'Fair' Welfare Comparisons with Heterogeneous Tastes: Subjective versus Revealed Preferences*

Alpaslan Akay, Olivier Bargain, H. Xavier Jara

August 2019

Abstract

Multidimensional welfare analysis has recently been revived by money-metric measures based on explicit fairness principles and the respect of individual preferences. To operationalize this approach, preference heterogeneity can be inferred from the observation of individual choices (revealed preferences) or from self-declared satisfaction following these choices (subjective well-being). We question whether using one or the other method makes a difference for welfare analysis based on income-leisure preferences. We estimate ordinal preferences that are either consistent with actual labor supply decisions or with income-leisure satisfaction. For ethical priors based on the compensation principle, we compare the welfare rankings obtained with both methods. The correlation in welfare ranks is high in general and reranking is insignificant for 77% of the individuals. The remaining discrepancies possibly pertain to a variety of factors including constraints (health issues, labor market rationing), irrational behavior and alternative life choices to the pursuit of well-being. We discuss the implications of using one or the other preference elicitation method for welfare analysis.

Key Words : fair allocation, money metric, decision utility, experienced utility, labor supply, subjective well-being.

JEL Classification : C35, C90, D60, D63, D71, H24, H31, J22

*__Acknowledgements:__ Akay is affiliated with the University of Gothenburg and IZA; Bargain is affiliated with Bordeaux University, IUF and IZA; Jara is affiliated with the University of Essex and ISER. We are grateful to two referees and to participants to seminars (LSE, ISER, AMSE) and conference (Royal Economic Society, HDCA, Warsaw International Economic Meeting) for useful suggestions. We are indebted to Jan Stuhler for his help on the application of a new method to correct rank correlation from the bias associated with measurement errors, as developped by him and his coauthors (Kitagawa et al., 2018). Corresponding author: Olivier Bargain, LAREFI, Bordeaux Université, rue Léon Duguit, Pessac, France, olivier.bargain@u-bordeaux.fr

1 Introduction

The recent years have witnessed a resurgence of interest in the measurement of well-being (Stiglitz et al., 2009, Fleurbaey, 2009) and the attempt to address the multidimensionality problem. In particular, the use of subjective well-being (SWB) – be it life satisfaction, happiness or mental health – has surged in social sciences as a broad welfare measure that possibly encompasses many other dimensions than income (see Senik, 2005, and Clark et al., 2008). In this approach, SWB is assumed to be an index that can lend itself to interpersonal comparison – a feature that is not well regarded by a large part of the economic profession (see the discussion by Ng, 1997). At the same time, considerable progress has been made in the measurement of multi-dimensional welfare based on money metric utility, an approach that is much more integrated in the standard microeconomic apparatus. Notably, the 'fair allocation' theory suggests ways to construct welfare indices that only require information about ordinal and non-comparable preferences while providing a rigorous framework to perform interpersonal comparisons and, hence, distributional analyses (Fleurbaey and Maniquet 2006, 2011).

A central feature of the 'fair allocation' approach is the respect for (ordinal) preference heterogeneity. It is possible to rank individual situations, when preferences differ, while escaping from most of the standard criticisms about money metrics. However, the devil might be in the empirics. One must make a choice about how to retrieve individual preferences when operationalizing this method for distributional analysis. Typical ways of doing so include attempts to elicit revealed, stated or subjective preferences. These approaches possibly capture only some aspects of 'authentic' preferences, which potentially differ. Recently, several applications of the 'fair allocation' theory have relied on *revealed preferences*, focusing on individual trade-offs between income and leisure.[1] Despite the focus on two dimensions of well-being, this domain is crucial for normative analyses because it is the place where redistributive policies operate, as made clear in the long tradition of second best policy design and optimal taxation.[2] Other studies have originally suggested ways to infer what we shall refer to as *subjective preferences*, i.e. ordinal preferences derived from SWB data.[3]

[1] See Bargain et al. (2013), Decoster and Haan, (2014) and Carpantier and Sapata (2016). The first two studies consider preference heterogeneity across countries and across groups within Germany, respectively. The third paper suggests a refined treatment of unobserved preferences.

[2] The bulk of this literature has assumed that individuals only differ in their abilities but have identical preferences otherwise (Boadway, 2012). Importantly, recent developments in optimal taxation have suggested to respect preference heterogeneity using fair allocation principles (Schokkaert et al., 2004, Jacquet and Van de Gaer, 2011, Fleurbaey and Maniquet, 2006, 2007, 2014).

[3] Decancq et al. (2014) suggest a method to construct money metric evaluation of "the good life", incorporating many dimensions beyond income, based on subjective data. Schokkaert et al. (2011) focus on income and job satisfaction. Decancq and Schokkaert (2013) and Decancq et al. (2015) follow similar approaches while

In this paper, we ask whether the way we elicit preference heterogeneity – i.e. from choices or from the subjective experience derived from these choices – makes a difference for welfare analysis. To the best of our knowledge, this question has never been addressed, making our contribution original in this respect. Nonetheless, it is closely related to the literature comparing decision and experienced utility. Recent studies – notably Benjamin et al. (2012, 2014) and Fleurbaey and Schwandt (2015) – conclude to an *overall* congruence between decision and experienced utility. When differences exist, they provide relatively intuitive explanations for them.[4] Our work further extends this question by asking about the *distributional* consequences – i.e. whether revealed or subjective preferences lead to different conclusions when used to rank people according to ethically-grounded money-metric welfare measures.

Our investigation focuses on a bidimensional measure of welfare comprising income and non-market time. Using the British Household Panel Study (BHPS), we first proceed with the estimation of ordinal preferences that are consistent either with the labor supply decisions made by observed individuals or with the subjective experience they derive from these choices, as proxied by a combined measure of income and leisure satisfactions. Section 2 presents the data and a brief outline of the procedure used to elicit preference heterogeneity in both approaches. Section 3 describes the welfare metrics and how we calculate them using estimated preferences. We focus on ethical priors that give priority to the "compensation principle" (for given preferences, inequalities due to non-responsibility factors should be compensated), with alternative views about individual responsibilities for work preferences (Fleurbaey, 2008). In the default case, we focus on a metric whereby some of the differences due to preferences should be compensated (Rent metric).

Results are presented and discussed in Section 4. We first check whether the characterization of welfare inequality varies when using revealed versus subjective preferences. For the main comparison, the anonymity property is removed and we confront the welfare ranks obtained with both preference elicitation methods. We check to which extent reranking may be due to modelling choices, noise or measurement errors. Reranking is statistically insignificant for 77% of the observations. Descriptive estimations show that the remaining discrepancies possibly

focusing on social progress and poverty respectively. Jara and Schokkaert (2017) assess tax reforms using SWB or equivalent income derieved from SWB.

[4]The first explicit comparison has been suggested by Benjamin et al. (2012, 2014), who proxy experienced utility using SWB and decision utility using stated or actual preferences in tailor-made studies. Regarding a broad range of life choices, Fleurbaey and Schwandt (2015) ask people if they can think of changes that would increase their SWB score. Akay et al. (2015) use large microdata to compare labor supply decisions and income-leisure SWB on average. Considering own income versus others' income, Clark et al. (2015) find similar relative concerns in happiness regressions and in hypothetical-choice experiments. Arguably, more divergence is found in other recent studies based on job satisfaction (Ferrer-i-Carbonell et al., 2010), residential choice (Glaeser et al., 2016) or consumption (Perez-Truglia, 2015).

pertain to a variety of factors including individual constraints (e.g., health issues) or labor market constraints (e.g. frictions, discrimination), irrational behavior and life choices that diverge from the pursuit of individual well-being. We find a fairly large overlap between the groups identified as the worst-off individuals. For a sensitivity analysis, we present results for a polar ethical view in terms of responsibility for work aversion (Wage metric). We end the paper with a thorough discussion about the implications of using one or the other preference elicitation method, in conjunction with particular ethical priors.

2 Estimation of Revealed and Subjective Preferences

We first present the empirical approach aimed to elicit revealed preferences from labor supply choices and subjective preferences from SWB information. As we shall see, the estimation methods are state-of-the-art in their respective fields. The estimation of revealed preferences follows the literature on structural model estimations in the presence of nonlinear budget constraints reflecting real-world taxes and benefits. The estimation of subjective preferences relies on the standard approach in the SWB literature but the functional form is slighlty more demanding than usual for the sake of comparability.

2.1 Data, Selection and Key Variables

Data and Selection. Our empirical application is based on data from the British Household Panel Survey (BHPS), a nationally representative survey collected in the United Kingdom between 1991-2008. It contains life satisfaction information since 1996 and standard information on socio-demographic characteristics that are used in our estimations. We restrict our analysis to single individuals, since extending the analysis to couples is extremely difficult.[5] This is not a particular problem since our empirical application does not aim to perform a nationally representative welfare analysis. We further exclude individuals in self-employment because their labor supply decisions may considerably differ from those of salaried workers and because income information from surveys is much less reliable in their case. We select people aged 18 to

[5] On the revealed preference side, to recover individual ordinal preferences requires the full identification of a collective model of labor supply with nonlinear taxation, which has very rarely been done (see Chiappori and Donni, 2011). Some of the rare attempts – Lise and Seitz (2011) and Bloemen (2018) – focus mainly on the sharing rule, which can be seen as a specific form of money metric utility. On the SWB side: one would need to recover individual income-leisure satisfaction within couples. However, when answering the income satisfaction question, it is unclear whether the wife (husband) expresses her (his) satisfaction about the household's total budget or whether she (he) talks about the resources available to her (him) in the household. Only the latter is the appropriate measure, to be put against her (his) level of leisure, to measure female (male) income-leisure preferences.

