) seem to be strong predictors of the latent class membership probability. For some groups (women, youth, older adults, and less educated persons, in particular), the chances of working part time or having an atypical job were higher than for the other groups. This was either by choice or the result of a lack of access to standard employment. Here, we examine whether being in these groups predicts job quality profiles. We assess this by including atypical employment and selected socio-demographic variables (and their interaction) in the LCA regression as predictors of the conditional latent class membership.

We computed the impact of a specific covariate on latent class membership probability through the estimated coefficients from the corresponding LCA regression (Appendix Tables A.3 and A.4). For example, the β coefficient on the part-time variable in Class 4 is 1.79 (or 5.97 in relative risk ratio). It reflects, for a part-time worker, that the odds of belonging to Class 4 (relative to reference Class 1) are nearly six times larger than the odds for a full-time worker.

Job Quality Profiles among Frequent Part Timers

In 2016, about one-third of young Canadian workers were in part-time jobs, whereas this was the case for one-tenth of prime-age workers. High rates of part-time work were

doi:10.3138/cpp.2018-030

Table 6: Job Quality Profiles by Selected Characteristics (Proportion of Workers in Each Class)

Characteristic	Jobs with			
	High Overall Quality: Class 1	Decent Work Benefits & Social Environment, Poor Working-Time and Skills: Class 2	Manageable Work Demands, Poor Benefits & Social Environment: Class 3	Poor Overall Quality: Class 4
Overall	0.368	0.122	0.278	0.232
Men	0.383	0.115	0.276	0.226
Women	0.373	0.109	0.279	0.239
Age group, y				
18–29	0.313	0.083	0.339	0.265
30–44	0.411	0.117	0.270	0.202
45–59	0.401	0.127	0.240	0.233
≥ 60	0.337	0.120	0.282	0.261
Education				
High school or less	0.292	0.109	0.325	0.274
Some post-secondary	0.369	0.130	0.269	0.232
University	0.462	0.095	0.246	0.197
Employment				
Full time	0.412	0.121	0.259	0.208
Part time (<30 hr/week)	0.247	0.079	0.348	0.326
Contract type				
Regular	0.398	0.117	0.270	0.214
Atypical	0.235	0.073	0.332	0.360
Firm size				
<20	0.263	0.076	0.372	0.289
20–99	0.362	0.118	0.283	0.237
100–500	0.443	0.144	0.215	0.198
≥500+	0.504	0.122	0.201	0.173
Industry				
Public administration	0.589	0.118	0.158	0.135
Primary	0.435	0.123	0.234	0.208
Finance or professional	0.449	0.096	0.270	0.184
Education	0.490	0.126	0.217	0.167
Trade or transportation	0.322	0.118	0.298	0.262
Construction	0.263	0.104	0.393	0.240
Health	0.399	0.143	0.248	0.210
Manufacturing	0.322	0.122	0.288	0.268
Hospitality	0.242	0.078	0.351	0.328
White collar				
Professionals	0.495	0.108	0.231	0.165
Clerks	0.376	0.116	0.282	0.226
Technicians	0.415	0.116	0.255	0.213
Managers	0.480	0.083	0.264	0.174
Operators	0.301	0.135	0.317	0.246
Elementary	0.309	0.116	0.264	0.311
Sales	0.299	0.097	0.317	0.287
Personal	0.276	0.120	0.315	0.289

Source: Authors' calculations based on the posterior probabilities derived from the baseline latent class analysis model.

also observed among older adults (31 percent), women (26 percent), and less educated persons (25 percent).[6] Figure 2 presents the predicted job quality profiles for these groups. Overall, the majority of part-time workers were predicted to be in either poor overall quality jobs (32 percent) or jobs with poor social environment and employment benefits (49 percent) compared with 17 percent and 26 percent, respectively, of full-time workers. The interactions of part-time and selected groups indicate that the risks of falling into the worst job quality class became even higher for 18- to 29-year-olds (44 percent) and older adults (49 percent), but less so for women (34 percent) and less educated persons (31 percent). These are 15 to 30 percentage points higher than the rates for their respective counterparts in full-time jobs. Meanwhile, their chances of being in high overall quality jobs were very limited. Those with less education (youth) in part-time employment, for example, had almost zero chance to be in the best job class. In stark contrast, the comparable figure was 58 percent for university graduates in full-time jobs and 45 percent for prime-age workers.