64 who are available for the labor market (not disabled nor full-time students or pensioners). Importantly, we exclude all job seekers, defined according to the questions about whether they have actively looked for a job within the last four weeks and are ready to take up a job within the next two weeks. While this steps aims to comply with the labor supply nature of the model, we probably do not discard all the persons facing labor market constraints (notably the discouraged workers or people not optimizing their work duration), as we will explain later. Finally, we keep individuals for whom all key characteristics are available for all years, and years in which all key variables are available (this leads to the exclusion of years 2006-7). We obtain a sample covering the years 1996-2005 and including $4,549$ observations (person $\times$ year variation).

Income and Leisure Time. The key variables for our analysis are disposable income and non-market time (the former is also referred to as 'leisure' for simplicity). Weekly working hours reported in the data are denoted h_{it}, for individual i in year t. Denote τ the maximum time available for work, so that 'leisure' is written $l_{it} = \tau - h_{it}$ for individual i in year t. Disposable income, denoted y_{it}, is calculated as $y_{it} = \psi_t(w_{it}h_{it}, \mu_{it}, \zeta_{it})$, using reported gross earnings $w_{it}h_{it}$ (hourly wage rates $\times$ work hours) and unearned income μ_{it}. Function ψ_t represents the aggregation of all incomes and the imputation of taxes and benefits, using numerical simulations of tax-benefit rules at each period $t = 1, \ldots, T$. The set ζ_{it} represents individual characteristics that matter for tax-benefit calculations and are extracted from the data, for instance the presence of children (which conditions the calculation of child benefits, increment of income support, tax credits, etc.).

SWB. SWB information is drawn from the answer to the life satisfaction questions. The main one, "How dissatisfied or satisfied are you with your life overall?", is measured on an ordered scale between 1 and 7 (1 means "not satisfied at all" and 7 means "completely satisfied").[6] While it could be used directly for our purpose, we aim to retrieve ordinal preferences that specifically concern the trade-off between income and leisure.[7] There is obviously no question about the relative well-being drawn from these two dimensions of life. Interestingly, however, the data contain satisfaction on life domains that can be combined for this purpose (see also van Praag et al., 2003, on how to combine the 'domains of satisfaction'). We rely on questions about how dissatisfied or satisfied respondents are regarding "the income of your household"

[6]Information on mental health is also available (namely the index from the General Health Questionnaire GHQ-12) as well as answers to the happiness question. These alternative SWB measures lead to relatively similar results regarding the estimation of ordinal income-leisure preferences (see Akay et al., 2015).

[7]Importantly, note that hours of work and gross income (used to compute disposable income) refer to the last week while subjective well-being indices correspond to the date of interview.

and "the amount of leisure time you have" (also on 1-7 scales). To combine these variables into an income-leisure satisfaction measure, we proceed as follows. We first regress overall satisfaction of individual i at time t, denoted S_{it}, on her income satisfaction S_{it}^y and her leisure time satisfaction S_{it}^l, i.e. we estimate the equation

$$S_{it} = \gamma^y S_{it}^y + \gamma^l S_{it}^l + e_{it}. \tag{1}$$

We then predicted value $\widehat{V}_{it}^E = \widehat{\gamma}^y S_{it}^y + \widehat{\gamma}^l S_{it}^l$, i.e. the experienced utility is an income-leisure 'concentrated' satisfaction measure. It aims to capture the share of life satisfaction that is driven by the satisfaction from income and from leisure. In sensitivity analyses, we will present alternative options, e.g. one where we simply rely on overall life satisfaction rather than the concentrated measure ($V_{it}^E = S_{it}$) or alternative concentrated approaches, notably allowing individual heterogeneity in coefficients γ in equation (1).

2.2 Estimation of Implicit Preferences from SWB and from Choices

General Model. We proceed with the estimation of ordinal preferences based on either subjective well-being or actual labor supply choices. We present here a summary of the estimation methods and of the main modelling choices – additional details are provided in Appendix A.1 and Akay et al. (2015). The deterministic function of income and leisure that defines ordinal preferences over these two dimensions is written $u_{it}^m(y_{it}, l_{it})$. We will consider either $m = E$ (experienced utility) or $m = D$ (decision utility). Estimations rely on the identity:

$$V_{it}^m = u_{it}^m(y_{it}, l_{it}) + \eta_{it}^m. \tag{2}$$

We use a box-cox specification of function $u_{it}^m(\cdot,\cdot)$ for $m = E, D$. The paper focuses on welfare metrics that deal with preference heterogeneity across individuals. Most often in the literature, welfare metrics correspond to group-level preferences, which are far from being individual preferences because only a few taste shifters are considered (see the discussion in Carpantier and Sapata, 2016). In our application, the utility function $u_{it}^m(\cdot,\cdot)$ varies with a wide range of characteristics including standard socio-demographic variables (dummies for gender, age above 40, higher education, presence of children aged 0 to 2, living in London, non-white ethnic origin, migrant) as well as two personality traits that provide some more heterogeneity across individuals, namely conscientiousness and neuroticism.[8] It also varies with a normally-distributed

[8] We use above-average dummies for ease of interpretation. Note that among the different personality traits, conscientiousness and neuroticism are shown to be what matters the most for labor supply choices (see Wichert and Pohlmeier, 2010). Neuroticism is a fundamental personality trait in the study of psychology characterized by anxiety, fear, moodiness, worry, envy, frustration, jealousy, and loneliness. Conscientiousness is the personality trait of being thorough, careful, or vigilant, implying the desire to do a task well.

random term for unobserved preferences, as specified in Appendix A1. While the form of $u_{it}^m(\cdot,\cdot)$ is common to both approaches, estimation methods – and the assumptions underlying the additive terms η_{it}^m, $m = D, E$ – are necessarily specific, as we now explain.

Estimation of Subjective Preferences. We first focus on preferences that are consistent with the SWB experienced at the observed income-leisure choice, namely function $u_{it}^E(\cdot,\cdot)$. The approach consists in a direct estimation of equation (2) whereby V_{it}^E is supposed to be known. As explained, our main proxy for it is the concentrated income-leisure satisfaction, calculated as $\widehat{V}_{it}^E = \widehat{\gamma}^y S_{it}^y + \widehat{\gamma}^l S_{it}^l$. It is assumed to be the well-being level experienced by individual i at period t working h_{it} hours per week and consuming y_{it}. The residual term is specified as $\eta_{it}^E = \lambda' z_{it} + \rho' \alpha_i + \epsilon_{it}$ to control for individual heterogeneity and subjectivities in well-being responses (Decancq et al., 2015). This comprises observable characteristics z_{it} corresponding to the usual determinants of well-being (cf. Clark et al., 2008),[9] individual effects α_i and i.i.d., normally distributed error terms ϵ_{it}. The individual effect α_i is not a fixed effect in the usual sense, as it would absorb all the time-invariant characteristics. We rather put more structure on it by making it a function of the period-average of most time-varying characteristics (a quasi-fixed effect à la Mundlak) and of the 'big five' personality traits. The latter have been shown to account for an important part of the individual variation in SWB (Boyce, 2010, Ravallion and Lokshin, 2001).[10] Utility $\widehat{V}_{it}^E$ is treated as continuous and the model is estimated by maximum likelihood (ML) to address the nonlinearity of the box-cox specification.

Estimation of Revealed Preferences. Then we elicit revealed preferences, namely function $u_{it}^D(\cdot,\cdot)$, as behavioral parameters consistent with the person's actual choice. The approach essentially requires information on the labor supply decision, deemed optimal for individual i at time t, and the resulting level of disposable income. Individuals are assumed to choose the bundle (y_{it}, l_{it}) according to a classic utility maximization problem:

$$(y_{it}, l_{it}) = \arg\max \left[u_i\left(y, l\right) | y \leq \psi(w_{it}(\tau - l), \mu_{it}, \zeta_{it}), l \leq \tau \right] \tag{3}$$

[9] Observed heterogeneity z_{it} includes gender, age (and age squared), education, health status, presence of children aged 0 to 2, living in London, non-white ethnic origin, migrant, family size, home ownership, region and year. Remark that some of these variables are allowed here to have a direct effect on SWB but also enter in income-leisure preference heterogeneity.

[10] Note also that at least three types of unobserved variables may limit the possibility of interpersonal comparability of SWB (see Fleurbaey and Blanchet, 2013): (i) omitted variables that make people perceive and interpret SWB scales differently, (ii) omitted personality traits that make them respond differently or adapt differently to their own conditions (e.g. the resilient poor – see the idea of 'physical-condition neglect' in Sen (1985) – or the grumpy rich, etc.), (iii) measurement errors. We assume that individual fixed effects – or a proxy based on time-invariant personality traits, in our case – can capture some of these unobservables and improve comparability.

with tax-transfer rules $\psi(.)$, wage rates w_{it}, unearned income μ_{it} and a set of individual characteristics ζ_{it} determining the level of disposable income y at any given level of leisure. We adopt modern techniques that address the presence of nonlinear taxation in function $\psi(.)$, which consists in discretizing potential work hours (Blundell et al, 2000, van Soest, 1995). In this approach, agents are assumed to face J pairs (y_{ijt}, l_{ijt}), $j = 1, ..., J$, and to choose the one maximizing utility.[11] As usual in this literature, the random component η^D_{ijt} is assumed to be i.i.d. and follow an extreme value type I (EV-I) distribution, such that the probability to observe individual i choosing alternative j at time t has an explicit conditional logit form that is directly used to construct the likelihood for ML estimations. This random term is not just capturing observational or optimization errors, it is also part of the utility attached to option j by individual i but unobserved by the analyst. Hence, it must be accounted for when calculating metrics, as discussed later. Note, however, that because of the independence assumption, it cannot be interpreted as reflecting heterogeneous preferences (as mentioned, unobserved preferences are specifically modelled as part of the deterministic utility function).

3 Welfare Metrics

Welfare metrics are defined and calculated on the basis of estimated ordinal preferences as characterized by individual indifference curves. This section explains the principles guiding the definition of welfare metrics and the way to derive them (more technical details on these procedures are provided in Bargain et al. 2013). Hereafter, we drop t from the suscripts in order to simplify notations.