Job Quality Profiles among Frequent Atypical Contract Workers

The literature has also shown that some groups (youth and less-educated persons, in particular) as well as workers in small firms were more likely to be in jobs with NSW arrangements (Galarneau 2010; Kapsalis and Tourigny 2004; OECD 2015). Figure 3 presents the predicted job quality profiles for selected groups on the basis of employment contract type. Similar to the patterns of part-time employment, as many as 44 percent of workers in atypical forms of employment were predicted to be in the lowest quality jobs, and 41 percent were predicted to be in jobs with poor social environment and limited employment benefits. The figures were only 19 percent and 28 percent, respectively, among regular contract workers.

When interacting NSW with selected groups, older adults in atypical contracts faced a significantly higher probability (59 percent) of falling into the worst job quality class, and more than one-half of young (and less-educated) workers in NSW were predicted to be in jobs with poor social conditions and limited employment benefits. At the same time, their chances of being in high overall quality jobs were relatively low: about 35 percentage points or lower compared with their respective counterparts in regular contract jobs.

Moreover, the incidence of non-standard employment also tended to be higher in small firms. This is because atypical contracts are often less costly and more flexible for small firms to cope with fluctuations in demand (Bentolila and Saint-Paul 1994) or are used as a screening process (Portugal and Varejao 2009). In Canada, about one in three non-standard workers worked in a company with fewer than 20 employees, whereas one in five standard workers did so (Kapsalis and Tourigny 2004).

Figure 3 shows a stark contrast in job quality profiles between atypical contract workers in small firms (less than 20 employees) and regular contract workers in larger firms. All non-standard workers in small firms faced

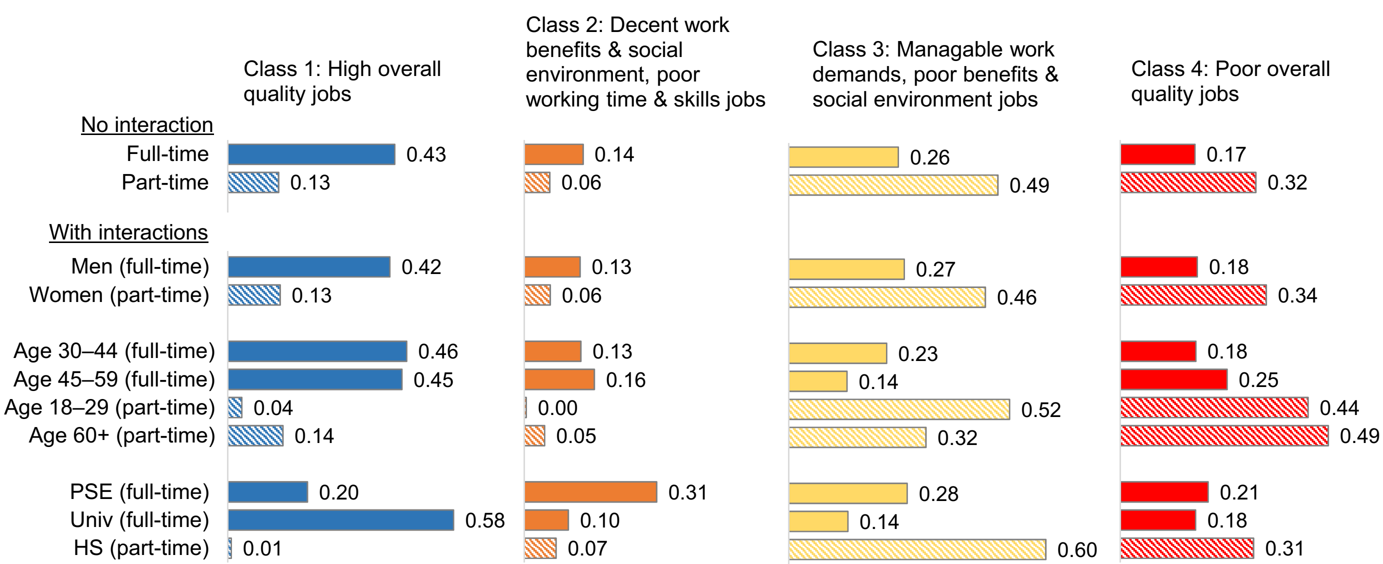

Figure 2: Predicted Probability of Job Quality Profiles for Selected Groups on the Basis of Full-Time or Part-Time Employment Status

Notes: The results were based on the latent class analysis in Table A.2. The values refer to predicted probabilities of belonging to a certain latent class with selected characteristics. HS = high school; PSE = post-secondary; Univ = university.

Source: Authors' calculations based on estimations from the 2016 General Social Survey.

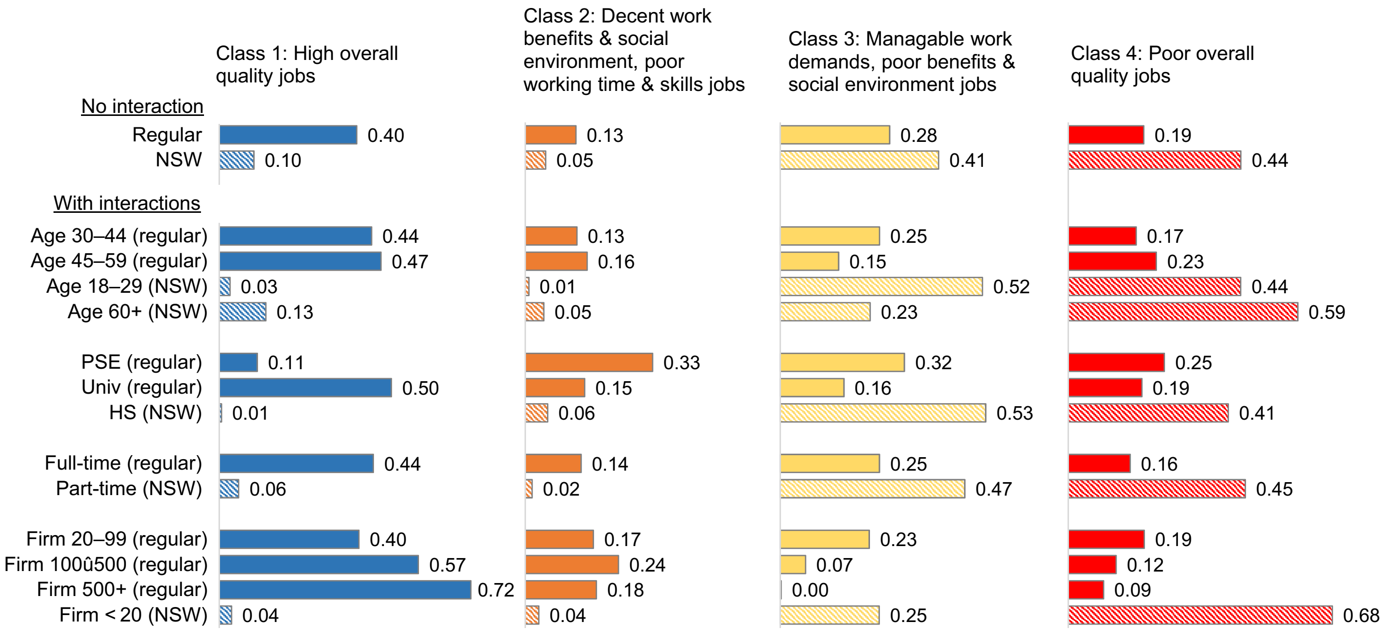

Figure 3: Predicted Probability of Job Quality Profiles for Selected Groups on the Basis of Employment Contract Type

Notes: The results are based on the latent class analysis in Table A.3. The values refer to predicted probabilities of belonging to a certain latent class with selective characteristics. HS = high school; NSW = non-standard work; PSE = post-secondary education; Univ = university.

Source: Authors' calculation based on estimations from the 2016 General Social Survey.

a very high risk (around 68 percent) of poor job quality outcome and a very low probability (around 4 percent) of belonging to the best job class. This may not be too surprising because small firms often lack many of the job features (e.g., unionization, benefit provision, training, or prospects) analyzed. Consequently, the growth of non-standard employment and its concentration in small firms will have profound implications for shaping the distribution of job quality in the modern era.