3.1 Overall Principles

We use welfare metrics as suggested in the growing literature on fair allocations (see Fleurbaey, 2006, 2008 for the axiomatic derivation and Thomson, 2011, for a survey). The first principle of the equivalence approach in the fair allocation theory is *nonpaternalism*, in the sense of a respect of individual preferences. It implies a rejection of Arrow's independence axiom, meaning that all the information about an individual's ordinal preferences, represented by her indifference curves, is taken into account. Then, the challenge of the fair allocation theory is to define equality when individuals have heterogeneous preferences $u_i(y, l)$ over the multiple

[11] We use a relatively thin discretization with $J = 7$ options corresponding to weekly work hours from 0 (inactivity) to 60 (overtime), with a step of 10 hours. For each option j, we specify decision utility as a function of (discrete) leisure l_{ijt} and income y_{ijt}. The latter is simulated as a function of the gross earnings generated when working $h_{ijt} = \tau - l_{ikt}$ hours and taking into account the taxes paid and benefits received at that income level (see Appendix A). Maximum available time τ is set to 80 hours per week in our application.

dimensions of a good life (two in our case), i.e. when indifference curves cross in the (y, l) space. The approach suggests to restrict interpersonal comparisons by applying a principle of *subset dominance* or *restricted dominance*. That is, it confines the dominance principle – i.e. a better bundle in all dimensions always reflects a better situation – to a reference set B_r, indexed by a parameter r, implicitly defined here by:

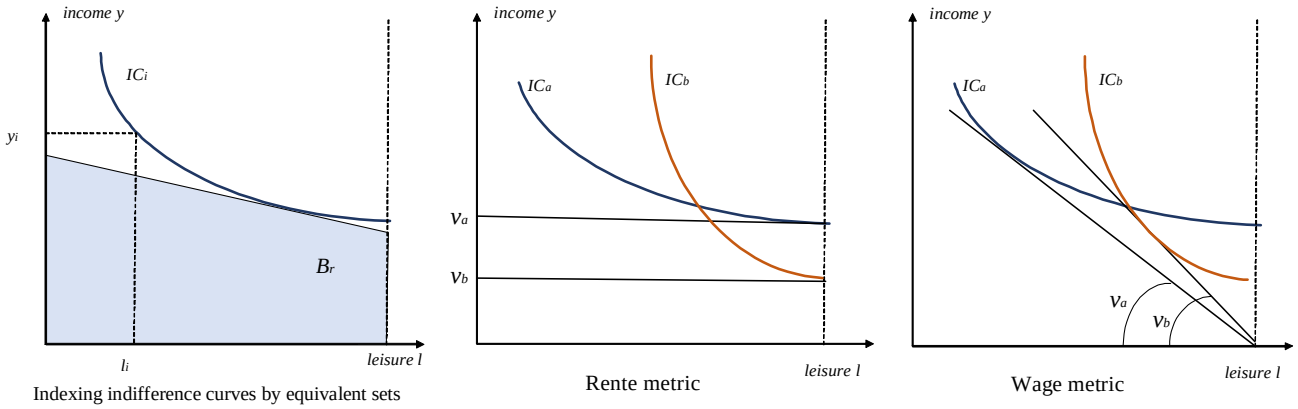
$$u_i\left(y_i, l_i\right) = \max[u_i(y, l)|(y_i, l_i) \subset B_r]. \tag{4}$$

In a relatively general formulation of the equivalence approach (Thomson, 1994), equivalent situations take the form of a collection of nested sets $(B_r)_{r\in\mathbb{R}_+}$, such that $r \leq r' \Leftrightarrow B_r \subseteq B_{r'}$. An individual's situation is evaluated by computing the equivalent set B_r, i.e. the set that would yield the same utility as her current situation. In our two-dimensional case, linearized budget curves, defined by their slope and intercept, allow indexing equivalent budget sets (see the first graph on Figure 1). Formally, the linearized budget constraint of an individual i choosing bundle (y_i, l_i) on a given indifference curve IC_i is written $y \leq \tilde{w}_i l + \widetilde{\mu}_i$, with virtual wage and nonlabor income $\tilde{w}_i$ and $\widetilde{\mu}_i$, so that the associated indirect utility function is:

$$v_i(\tilde{w}_i, \widetilde{\mu}_i) = \max[u_i(y, l)|y \leq \tilde{w}_i(\tau - l) + \widetilde{\mu}_i]. \tag{5}$$

The ordinal equity concept of *egalitarian-equivalence* (Pazner and Schmeidler, 1978) consists in retrieving a configuration where the actual allocation of individual bundles is Pareto equivalent to an egalitarian allocation indexed r, which defines the reference set B_r. 'Fair' allocations imply that this set needs not be arbitrary (Fleurbaey, 2008), i.e. it can be chosen according to explicit fairness criteria, which we now make explicit.

Figure 1: Welfare Metrics: Graphical Representation



3.2 Defining and Recovering Welfare Metrics

Definitions. In our setting, we consider reference sets based on the evaluation of individual situations according to hypothetical, linear budget constraints, as indicated above. Our welfare metrics belong to the domain of *responsibility-sensitive egalitarianism*, an approach that helps to rank individuals when their outcomes differ because of differences both in endowed circumstances and in individual preferences (Fleurbaey 2008). We give priority to the *compensation principle*, which states that for given preferences, inequalities arising from endowed circumstances – e.g. innate ability – should be removed. In the domain of income-leisure preferences, it means that people should be compensated for differences in their productive abilities, as the latter are deemed non-responsibility factors.[12] Then, it is possible to vary the degree of neutrality with respect to work preferences: reference wages or reference incomes can be defined according to specific priors that hold people more or less responsible for their work preferences.

Rent Metric. To simplify the exposition, we will focus on a specific metric where the reference parameter r is the wage rate $\widetilde{w}$ set at equal level for all. We consider a polar case where $\widetilde{w} = 0$ for all (less radical cases are examined in Bargain et al., 2013). Thus, interpersonal comparison is conducted in a counterfactual situation where inequalities from productivity differences are ignored. With this *Rent metric*, we must search the nonlabor income level μ_i^* that allow each individual to reach her/his current utility level. Note that in this situation, the same unearned income for all would lead to equal welfare for all, i.e. differences in preferences are neutralized. In other words, with the Rent metric, people are *held minimally responsible for their work aversion* so that actual differences in outcome due to preferences – and not only those due to responsibility factors (wages) – should be compensated. It is illustrated by the second graph of Figure 1. The metric is the vertical intercept of the actual indifference curve in the case of well-behaved preferences.[13] Individual b (work averse) is deemed worse off than a (industrious), so a-to-b redistribution is justified.

Wage Metric. For sensitivity check, we will also adopt a somewhat polar case where the reference parameter r will be nonlabor income $\widetilde{\mu}$, set equally to 0 for all. With the *Wage metric*, the money metric is going to be the wage level w_i^* allowing each individual to reach her/his current utility level. As shown on the last graph of Figure 1, the metric is the slope of

[12] This is a rather strong assumption made for practical reasons here (and in line with the long tradition of second-best optimal policy design). Hourly wages depend, to some extent, on past decisions regarding the accumulation of human capital and, hence, on individual preferences and past efforts.

[13] More generally, it is defined as a 'min criterion', i.e. the unearned income that would suffice if working did not bring any wage. This metric is extensively discussed in Fleurbaey and Blanchet (2013, Appendix A3).

the tangent through the origin at the actual indifference curve.[14] In an hypothetical world with equal wage rates and zero non-labor income, laisser-faire is the best possible allocation, i.e. remaining inequalities are solely due to differences in preferences and are legitimate (Fleurbaey and Maniquet, 2006). In that sense, the Wage metric can be interpreted as holding people maximally responsible for their willingness-to-work. Note, however, that its properties are not clear in terms of favoring the industrious or the work averse. On the last graph of Figure 1, b-to-a redistribution is justified – this situation will prevail, in particular, when indifference curves cross at maximal leisure – but one might easily construct a situation where two agents with crossing preferences are evaluated in the opposite direction.

Calculating Welfare Metrics. We can formally define the Rent metric as:

$$\nu_i^W(u, \mu^r = 0) = \min_{\tilde{w}_i}[\tilde{w}_i | v_i(\tilde{w}_i, \mu^r = 0) \geq u_i] \tag{6}$$

and the Wage metric as:

$$\nu_i^R(u, w^r = 0) = \min_{\widetilde{\mu}_i}[\widetilde{\mu}_i | v_i(w^r = 0, \widetilde{\mu}_i) \geq u_i]. \tag{7}$$

To obtain welfare metrics, we first retrieve individual indifference curves, implicitly defined as $\widehat{u}_i^m(y, l) = \overline{u}_i$ for each observation i in the data and based on estimated utility functions $\widehat{u}_i^m(\cdot, \cdot)$, for $m = E, D$.[15] Metrics are then calculated by iterative procedures, i.e. by incrementing hours using very small steps of 0.01 hours/week (note that this is different from moving across discrete categories $j = 1, ..., J$ as used for the labor supply estimation). The Wage metric is obtained by numerical search of the slope of the indifference curve that equals $\frac{y}{\tau - l}$. The Rent metric is simulated as the minimum unearned income allowing us to reach the indifference curve (see detailed explanations in Bargain et al., 2013).

[14] This measure, introduced by Pencavel (1977), is taken up in a few applications, like the recent work of Ooghe and Peichl (2010), and is grounded in the fair allocation approach by Fleurbaey and Maniquet (2006).

[15] Because the derivation of welfare metrics requires tangency conditions, indifference curves must be based on the deterministic part of utility functions (see Appendix A1 about the respect of monotonicity and concavity in the empirical application). This is not an issue for subjective preferences: random terms η_{it}^E do not vary with hour alternatives, and metrics can be calculated at the observed choice so that $\overline{u}^E = \widehat{u}_i^E\left(y^{obs}, l^{obs}\right)$. For revealed preferences, as indicated above, stochastic components η_{ijt}^D are unobserved attributes of work alternatives j (discrete choice formulation) and explain why certain hour options are sometimes chosen despite not yielding the highest deterministic utility. The probabilistic nature of the model leads to a frequency distribution of hour choices (rather than to a perfect prediction of the observed choice), which is taken into account when calculating the metrics. We simply calculate $\overline{u}^D$ as an average utility over the discrete hour alternatives weighted by their estimated probabilities. Similar approaches are also possible. Decoster and Haan (2014) calculate the welfare metric at each discrete hour, and average over all hours using estimated hour probabilities as weights. Bargain et al. (2013) compute the expected utility over many draws of the sets of EV-I terms, keeping each time the utility level attained at the optimal choice. A sensitivity analysis by these authors shows that the different methods lead to very similar welfare orderings.

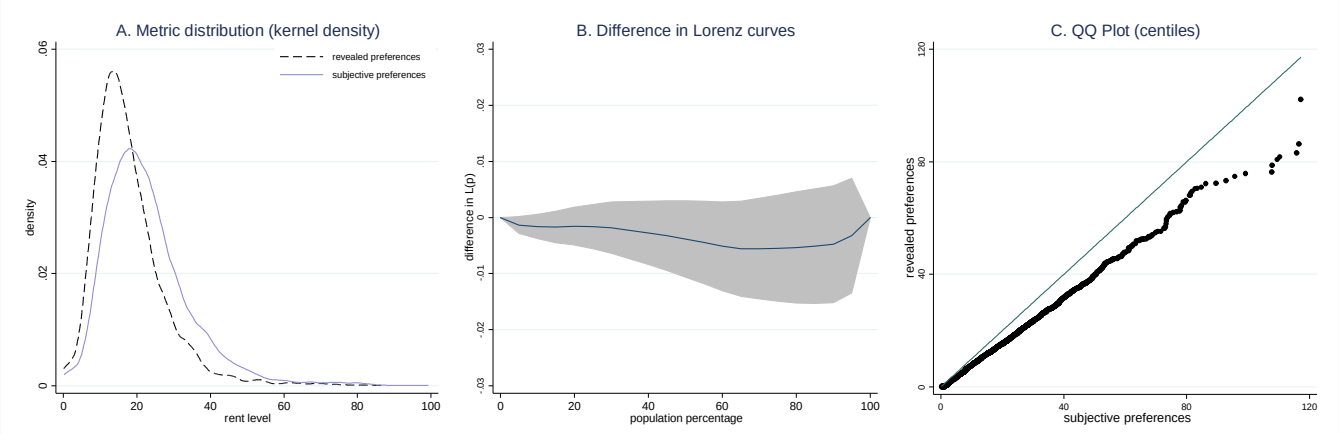
4 Results

We present the results in three steps, with the objective of assessing whether the way we capture preference heterogeneity makes a difference for distributional welfare analysis. We start with the implications for standard inequality analyses. We move to our main contribution, namely a direct confrontation of the two welfare orderings, investigating the possible explanatory channels for the reranking of individual situations. Finally, we bring this comparison to the level of the poorest, checking whether preference estimations affect the definition of the worst off. Baseline results focus on the Rent metric while we consider the alternative measure (Wage metric) in sensitivity analyses.

4.1 Welfare Inequality Analysis

We start with standard inequality measures, yet using a broader welfare concept than income. The densities of money metric utility obtain with revealed versus subjective preferences are depicted on the first graph of Figure 2. Kernel distributions look rather similar and log-normal. Choice-based welfare levels (dashed line) are slightly more concentrated while SWB-based welfare levels (solid line) are more right-skewed. An alternative representation could be the c.d.f. or the Lorenz curves. Thus, in the intermediary graph of Figure 2, we show the difference in Lorenz curves derived from revealed versus subjective preferences. The distance between the Lorenz curves is never significant along the cumulated distribution. We also find marginal differences in inequality measures (Gini, variance).

Figure 2: Welfare Distributions (Rent Metric) using Revealed vs. Subjective Preferences



Anticipating on the next section, we can directly compare the distributions to one another using quantile-quantile plots (QQ Plots) on the last graph of Figure 2. They compare two distributions by plotting their centiles against each other. If the two distributions are similar, the points will approximately lie on the 45° line. Our graphs show much overlap for rent values below around £30 per week, which correspond to the vast majority of observations (i.e. the lower 80% of the distribution). Upper quantiles show more dispersion according to subjective preferences, which is consistent with the first graph on kernel densities. A key point is that, even if money metrics diverge in levels, there is hardly any difference in terms of ranks. Thus, for inequality or relative poverty measures, both methods would lead to similar conclusions – at least when individual variation is averaged up at centile level. The next section prolongs this comparison at the individual level while focusing on potential welfare reranking.

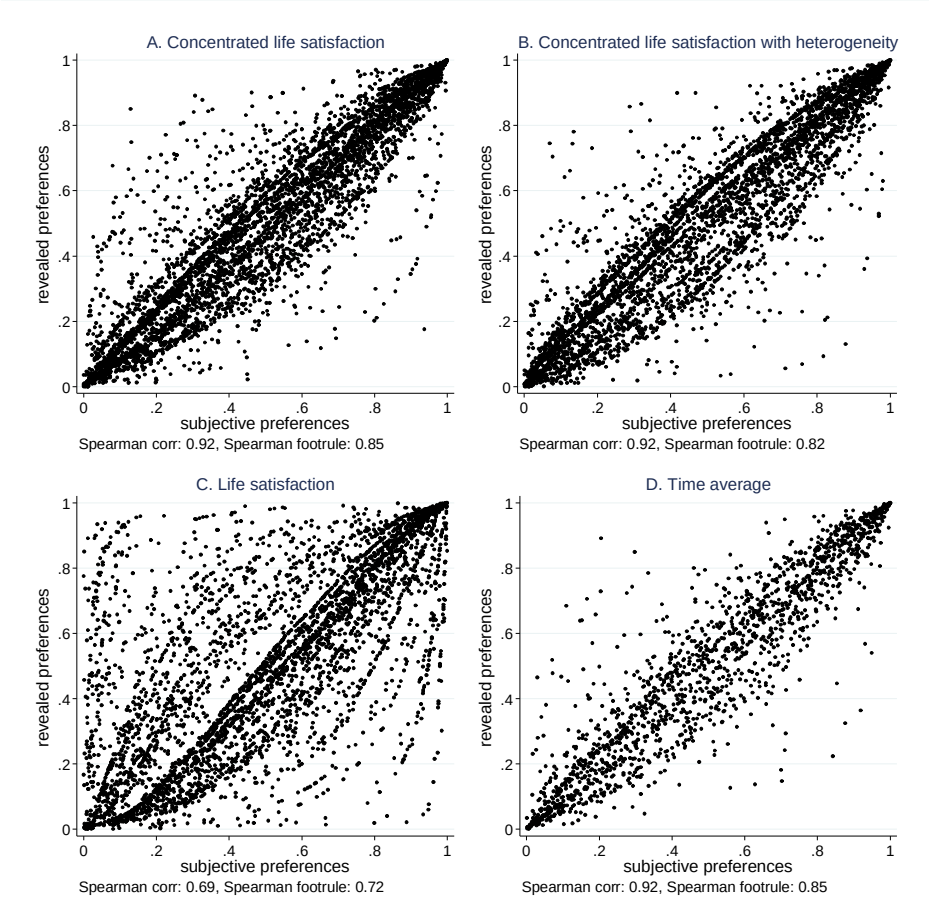
4.2 Reranking: General Results

We move to our core results whereby we directly compare the ranks of each observation according to revealed versus subjective preferences. This comparison is more demanding than in the previous section because the anonymity property is removed. We characterize the extent of reranking in a series of graphs that also indicate two summary indices, namely the Spearman rank correlation (a function of the sum of squared distances between ranks) and the Spearman footrule (the sum of absolute distances between ranks).

General Characterization. For the overall sample, results are presented in Figure 3. We plot individual ranks according to revealed preferences against ranks according to subjective preferences. For the latter, the baseline makes use of the concentrated income-leisure satisfaction (first graph). The correlation is relatively high, with a Spearman's ρ (footrule) of .92 (.85). A relatively basic test of whether the two distributions are similar can be performed with procedures dealing with two dependent distributions. The Wilcoxon signed-rank test is a non-parametric test precisely used when comparing two matched samples, or repeated measurements on a single sample, to assess whether their population mean ranks differ. It tests the equality of matched pairs of observations. It turns out that we cannot reject that the two distributions are similar (p-value of .985).

Sensitivity Checks. We acknowledge the fact that the difference between estimations based on behavior and those based on SWB may not just reflect true differences between revealed and subjective preferences. Part of it could indeed be due to some noise in the different measures (notably SWB) and to measurement errors: we provide an extended sensitivity analysis along

Figure 3: Welfare Rank Correlation (Rent Metric): Whole Sample



Note: for the Rent metrics, these graphs compare the welfare ranks obtained with revealed vs. subjective preferences, i.e. income-leisure ordinal preferences from actual choices vs. from the SWB experienced at these choices. Preferences are modelled using box-cox utility functions with preference heterogeneity (male, age, education, presence of young children, London, non-white, migrant, conscientious, neurotic).

these lines.[16]

First, we study the robustness of our results to alternative definitions of what the appropriate SWB measure could be. In Figure 3 (2nd graph), we proxy V_{it}^E using a concentrated measure whereby the relative weights on income and leisure satisfactions are heterogeneous, i.e. when $\widehat{V}_{it}^E = \widehat{\gamma}_{it}^y S_{it}^y + \widehat{\gamma}_{it}^l S_{it}^l$ with $\widehat{\gamma}_{it}^s = \widehat{\gamma}_0^s + \widehat{\gamma}_1^{s\prime} x_{it}$ for $s = y, l$ (the set of demographics x_i is the same as preference shifters in the structural models). The degree of reranking is hardly changed.[17] In Figure 3 (3rd graph), the dispersion increases when using the overall life satisfaction measure in place of our concentrated measure, i.e. $\widehat{V}_{it}^E = S_{it}$. This is expected given that the general satisfaction question implicitly covers many dimensions of well-being and, hence, adds considerable noise to our welfare characterization. Yet this check is important because the other dimensions possibly relate to the income-leisure tradeoff (for instance, the satisfaction with housing and distance to the working place.