Conclusion

In this study, we assessed job quality in Canada using an internationally inspired multidimensional framework that covers six broad aspects: income and benefits, career prospects, work intensity, working-time quality, skills and discretion, and social environment. The descriptive results show diverse patterns of job quality across sectors and socio-demographic groups. In particular, some of the largest labour market segments, such as hospitality and personal services, exhibited lower job quality features in multiple dimensions. We also showed that the concerns relating to the youth labour market involve more than just a high level of unemployment or low participation. Marked differences in job quality were more apparent along the education line and, to a lesser extent, the gender line.

LCA identified four very distinct job quality profiles. About 37 percent of all workers were predicted to hold a high-quality job associated with many good job features in all six quality dimensions, whereas 23 percent had a overall poor-quality job, which lacks many good features in most dimensions. On average, 28 percent of workers held a job with manageable work demands but poor benefits and social environment, and the remaining 12 percent were considered to have a job with decent work benefits and social context but less in terms of working-time quality as well as skills and training opportunities.

The results indicate that non-standard employment arrangements are strong predictors of job quality classes: the majority of part-time or atypical contract workers were predicted to be either in the overall poor-quality jobs or in jobs with poor social environment and employment benefits. Moreover, firm size was positively associated with job quality, a result that suggests demand-side factors also play a role.

Finally, findings on job quality profiles for groups at risk of NSW suggest that workplace exclusion, either through lower pay or poorer access to other job quality features, can pose challenges to inclusive growth. Because many women, youth, and less-educated people today are engaging in part-time or non-standard contracts, either voluntarily or involuntarily, a growing dispersion of job quality along these lines could stand in the way of inclusive growth and a robust economy. The work done as part of this study provides partial guidance for future studies that could shed more light on gender, youth, occupational, or sectoral patterns of job quality.

Notes

1 Although we follow the Eurofound framework and guidelines to construct job quality indicators, we should note that the proposed indicators from the GSS are not exactly the same as those from the EWCS surveys. The GSS does not collect all workplace-related questions as does the EWCS. However, the information is sufficient enough to allow construction of indicators that cover the majority of the job quality dimensions mentioned.

2 Note that the GSS did not include data on earnings or wages. As a result, we derived information on hourly wage from an auxiliary source (the March 2016 LFS). This study focussed on hourly wages rather than monthly earnings (as in Eurofound) because the latter also depends on hours of work. High monthly income therefore cannot be regarded as a good job quality if it was a result of involuntary long working hours.

3 The industry groups were constructed on the basis of the 2012 one-digit North American Industry Classification System. We bundled the "other services" category with "hospitality" because of small sample size. The occupational groups were constructed on the basis of a variant version of the National Occupational Classification, which consists of 10 categories similar to the one-digit level in the International Standard Classification of Occupations. For the sake of parsimony, we combined the categories "industrial and equipment operation trades" and "workers in transport and construction" into "machine operators"; similarly, we combined "nature resources, agriculture occupations" and "processing, manufacturing and utilities labourers" into "elementary."

4 In this study, a cell in question is regarded as significantly better than average if its value is greater than the average for positive indicators (or lower than the average for negative indicators) with the estimated difference (between-groups mean and overall mean) being significant at the 1 percent level (i.e., $p \leq 0.01$), better than average if the difference is significant between the 1 percent and 10 percent level, and within the average if the difference is not significant at the 10 percent level. The notion of significantly worse than average and worse than average is defined in an analogous manner.

5 For a negative job quality indicator (e.g., may lose job in six months), workers without such a quality are considered to have a "good" feature in that dimension.

6 In comparison, part-time rates were lower for men (12 percent) and for highly educated workers (14 percent). See Statistics Canada Tables 14-10-0018-01 and 14-10-0020-01.

References

Arends, I., C. Prinz, and F. Abma. 2017. "Job Quality, Health and At-Work Productivity." OECD Social, Employment and Migration Working Paper No. 195, OECD Publishing, Paris.

Autor, D.H., and D. Dorn. 2013. "The Growth of Low Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103(5):1553–97. https://doi.org/10.1257/aer.103.5.1553.

Bardasi, E., and J. Gornick. 2008. "Working for Less? Women's Part-Time Wage Penalties across Countries." *Feminist Economics* 14(1):37–72. https://doi.org/10.1080/13545700701716649.

Bentolila, S., and G. Saint-Paul. 1994. "A Model of Labor Demand with Linear Adjustment Costs." *Labour Economics* 1(3):303–26. https://doi.org/10.1016/0927-5371(94)90015-9.