Next, recall that we pool several years of a panel. This choice is mainly driven by the attempt to get estimates as precise as possible and, also, because identification of the empirical model relies on time-variation in socio-fiscal rules (see Appendix A.1). Yet, we wonder if having the same persons repeated several times in our reranking measures has some influence on the results, especially if measurement errors affect welfare measures over time. Figure 3 (4th graph) reports welfare comparisons when collapsing observations into time-average welfare levels for each person in our sample. It turns out that the picture is very similar to the baseline (Spearman rank correlation of .92). We also address the question of measurement errors directly. We suggest an application of the bias correction formula for Spearman's ρ as suggested by Kitagawa et al. (2018). In Figure 4, we report bias-corrected estimates for a realistic range of possible values of the error variance, namely between 0 and 10% of the variance of the welfare measures. For zero errors, the bias-corrected Spearman's ρ is our baseline correlation of 0.92. It then gets closer to 1 for an error variance above 5%.[18] Whether our data could contain this level of error is unknown but this is surely far below any plausible upper bound.[19]

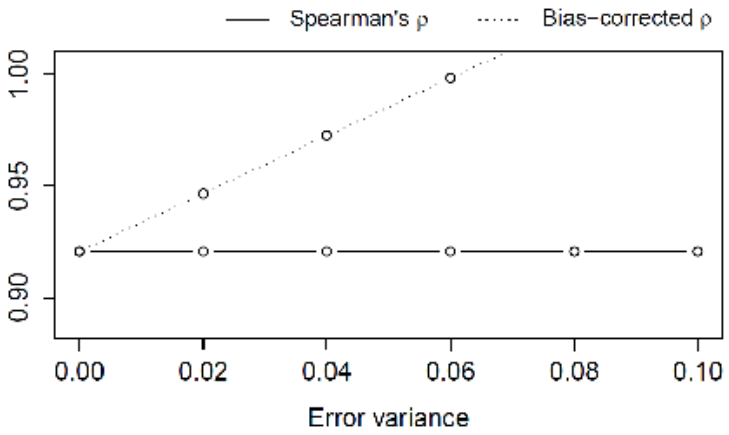
[16] Other sources of bias may relate to different treatments of the random terms, which, by construction, cannot be perfectly symmetrical in the two approaches. Also, we extensively discuss in Appendix A the fact that preference estimations based either on behavior or SWB may be biased by unobservables. We use several tools to address this issue: spatial and temporal variation in net wages, taste-shifters including psychological traits, random preference-for-leisure parameters. Still, one can never exclude that remaining biases cause some of the observed differences between preferences derived from both methods.

[17] We have also tried more flexible specifications than the linear form in equation (1), namely the addition of interaction terms between S_{it}^y and S_{it}^l (the coefficient of which proved insignificant) and/or quadratic terms. The results are again very similar (unreported).

[18] The bias-correction procedure increases the variance of the estimator, but our sample size is sufficiently large for this to be negligible.

[19] Using panel information, we find a correlation of .80 between an absolute welfare measure (i.e. before taking

Figure 4: Error-Corrected Rank Correlations at Different Error Variances (Rent Metric)



Note: Spearman rank correlation corrected from measurement error bias using Kitagawa et al. (2018). Spearman's ρ is reported for various hypothetical levels of error variance.

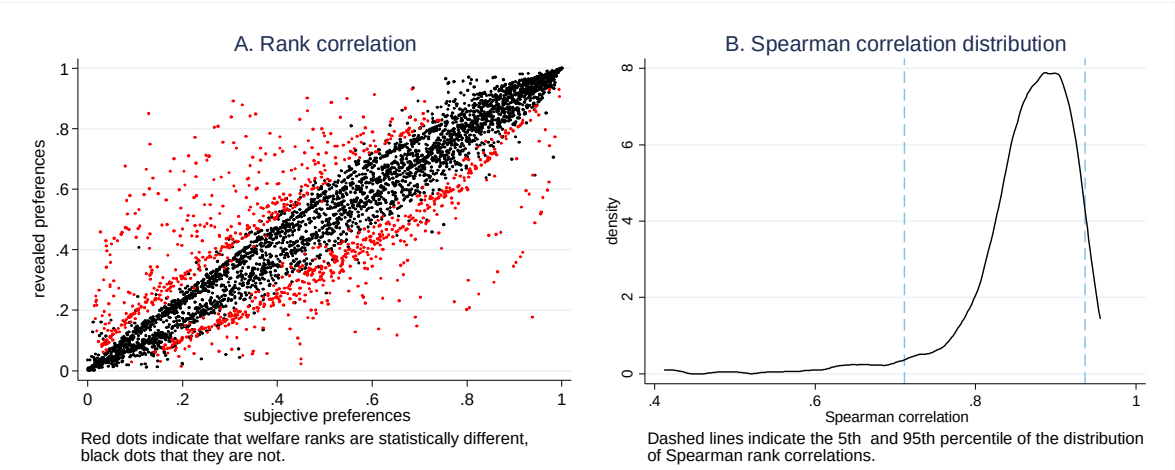
Inference. Given that there are two (unnested) models, one must fix one and check the role of sampling variance through the other. We fix the parameters of the choice-based model, which is the more precisely estimated one, and bootstrap preference parameters of the SWB-based model in its empirical distribution in order to derive a 95% confidence interval for the welfare ranks. In Figure 5, we first reproduce the main reranking graph but now indicate in black (red) the observations for which welfare ranks based on revealed preferences and subjective preferences are not (are) significantly different. As expected, cases where the welfare difference is not significant are close to the 45° line while those where it is gravitate further away from the line. The former represent 77% of the sample and yield a Spearman rank correlation of .98. The second graph of Figure 5 depicts the distribution of Spearman rank correlations obtained by bootstrapping preference parameters. It turns out that 95% of the rank correlations lie in a relatively narrow interval around the point estimate, as indicated by the dash lines.

4.3 Reranking: Heterogeneity and Tentative Interpretations

The previously analysis suggests that the observed reranking may reflect genuine differences between revealed and subjective preferences. This is all the more plausible as reranking is not pure noise but can be explained by observed characteristics, as shown hereafter. We also attempt to interpret these differences and their implications for welfare analyses.

ranks) and its lag. Hence, in the extreme case where time variation was entirely due to errors, the error variance would be around 20% of the variance of the observed welfare measures.

Figure 5: Inference for Welfare Rank Correlations (Rent Metric)



Note: Graphs based on bootstrapped preference parameters to generate confidence intervals for welfare ranks. First graph: ranks not statistically different in 77% of the cases. When insignificant (significant) difference, Spearman correlation is 0.98 (0.57), Spearman footrule is 0.91 (0.55).
Second graph: bootstrapped distribution of Spearman rank correlations.

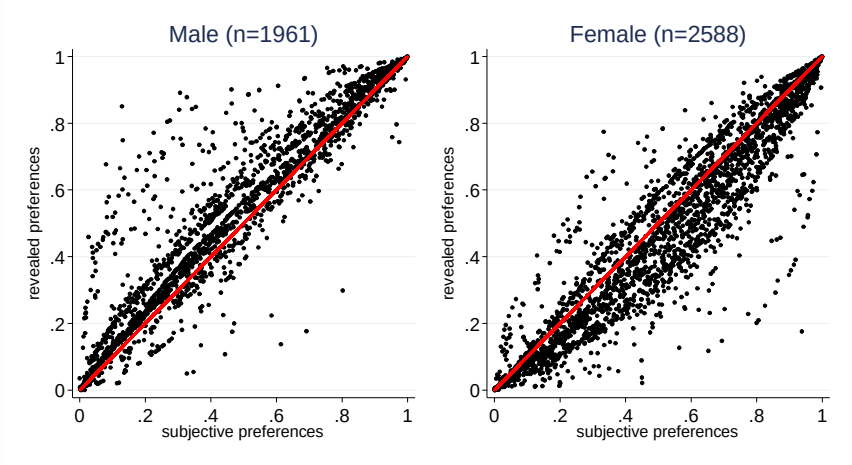
Reranking by Broad Groups. Reranking may be driven by certain aspects of preference heterogeneity. Thus, we start with a comparison of individual welfare ranks based on revealed versus subjective preferences when considering broad population groups, i.e. focusing on the main dimensions used in our taste shifters. We can characterize *within-group reranking* using ranks redefined among observations of the same group (for instance among males). The extent of reranking sometimes vary across subgroups: it is for instance larger among Londoners than in the rest of the country, or among highly education or those with children (detailed graphical results are presented in Online Appendix Figure B.2).

The contribution of each group to *overall reranking* can be investigated using population ranks (see Figure B.3). Similar differences can be seen but a more stricking observation is the specific asymmetries that come out. We examplify it for gender differences in Figure 6 below. We see that men (women) are deemed better off with revealed (subjective) preferences than with subjective (revealed) preferences, i.e. they tend to be concentrated above (below) the 45° line. These patterns can intuitively be explained by average indifference curves, as depicted in Figure 7 for the whole sample (left) or by gender (right). The first graph shows an overlap between revealed and subjective ordinal preferences, as discussed in Akay et al. (2015). In the second, 'revealed' indifference curves are relatively flat for men, rationalizing high working hours in this group. 'Subjective' indifference curves are steeper, revealing that the experience of working long hours implies larger compensation. The reverse reasoning applies to women. In

particular, forced under-employment may create distress (Clark and Oswald, 1994) – in Figure 7, it would explain a higher income compensation at zero work hours with the SWB approach compared to 'revealed' preferences. In brief, men (women) tend to work 'too much' ('too little') from a SWB perspective, so they are deemed better off when revealed (subjective) preferences are used with the Rent metric.

In the Online Appendix, Figure B.1 also show that the highly educated, those in London or the above-40 tend to over-work while those with young children tend to under-work according to SWB. These trends transpire in the estimates of our structural models, as reported in Table A.1: we see that the coefficients on leisure are negative for men or highly educated in the labor supply model while being insignificant in the SWB regression; the coefficients on London or above-40 are positive in both models but larger in the SWB estimation. Another way to capture these differences will be suggested later on, namely by regressing the difference in welfare ranks based on revealed versus subjective preferences on a broad set of individual characteristics including the taste shifters used in the models.

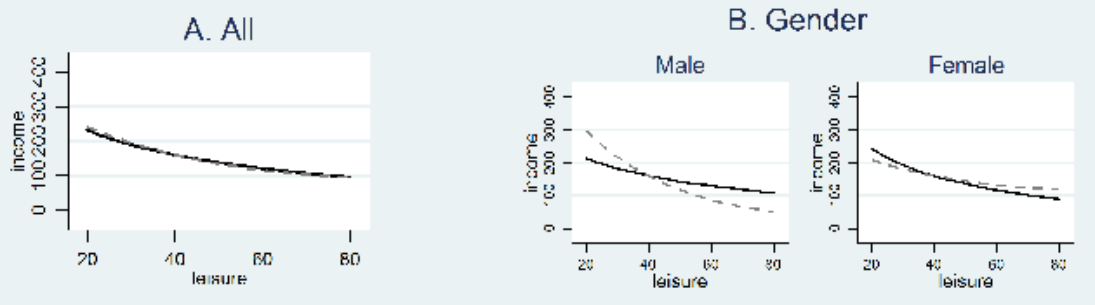
Figure 6: Welfare Rank Correlation (Rent Metric) by Gender, using Overall Ranks



Note: for the Rent metric, the graph compares welfare ranks with revealed versus subjective preferences, i.e. income-leisure ordinal preferences from actual choices versus from SWB experienced at these choices. Observations are grouped by gender type, using overall ranks.

Interpreting Reranking. While the previous analysis interprets the direction of the divergence, the underlying factors explaining differences between revealed and subjective preferences are complex. There are (at least) three types of factors potentially explaining the dissonance. First, there may be *constraints* at the individual level (e.g. health-related restrictions) or due

Figure 7: Indifference Curves with Revealed vs. Subjective Preferences



Note: solid (dash) lines indicated indifference curves for revealed (subjective) preferences. Indifference curve representations on these graphs are obtained using estimated parameters of the income-leisure utility functions, overall or for particular groups (e.g. women), and averaging individual indifference curves drawn through a common point, defined as $(\overline{y}(40), 40)$.

to the labor market (rationing, discrimination), which prevent people from reaching their optimal choice. The current state of our models is not able to 'reveal' genuine preferences from choices. Second, people possibly make *suboptimal choices*. This aspect is extensively investigated through numerous experiments in the behavioral economics literature, exploring different dimensions of suboptimality such as 'projection errors' (Loewenstein et al., 2003) or 'focusing illusions' (for instance giving too much importance to income compared to other aspects of a good life, cf. Kahneman et al., 2006). The third category is of a somewhat opposite nature: actual decisions may be more relevant than SWB if they reveal *other life goals* than the pursuit of personal satisfaction (as we measure it), for instance moral objectives.[20]

There are many difficulties. First, it may be difficult to conceptually distinguish these broad factors one from another. For instance, alternative life objectives may be associated with constraints (e.g. moral obligations to care for a sick mother, leading to 'under-work' according to SWB, or to financially support some relatives, leading to 'over-work'). Suboptimal behavior in terms of SWB (e.g. workaholism) may be seen as constraints (influenced by 'bad' norms)

[20] Another aspect is the role of dynamics and forward-looking decisions, including intertemporal substitution in labor supply choices (e.g. work harder now to save for a future time when productivity declines). The structural models used to retrieve revealed and subjective preferences are both static. Thus, the labor supply model can be seen as misspecified or based on strong (intertemporal separability) assumptions. The SWB model can also be characterized as misspecified if the current situation only partly correlates with instantaneous income-leisure satisfaction (i.e. if people who currently work hard and earn a little are nevertheless happy because they think prosperity is around the corner). Yet, SWB often pertains to myopic attitudes. Several studies attempt to show the extent to which people make systematic prediction errors regarding the future impact of choices/events on their life satisfaction, partly because of unforeseen adaptation (Loewenstein et al., 2003, Frijters, 2000, Frijters et al., 2009, Benjamin et al., 2012, Odermatt and Stutzer, 2015).

or alternative goals (honor, recognition, leaving an inheritance, etc.) that should be respected. Moreover, it seems impossible to disentangle them in a non-experimental set-up (for experimental approaches extracting some of the explanatory channels, see Fleurbaey and Schwandt, 2015, or Benjamin et al., 2012). Hereafter, we only suggest a descriptive analysis based on a regression of the welfare rank gaps, defined as the distance between a person's rank based on revealed preferences and her rank based on subjective preferences. Table 1 reports the results. Welfare rank gaps are estimated on the taste shifters used in the models and a few additional variables. These observed characteristics explain a substantial fraction of reranking: the R^2 reaches .39. Most factors are statistically significant and carry intuitive interpretations in line with the previous discussion. For instance, we find again that being a man (woman) contributes positively (negatively) to welfare rank gaps, as it leads to over-work (under-work) from a SWB perspective.

The first six variables are individual characteristics that cannot be changed – gender, age, ethnicity, psychological traits, etc. – and, hence, could be seen as non-responsibility factors. Yet, each of them may be related to several of the broad factors outlined above (constraints, suboptimality, other life goals) and characterize different degrees of free choice. Single women may face labor market constraints (as analyzed by Petrongolo, 2004, in the British context) or anticipate social norms regarding family-work balance. Single men can be socially pressurized to 'make a career' or seen as overly ambitious. Elderly workers may implicitly feel they should be more compensated than they are when working long hours. Non-white or migrants tend to under-work according to SWB, which might be due to labor market discrimination, or because of different norms (i.e. perceiving less of a need for compensation when actually working long hours).

The second column adds variables on which individuals may act to 'some' extent, but again the degree of responsibility is hard to establish. The highly educated or those in London may be led, through local norms, to demanding jobs,[21] while the low-skilled or those outside London may have less labor market opportunities – or may experience genuine work aversion due to the bad quality of the jobs they can find. People with bad health or those who have experienced long term unemployment in the past tend to under-work from a SWB perspective.[22] Those

[21] Hamermesh and Slemrod (2008) point to workaholism as an issue affecting the high skilled primarily, generated by biased beliefs about the well-being effects of work. Loewenstein et al. (2003) argue that individuals fail to appreciate how habit formation will affect future preferences and show that such a 'projection bias' might create a tendency to repeatedly increase labor and decrease leisure relative to earlier plans.

[22] The individual responsibility regarding these possibly inherited conditions is again a difficult question. Trannoy (2016) writes: "In the lifespan, maybe we can claim that the degrees of freedom of an individual are more important but still the analyst has to cope with the dependency of the trajectory of the individual to initial conditions. An individual starting with a long spell of unemployment just due to bad luck will have a

facing long commuting may be constrained by job availability or housing prices, or simply make irrational choices (as shown in Stutzer and Frey, 2008). Finally, having young children may lead to situations where work does not pay, notably when wage prospects are low and childcare costs are high.[23] We decompose the effect according to a question about whether pre-school children suffer if mothers work. The presence of children is consistent with a strong sense of under-work from a SWB perspective for women who disagree with this statement (and an opposite effect for those who think that children suffer).

4.4 A Focus on the Worst-Off

Limited Dissonance in the Identification of the Worst-off. We finally check the profile of the most deprived in our sample. This characterization is especially relevant from a policy perspective, when aiming to target the worst-off in a society. Similar exercises have been conducted in other studies that attempt to compare welfare measures.[24] For simplicity, we define the most deprived (equivalently: worst off, poor) as the bottom quintile of the welfare metric distributions. Note that this is different from standard poverty analyses that rely on poverty lines: the latter add another degree of arbitrariness while our approach allows comparing a group of the same size for both preference elicitation methods. We find a high degree of overlap, namely 80%, between those identified as worst off according to revealed preferences and those identified as such according to subjective preferences. The Rent metric gives a non-zero value to leisure so that the income-poor are not necessarily worst off. The overlap between income poverty and being worst off according to revealed (subjective) preferences is 60% (58%).

Profiles of the Worst Off. Table 2 suggests a portrait of the worst off. We report the mean characteristics of the bottom quintile defined according to income or to the Rent metric with revealed or subjective preferences. Since welfare metrics give a weight to leisure, the welfare-poor have higher incomes, a lower leisure time and a higher income-leisure satisfaction

stigma which will take time to be rubbed out."

[23] The UK is often described as a country with little support for maternal employment due to little public childcare provision, pushing maternal workforce into inactivity or low paid part-time employment (see Viitanen, 2005, for instance).

[24] In particular, Decancq and Neuman (2015) confront a variety of measures of the "good life". They show a high degree of reranking, and almost no correlation in the definitions of the worst off, when using current measures available in the literature. Decanq et al. (2015) for Russia also find low overlap between worst-off definitions according to income, life satisfaction and equivalent income. Given our previous results and the fact that we focus on a bidimensional welfare measure (income-leisure), we expect to find more overlap than in these studies. Carpantier and Sapata (2016) are in a similar situation. They also focus on income-leisure preferences, using the revealed preference approach only but a larger variety of fairness criteria. They find a great overlap in the identity of the worst-off across these criteria.