Brisbois, R. 2003. *How Canada Stacks Up: The Quality of Work: An International Perspective*. Research Paper W|23, Work Network. Ottawa: Canadian Policy Research Networks.

Catalina-Romero, C., J.C. Sainz, J.I. Pastrana-Jimenez, N. Garcia-Dieguez, I. Irizar-Munoz, J.L. Aleixandre-Chiva, A. Gonzalez-Quintela, and E. Calvo-Bonacho. 2015. "The Impact of Poor Psychosocial Work Environment on Non-Work-Related Sickness Absence." *Social Science and Medicine* 138:210–16. https://doi.org/10.1016/j.socscimed.2015.06.009.

Cazes, S., A. Hijzen, and A. Saint-Martin. 2016. "Measuring and Assessing Job Quality: The OECD Job Quality Framework." OECD Social, Employment and Migration Working Paper No. 174. OECD Publishing, Paris.

Chappell, D., and V. Di Martino. 2006. *Violence at Work*, 3rd edition. Geneva: International Labour Office.

Chen, W.-H., and T. Mehdi. 2018. "Assessing Job Quality in Canada: A Multidimensional Approach." Analytical Studies Branch Research Paper Series, Cat. No. 11F0019M – No. 412. Ottawa: Statistics Canada.

Drobnic, S., B. Beham, and P. Prag. 2010. "Good Jobs, Good Life? Working Conditions and Quality of Life in Europe." *Social Indicators Research* 99(2):205–25. https://doi.org/10.1007/s11205-010-9586-7.

European Foundation for the Improvement of Living and Working Conditions (Eurofound). 2016. *Sixth European Working Conditions Survey: Overview Report*. Luxembourg: Publications Office of the European Union.

Galarneau, D. 2010. "Temporary Employment in the Downturn." *Perspectives on Labour and Income* 11(11):5–17. Statistics Canada Cat. No. 75-001-X.

Garcia-Mainar, I., V. Montuenga, and M. Navarro-Paniagua. 2015. "Workplace Environmental Conditions and Life Satisfaction in Spain." *Ecological Economics* 119:136–46. https://doi.org/10.1016/j.ecolecon.2015.08.017.

Goos, M., A. Manning, and A. Salomons. 2009. "Job Polarization in Europe." *American Economic Review* 99(2):58–63. https://doi.org/10.1257/aer.99.2.58.

Green, D., and B. Sand. 2015. "Has the Canadian Labour Market Polarized?" *Canadian Journal of Economics/Revue canadienne d'économique* 48(2):612–46. https://doi.org/10.1111/caje.12145.

Green, F. 2010. "Wellbeing, Job Satisfaction and Labour Mobility." *Labour Economics* 17(6):897–903. http://doi.org/10.1016/j.labeco.2010.04.002.

Horowitz, J. 2016. "Dimensions of Job Quality, Mechanisms, and Subjective Well-Being in the United States." *Sociological Forum* 31(2):419–40. https://doi.org/10.1111/socf.12251.

International Labour Office (ILO). 2016. *Non-Standard Employment around the World: Understanding Challenges, Shaping Prospects*. Geneva: ILO.

Jackson, A., and P. Kumar. 1998. *Measuring and Monitoring the Quality of Jobs and the Work Environment in Canada*. Paper presented at the CSLS Conference on the State of Living

Standards and the Quality of Life in Canada, Ottawa, 30–31 October.

Kapsalis, C., and P. Tourigny. 2004. "Duration of Non-Standard Employment." *Perspectives on Labour and Income* 5(12):5–13. Statistics Canada Cat. No. 75-001-X.

Lanza, S.T., J.J. Dziak, L. Huang, A.T. Wagner, and L.M. Collins. 2015. *LCA Stata Plugin Users' Guide* (Version 1.2). University Park: Methodology Center, Pennsylvania State University. At http://www.methodology.psu.edu.

Lowe, G. 2007. *21st Century Job Quality: Achieving What Canadians Want*. CPRN Research Report W|37. Ottawa: Canadian Policy Research Networks.

Lowe, G., and G. Schellenberg. 2001. *What's a Good Job? The Importance of Employment Relationships*. CPRN Study W|05. Ottawa: Canadian Policy Research Networks.