Table 1: Factors contributing to Welfare Rank Differences

Dependent var. : welfare rank gap	
Male	10.29***
	(0.281)
Over 40	2.498***
	(0.285)
Non-white	-8.152***
	(1.250)
Migrant	-6.513***
	(0.886)
Conscientious	-4.303***
	(0.296)
Neurotic	-0.425
	(0.280)
Poor health	-1.066***
	(0.317)
Past long term unemployment	-2.531***
	(0.727)
High education	4.281***
	(0.351)
London	17.82***
	(0.532)
Excessive commuting	-1.593***
	(0.280)
Presence of young child	
x child doesn't suffer if mother works	-5.052***
	(1.204)
x child suffers if mother work	3.237***
	(0.775)
Observations	4,549
Adjusted R-squared	0.393

Linear regressions of the welfare rank difference (welfare rank based on revealed preferences - welfare rank based on subjective preferences). Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

than the income-poor. The income-poor are more likely to be women, low educated and single parents, which is consistent with low labor market outcomes in this group. For our main comparison, i.e. the welfare-poor according to revealed versus subjective preferences, we report significant differences in the last column. Consistently with the high overlap rate previously reported, only few characteristics are significantly different. The worst-off group is more often composed of women and non-Londoners when using revealed rather than subjective preferences. As seen in Table 1, these were the main characteristics explaining welfare rank gaps – and they consistently matter for the bottom of the distribution. Men and Londoners are associated with a higher labor market participation, rationalized as low work aversion and deemed better off with revealed preferences than with subjective preferences when using the Rent metric.

Table 2: Characteristics of the Most Deprived (Rent Metric)

	Income	Rent metric		
		Rev. pref.	Subj. pref.	Δ
Disposable income	117.40	147.78	154.73	
	(36.46)	(68.51)	(67.58)	
Worked hours	13.48	28.16	29.40	
	(16.67)	(18.26)	(17.59)	
Income-leisure satisfaction	4.57	4.66	4.69	
	(0.84)	(0.82)	(0.83)	
Male	0.27	0.34	0.48	***
	(0.45)	(0.47)	(0.50)	
Over 40	0.52	0.58	0.57	
	(0.50)	(0.49)	(0.50)	
High education	0.05	0.08	0.11	
	(0.22)	(0.27)	(0.31)	
Child 0-2	0.38	0.13	0.12	
	(0.49)	(0.34)	(0.32)	
London	0.06	0.07	0.14	***
	(0.23)	(0.26)	(0.35)	
Non-white	0.01	0.01	0.01	
	(0.09)	(0.09)	(0.09)	
Migrant	0.02	0.02	0.02	
	(0.14)	(0.15)	(0.14)	
Conscientious	0.28	0.28	0.25	
	(0.45)	(0.45)	(0.43)	
Neurotic	0.53	0.53	0.49	
	(0.50)	(0.50)	(0.50)	

Notes: income is in pounds per week, hours are weekly, satisfaction is a weighted average of financial and leisure satisfactions on a 1-7 scale. Standard deviations in brackets. Δ: *, **, *** indicates significant difference in mean characteristics of the worst-off between revealed and subjective preferences at the 10%, 5%, 1% significance levels respectively.

4.5 Alternative Welfare Metrics

We finally check whether the previous conclusions are broadly preserved when moving to a polar case in terms of ethical priors. With the Wage metric, people are held maximally responsible for their work aversion – the metric gives more weight to leisure. The main results can be found in Online Appendix B.3. As shown in Figure B.4, the degree of reranking is a bit larger than with the Rent metric (the Spearman rank correlation is now .76, the Spearman footrule .70). The main reason goes as follows. For those not working, indifference curves based on revealed and subjective preferences cross at zero work hour so that there is no welfare difference between the two preference elicitation methods according to the Rent metric while there can be large differences according to the Wage metric. Inference calculations show that 65% of the welfare rank gaps are not significant, giving a Spearman correlation of .94 in these cases.

As discussed in section 3.2, crossings of indifference curves tend to give reversed welfare classifications. We indeed see that the asymmetry in terms of reranking is broadly reversed, as examplified in Figure B.5 in the case of gender. Those who tend to work 'too much' from a SWB perspective – e.g. men – are often deemed worse (better) off when revealed (subjective) preferences are used. Consistently, the regression of welfare gaps gives opposite signs for most factors compared to Figure 1 (unreported). Since the extent of reranking is larger with the Wage metric, these factors play a larger role: the R^2 goes as high as .65.

Finally, with the Wage metric, we find an overlap of 63% between the groups of individuals identified as worst off according to revealed preferences and those identified as such according to subjective preferences. A few variables significantly explain the mismatch including male and london, as for the Rent metric, but also age, the presence of young children and personality traits (Figure B.1) – all in the opposite direction compared to the Rent metric.

5 Summary and Concluding Discussion

The literature tends to show that for standard decisions in life (work choices), there is an overall congruence between decision and experienced utility (e.g., Benjamin et al., 2012, 2014, Fleurbaey and Schwandt, 2015). The present paper brings this question to the field of welfare analysis by characterizing the implications of using preferences derived from life satisfaction questions upon income and leisure (subjective preferences) rather than derived from actual labor supply choices (revealed preferences). Preference estimations are used to derive money metrics based on a 'fair allocation' approach in which the compensation principle prevails. We find that the correlation between welfare ranks based on revealed versus subjective preferences is high. This result holds for different subjective well-being (SWB) measures, when accounting

for measurement errors, and for different ethical views about how much people should be compensated due to differences in work preferences.

A broad set of observable characteristics explains a substantial amount of the remaining gaps in welfare ranks. The direction of the reranking pertains to the fact that specific groups tend to works 'too much' or 'too little' from a SWB perspective. For most of the characteristics, their contribution to reranking intuitively relates to general factors including health or labor market constraints, suboptimal decisions and life goals that differ from the pursuit of individual SWB. Unfortunately, it is far beyond our capacity to identify which of these factors prevail. Nonetheless, we can derive some implications for welfare analyses, as follows. *First*, if constraints prevail, "revealed" preferences cannot reflect the desired income-leisure balance, while subjective preferences may get closer to the true preferences (at least, to the extent that income-leisure satisfaction does not internalize these constraints). This is the case for instance with some of the characteristics leading to a sense of under-work from a SWB perspective (e.g. discrimination or rationing affecting women, the low skills or nonwhite ethnic groups, or rational 'under-employment' for single mothers facing high cost of work due to expensive child-care). *Second*, dissonance due to other life plans than the mere pursuit of hedonic well-being (e.g. moral obligations) may be respected too – this may be the case of single mothers who believe their child suffers from them working or the case of ethnic groups with different work norms than what is offered on local labor markets (in both cases, these groups are characterized as over-working according to subjective preferences).[25] However, identifying these alternative life objectives from constraints seems difficult on the basis of non-experimental data (such as traditional demographic variables used in structural models or in life satisfaction estimations). *Third*, suboptimal choices have received enormous attention in psychology and behavioral economics. Their normative implications remain unclear. One may be tempted to hold people responsible for their choices (and use revealed preferences). Yet, the view that human choices are frequently irrational – and hence the limits of the revealed preference approach – is precisely the motivation underlying the SWB literature and seems in line with not holding people responsible for the underlying welfare function leading to bad decisions. This said, it is not clear to which extent income-leisure satisfactions used in our application are not themselves tainted by suboptimal views.

Two key implications derive from this discussion. The first one is normative: the choice of the preference elicitation method may not be independent from the choice of the ethical prior. In particular, if work decisions are strongly constrained by the labor market, e.g. some individuals tend to under-work from a SWB perspective, then: (i) one may not want to punish them by

[25] The normative debate about whether these preferences are justified or not – for instance inherited from 'bad' labor market equilibria – is open. Yet imposing subjective well-being as a norm seems equally arbitrary.

putting too much weight on the value of their leisure, so that the egalitarian reference wage should be set to low values – zero with the Rent metric – if the welfare analysis is based on actual choices to derive preferences;[26] or (ii) subjective preferences may be preferred to 'revealed' preferences if the society wants to hold people responsible for (apparent) work aversion, i.e. wants the Wage metric to be used.[27] The second implication is rather methodological. Even state-of-the-art labor supply estimation methods, as used here, are not necessarily able to incorporate job opportunities, rationing and work costs, i.e. to identify what pertains to individual preferences from what is due to institutional or demand side constraints.[28] In fact, discrepancies between 'revealed' and subjective preferences could actually been used as an original way to elicit labor market frictions (at least if we believe that income and leisure satisfaction questions do not internalize labor market constraints).[29]

Further research could also address the following points. First, if revealed preferences are affected by suboptimal choices that divert individuals from the maximization of their well-being, the interval between revealed and subjective preferences might be used to define incomplete preference relations – in the vein of Bernheim and Rangel (2009) and Fleurbaey and Schokkaert (2013) – and see whether distributional judgments can still be made on the basis of partial or-

[26] This is in line with Fleurbaey and Maniquet (2014) who suggest that if work aversion is partly due to non-responsibility factors, for instance low job quality (unpleasant, dangerous, etc.) for the unskilled, it may be "prudent or charitable" to choose a low value for the equivalent wage. After all, involuntary under-work may be viewed as reducing the agents' earning ability (Fleurbaey and Maniquet, 2006).

[27] The Laissez-Faire principle underlying the Wage metric is acceptable if individual preferences are fully respectable. This is not the case if they reflect external factors (e.g. constraints) but more debatable in cases discussed above (e.g. moral obligation to support relatives, workaholism due to social pressure, etc.).