Milner, A., P. Butterworth, R. Bentley, A.M. Kavanagh, and A.D. LaMontagne. 2015. "Sickness Absence and Psychosocial Job Quality: An Analysis from a Longitudinal Survey of Working Australians, 2005–2012." *American Journal of Epidemiology* 181(10):781–8. https://doi.org/10.1093/aje/kwu355.

Morissette, R. 2016. "Perspectives on the Youth Labour Market in Canada." Statistics Canada Catalogue No. 11-631-X. Ottawa: Statistics Canada.

Moyser, M. 2017. "Women and Paid Work." In *Women in Canada: A Gender-Based Statistical Report*. Statistics Canada Catalogue No. 89-503-X. Ottawa: Statistics Canada.

O'Dorchai, S., R. Plasman, and F. Rycx. 2007. "The Part-Time Wage Penalty in European Countries: How Large Is It for Men?" IZA Discussion Paper No. 2591. Bonn, Germany: Institute for the Study of Labor.

Organisation for Economic Co-operation and Development (OECD). 2013. *How Is Life? Measuring Well-Being*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. 2014. "How Good Is Your Job? Measuring and Assessing Job Quality." In *OECD Employment Outlook 2014*, 79–139. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). 2015. "The Quality of Working Lives." In *OECD Employment Outlook 2015*, 167–202. Paris: OECD Publishing.

Park, J. 2010. "Health Factors and Early Retirement among Older Workers." *Perspectives on Labour and Income* 11(6): 5–13. Statistics Canada Cat. No. 75-001-X.

Portugal, P., and J. Varejao. 2009. "Why Do Firms Use Fixed-Term Contracts?" IZA Discussion Paper No. 4380. Bonn, Germany: Institute for the Study of Labor.

Salvatori, A. 2010. "Labour Contract Regulations and Workers' Wellbeing: International Longitudinal Evidence." *Labour Economics* 17(4):667–78.

Shields, M. 2006. "Unhappy on the Job." *Health Reports* 17(4):33–37.

Shields, M., and K. Wilkins. 2009. "Factors Related to On-the-Job Abuse of Nurses by Patients." *Health Reports* 20(2):7–17.

Statistics Canada. General Social Survey: Canadians at Work and Home (GSS). Ottawa: Statistics Canada. At http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5221.

Tal, B. 2016. "On the Quality of Employment in Canada." Toronto: Canadian Imperial Bank of Commerce.

Turcotte, M., and G. Schellenberg. 2005. "Job Strain and Retirement." *Perspectives on Labour and Income* 6(7):13–17. Statistics Canada Cat. No. 75-001-X.

Vermunt, J.K., and J. Magidson. 2002. "Latent Class Cluster Analysis." In *Applied Latent Class Analysis*, ed. J.A. Hagenaars and A.L. McCutcheon, 89–106. Cambridge: Cambridge University Press.

Appendix

Table A.1: Job Quality and Life and Job Satisfaction

Job Quality Dimension and Indicators	Positive (+) or Negative (–) Indicator	Answer to Job Quality Question: Life Satisfaction Score of at Least 8/10		Answer to Job Quality Question: Job Satisfaction Rating of Very Satisfied or Satisfied	
		Yes	No	Yes	No
Prospects					
May lose job in the next 6 months	–	51.6	67.7	72.6	86.9
Job offers good prospects for career advancement	+	73.6	55.6	93.3	75.6
Work intensity					
Workload not often manageable	–	54.2	70.2	75.1	89.0
Can't often finish assigned work during regular working hours	–	56.2	67.8	78.5	86.9
Working-time quality					
Involuntary irregular (rotating, split, on-call) shift job	–	58.0	67.0	74.2	86.8
Can choose start and end time of your work day	+	67.6	63.2	87.7	82.7
Easy to take 1 or 2 hours off for personal matters	+	67.9	57.6	88.0	76.6
Skills and discretion					
Can choose the sequence of tasks	+	69.6	58.9	90.3	75.6
Have opportunities to provide input into decision	+	69.6	49.6	91.0	64.1
Had formal training paid by employer	+	69.2	62.0	88.3	82.3
Had informal or on-the-job training	+	65.1	64.8	85.8	83.3
Social environment					
Experienced verbal, sexual or physical violence at work	–	52.6	68.3	72.9	87.6
Received support from managers	+	70.4	60.2	91.3	77.2
Had a formal job performance assessment	+	67.1	61.9	86.1	82.8
Covered by a union contract or collective agreement	+	67.8	63.7	85.4	84.5
Income and benefits					
Hourly earnings [a]	+	n.a.	n.a.	n.a.	n.a.
Workplace pension plan	+	72.9	60.0	89.3	81.8
Paid sick leave	+	72.3	59.7	89.0	81.6
Paid vacation leave	+	71.0	57.4	87.9	80.7
Disability insurance	+	72.3	59.7	89.2	81.5
Supplemental medical/dental care	+	70.6	60.2	87.9	82.0
Worker's compensation	+	70.9	59.3	88.1	81.5
Maternity, parental or layoff benefits	+	72.0	59.9	88.6	81.9

Notes: The sample was restricted to workers (including those who were self-employed) aged 18 y and older. The number of overall observations is 10,680. Actual sample size, however, varies across indicators because not every person in the sample answered all the job quality questions. The life satisfaction question was answered on a scale ranging from 0 (*very dissatisfied*) to 10 (*very satisfied*). The job satisfaction question was answered on a scale ranging from 1 to 5 (1 = *very satisfied,* 2 = *satisfied,* 3 = *neither satisfied nor dissatisfied,* 4 = *dissatisfied,* 5 = *very dissatisfied*). n.a. = not available.

[a] Answers to the satisfaction questions were not available for hourly wages indicator, for which information was drawn from the Labour Force Survey, not the General Social Survey.

Source: Authors' calculations from Statistics Canada, 2016 General Social Survey.

Table A.2: Comparison of Baseline Models (*N* = 8,004)

No. of Classes	Log-Likelihood	Degrees of Freedom	No. of Estimated Parameters	AIC	BIC
2	–833.6	703	25	1,717.2	1,891.9
3	–646.6	690	38	1,369.1	1,634.6
4	–567.3	677	51	1,237.0	1,593.4
5	–521.6	664	64	1,171.2	1,618.4
6	–482.6	651	77	1,119.1	1,657.1

Notes: The sample was restricted to workers aged 18 y and older, with valid answers to all job quality questions in the General Social Survey. Total number of observations = 8,001. The degree of freedom is calculated as the number of unique observed job quality patterns, minus the number of parameters that are freely estimated, minus 1. In our case, with 6 (3-category) job quality response variables, there are 3 x 3 x 3 x 3 x 3 x 3 = 729 possible patterns. So the degree of freedom in a 4-class model, for example, would be 729 − 3 (number of latent class membership probabilities) – 48 (number of item-response probabilities estimated) − 1 = 677.

Source: Statistics Canada, 2016 General Social Survey.

Table A.3: Estimated Coefficients for Four-Class LCA Regressions, by Specification Relating to Part-Time and Full-Time Status (*N* = 8,001)

Specification and Covariate	Relative Risk Ratios for Jobs with		
	Decent Work Benefits & Social Environment, Poor Working-Time and Skills: Class 2	Manageable Work Demands, Poor Benefits & Social Environment: Class 3	Poor Overall Quality: Class 4
LCA 1_1			
Intercept	0.261*	0.486*	0.486*
Part time (vs. full time)	1.423	6.272*	5.968*
LCA 1_2			
Intercept	0.258*	0.520*	0.515*
Women (vs. men)	0.918	0.822	0.927
Part time (vs. full time)	0.915	7.187*	6.097*
Women × part time	1.712	0.893	1.040
LCA 1_3			
Intercept	0.240*	0.393*	0.431*
Ages 18–29 (vs. Ages 30–44)	0.960	4.045*	2.111*
Ages 45–59 (vs Ages 30–44)	1.274	0.612*	1.454*
Ages ≥ 60 (vs. Ages 30–44)	1.675*	1.091	1.863*
Part time (vs. full time)	3.792*	10.744*	13.779*
Ages 18–29 × part time	0.117	0.664	1.095
Ages 45–59 × part time	0.195*	0.278*	0.212*
Ages ≥ 60 × part time	0.184*	0.388	0.349*
LCA 1_4			
Intercept	1.436	1.270	1.209
High school (vs. post-secondary)	2.655	4.528*	3.328
University (vs. post-secondary)	0.117*	0.176*	0.302*
Part time (vs. full time)	6.284*	24.355*	19.052*
High school × part time	0.335	0.459	0.558
University × part time	0.745	0.431	0.284

Notes: Class 1 is the reference class. The sample was restricted to workers aged 18 y and older, with valid answers to all job quality questions in the General Social Survey. LCA = latent class analysis.

* Statistically significant relative risk value from the 95% confidence interval.

Source: Authors' calculations from Statistics Canada, 2016 General Social Survey.

Table A.4: Estimated Coefficients for Four-Class LCA regressions, by Specification Relating to Employment Contract Type (N = 8,001)

Specification and Covariate	Relative Risk Ratio for Jobs with		
	Decent Work Benefits & Social Environment, Poor Working-Time and Skills Jobs: Class 2	Manageable Work Demands, Poor Benefits & Social Environment: Class 3	Poor Overall Quality: Class 4
LCA 2_1			
Intercept	0.270*	0.602	0.579*
Atypical contract (vs. regular)	1.609	5.754*	9.082*
LCA 2_2			
Intercept	0.257*	0.469*	0.460*
Ages 18–29 (vs. Ages 30–44)	0.970	3.960*	2.669*
Ages 45–59 (vs Ages 30–44)	1.128	0.555*	1.221
Ages ≥ 60 (vs. Ages 30–44)	1.558	1.133	1.667*
Atypical contract (vs. regular)	1.510	4.239*	6.920*
Ages 18–29 × NSW	0.721	1.724	1.953
Ages 45–59 × NSW	1.831	1.487	1.280
Ages ≥ 60 × NSW	0.504	0.617	0.957
LCA 2_3			
Intercept	3.394*	3.106*	2.899*
High school (vs. post-secondary)	7.673	13.449*	11.118*
University (vs. post-secondary)	0.103*	0.113*	0.169*
Atypical contract (vs. regular)	2.736	7.651*	11.274*
High school × NSW	0.149	0.282	0.233
University × NSW	0.532	1.761	0.970
LCA 2_4			
Intercept	0.267*	0.475*	0.427*
Atypical contract (vs. regular)	1.524	5.243*	7.392*
Part time (vs. full time)	1.486	6.146*	6.054*
NSW × part time	0.449	0.460	0.509
LCA 2_5			
Intercept	0.382*	0.500*	0.637*
Firm size <20 (vs. 20–99)	0.762	3.586*	2.507*
Firm size 100–500 (vs. 20–99)	0.958	0.200*	0.442*
Firm size ≥500 (vs. 20–99)	0.580	0.006*	0.259*
Atypical contract (vs. regular)	0.930	4.842*	6.185*
Firm size <20 × NSW	3.263	0.732	2.563
Firm size 100–500 × NSW	0.719	0.352	1.172
Firm ≥500 × NSW	1.557	67.046*	1.093

Note: Class 1 was the reference class. Sample was restricted to workers aged 18 y and older, with valid answers to all job quality questions in the General Social Survey. LCA = latent class analysis; NSW = non-standard work.

* Statistically significant relative risk value from the 95% confidence interval.

Source: Authors' calculations from Statistics Canada, 2016 General Social Survey.

Appendix B: Ethics review

FORM UPR16

Research Ethics Review Checklist

Please include this completed form as an appendix to your thesis (see the Research Degrees Operational Handbook for more information

UNIVERSITY OF PORTSMOUTH

Postgraduate Research Student (PGRS) Information		**Student ID:**	909418
PGRS Name:	Tahsin Fazle Mehdi		
Department:	ECFIN	**First Supervisor:**	Lester Hunt
Start Date: (or progression date for Prof Doc students)		**February 1, 2019**	
Study Mode and Route:	Part-time ☒ Full-time ☐	MPhil ☐ PhD ☒	MD ☐ Professional Doctorate ☐

Title of Thesis:	Robust comparisons of socio-economic well-being
Thesis Word Count: (excluding ancillary data)	8,666

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a) Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES ☒ NO ☐
b) Have all contributions to knowledge been acknowledged?	YES ☒ NO ☐
c) Have you complied with all agreements relating to intellectual property, publication and authorship?	YES ☒ NO ☐
d) Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES ☒ NO ☐
e) Does your research comply with all legal, ethical, and contractual requirements?	YES ☒ NO ☐

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	N/A

If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:

This is a PhD by publications.

Signed *(PGRS)*:	Tahsin Mehdi	**Date:**	July 27, 2020