[28] In the present study, we have discarded job seekers – deemed as involuntary unemployed – from the analysis. Nonetheless, labor constraints are still present among part-timers and involuntarily idle workers (e.g. discouraged workers or single mothers facing zero or negative gains from work). As noted, a fundamental difficulty is to identify demand-side and institutional constraints on the basis of standard characteristics observed in survey data. Exclusion restrictions are never satisfying, and more (quasi)experimental variation should be used in order to recover actual preference parameters. Recent approaches characterize frictions by comparing long term and short term adjustments, assuming people are less constrained in the long run (Chetty, 2012). Some studies have explicitly accounted for labor market rationing within labor supply models, for instance by modelling the probability of involuntary unemployment (e.g., Haan and Uhlendorff, 2013), the demand-side of the labor market (Peichl and Siegloch, 2012) or the distribution of job opportunities (see a modern account in Beffy et al., 2016, and Capéau et al. 2016). It is difficult, however, to account for all these aspects simultaneously – and the range of 'constraints' may be large: discrimination, rationing (e.g. productivity below minimum wage), frictional unemployment, discouragement, low-quality jobs, wrong belief about job opportunities, etc. As for work costs, they are also typically not identified on the basis of standard observed characteristics (cf. van Soest et al., 2012).

[29] In further work, more systematic characterization of these SWB-revealed frictions could be obtained for different countries and points in time in order to check if they are indeed correlated with the business cycle (ex: larger frictions in times of strong demand-side constraints).

derings. Second, one may check the implication of using revealed versus subjective preferences when aggregating welfare metrics in a social welfare function. The well-known issue that equivalent measures are not necessarily concave in income and, hence, may induce antiegalitarian policy implications (see Blackorby and Donaldson, 1988). This problem can be overcome, as recently suggested by Bosmans et al. (2017).[30]

References

[1] Akay, A., O. Bargain & H.X. Jara (2015): "Back to Bentham, Should We? Large-Scale Comparison of Experienced versus Decision Utility", ISER discussion paper.

[2] Bargain, O. (2012): "Decomposition analysis of distributive policies using behavioural simulations", International Tax and Public Finance, 19(5), 708-731

[3] Bargain, O., A. Decoster, M. Dolls, D. Neumann, A. Peichl, & S. Siegloch (2013): "Welfare, labor supply and heterogeneous preferences: Evidence for Europe and the US", Social Choice and Welfare, 41 (4), 789–817.

[4] Bargain, O., K. Orsini & A. Peichl (2014): "Labour supply elasticities: A complete characterization for Europe and the US", Journal of Human Resources, 49(3),723-838

[5] Beffy, M., Blundell R, Bozio A, Laroque G, & To M. (2016): "Labour supply and taxation with restricted choices", IFS working paper 15/02

[6] Benjamin, D., O. Heffetz, M. Kimball & A. Rees-Jones (2012): "What Do You Think Would Make You Happier? What Do You Think You Would Choose?" American Economic Review, 102(5), 2083–2110

[7] Benjamin, D., O. Heffetz, M. Kimball & A. Rees-Jones (2014a): "Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices", American Economic Review 104, 3498–3528.

[8] Benjamin, D., O. Heffetz, M. Kimball & N. Szembrot (2014b): "Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference", American Economic Review, 104(9), 2698–2735

[9] Bernheim, D., & A. Rangel (2009): "Beyond revealed preference: choice theoretic foundations for behavioral welfare economics". Quarterly Journal of Economics 124, 1, 51-104

[30] The fact that these metrics need not satisfy the Pigou-Dalton principle everywhere is not necessarily a strong argument against using them to construct a social welfare function that is less extreme than the maximin. Indeed, the violation of the Pigou-Dalton principle occurs only when indifference curves change shape when utility increases, in a way that makes the violation of the principle not so shocking.

[10] Blackorby, C. & Donaldson, D. (1988): "Money Metric Utility: A Harmless Normalization", Journal of Economic Theory 46: 120–129

[11] Bloemen, H.G. (2018), Collective Labour Supply, Taxes, and Intrahousehold Allocation: An Empirical Approach, Journal of Business and Economic Statistics

[12] Blundell, R., A. Duncan, J. McCrae, & C. Meghir (2000): "The labour market impact of the working families' tax credit", Fiscal Studies, 21 (1), 75-103

[13] Blundell, R., Duncan, A. & Meghir, C. (1998): "Estimating labor supply responses using tax reforms", Econometrica 66(4), 827–861

[14] Boadway, R. (2012): "Review of 'A Theory of Fairness and SocialWelfare' by Fleurbaey and Maniquet". Journal of Economic Literature, 50(2), 517–521

[15] Bosmans, K., K. Decancq and E. Ooghe (2017): "Who's afraid of aggregating money metrics?", mimeo

[16] Boyce, C.J. (2010): "Understanding Fixed Effects in Human Wellbeing", Journal of Economic Psychology, 31: 1-16.

[17] Capéau, B., Decoster, A., Dekkers, G. (2016), Estimating and Simulating with a Random Utility Random Opportunity Model of Job Choice Presentation and Application to Belgium, International Journal of Microsimulation, 9 (2), 144-191.

[18] Carpantier, J.F. and C. Sapata (2016): "Empirical Welfare Analysis: When Preferences Matter", Social Choice and Welfare Economics, 46(3), 521-542

[19] Chetty, R. (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Microand Macro Evidence on Labor Supply."Econometrica, 80, 969-1018

[20] Chiappori, P-A. and O. Donni (2011). Non-Unitary Models of Household Behavior: A Survey of the Literature. In: A. Molina (ed.), Household Economic Behaviors, Berlin: Springer.

[21] Clark, A. and A. Oswald (1994): "Unhappiness and Unemployment", Economic Journal, 104, 424, 648-659

[22] Clark, A. E., Frijters, P. & M. Shields (2008): "Relative income, happiness and utility: an explanation for the Easterlin paradox and other puzzles", Journal of Economic Literature, 46(1), 95-144

[23] Clark, A., C. Senik, & K. Yamada (2015): "When Experienced and Decision Utility Concur: The Case of Income Comparisons", IZA DP No. 9189

[24] Dagsvik, J.K. and S. Strøm (2006): "Sectoral labour supply, choice restrictions and functional form," Journal of Applied Econometrics, 21 (6), 803–826.

[25] Decancq, K., M. Fleurbaey and E. Schokkaert (2014): "Happiness, equivalent incomes, and respect for individual preferences", forthcoming in Economica

[26] Decancq, K., M. Fleurbaey & E. Schokkaert (2015): "Inequality, income, and well-being". in: Anthony B. Atkinson & François Bourguignon (eds.) Handbook on Income Distribution, Volume 2A. Elsevier. pp. 67-140

[27] Decancq, K. & D. Neumann (2016): "Does the choice of well-being measure matter empirically?" in: Matthew Adler and Marc Fleurbaey (eds.), Oxford Handbook of Well-Being and Public Policy. Oxford University Press

[28] Decoster, A. & P. Haan (2015): "Empirical welfare analysis with preference heterogeneity", International Tax and Public Finance, 22 (2), 224-251

[29] Ferrer-i-Carbonell, A. & P. Frijters (2004): "How Important is Methodology for the Estimates of the Determinants of Happiness?", Economic Journal, 114(497), 641-659

[30] Ferrer-i Carbonell, A., B.M.S. van Praag & I. Theodossiou (2010): "Vignette Equivalence and Response Consistency: The Case of Job Satisfaction", IZA DP 6174

[31] Fleurbaey, M. (2006): "Social welfare, priority to the worst-off and the dimensions of individual well-being". In: Farina F, Savaglio E (eds) Inequality and economic integration. Routledge, London

[32] Fleurbaey, M. (2008), Fairness, Responsibility and Welfare, Oxford University Press.

[33] Fleurbaey, M. (2009): "Beyond GDP: The Quest for a Measure of Social Welfare", Journal of Economic Literature 47: 1029–1075.

[34] Fleurbaey M. & F. Maniquet (2006): "Fair Income Tax", Review of Economic Studies, 73(1), 55-83

[35] Fleurbaey, M. & F. Maniquet (2007): "Help the low skilled or let the hardworking thrive? A study of fairness in optimal income taxation", Journal of Public Economic Theory, 9, 467–500

[36] Fleurbaey M. & F. Maniquet (2011), A Theory of Fairness and Social Welfare, Cambridge University Press

[37] Fleurbaey M. & F. Maniquet (2014): "Optimal Taxation Theory and Principles of Fairness", Core Discussion Paper.

[38] Fleurbaey, M., & Blanchet, D. (2013), Beyond GDP, Measuring Welfare and Assessing Sustainability, Oxford University Press.

[39] Fleurbaey, M. & H. Schwandt (2015): "Do People Seek to Maximize Their Subjective Well-Being?", IZA DP No. 9450

[40] Fleurbaey, M. and E. Schokkaert (2013): "Behavioral welfare economics and redistribution", American Economic Journal: Microeconomics, 5(3):180-205, 201

[41] Frijters, P. (2000). "Do individuals try to maximise satisfaction with life as a whole" Journal of Economic Psychology, 21, 281-304.

[42] Frijters, P., Greenwell, H., Shields, M.A., & J.P. Haisken-DeNew (2009). "How rational were expectations in East Germany after the falling of the wall?", Canadian Journal of Economics, 42(4), 1326-1346

[43] Glaeser, E.L., J.D, Gottlieb & O. Ziv (2016): "Unhappy cities", Journal of Labor Economics, 34(S2), S129 - S182

[44] Haan, P. & A. Uhlendorff (2013): "Intertemporal Labor Supply and Involuntary Unemployment, Empirical Economics, 44, 2, 661-683

[45] Hamermesh, D.S. and Slemrod, J.B. (2008): "The economics of workaholism: we should not have worked on this paper", B.E. J. Econ. Anal. Policy 8 (1)

[46] Hoynes, H. W. (1996): "Welfare transfers in two-parent families: Labor supply and welfare participation under AFDC-UP", Econometrica, 64(29), 295-332

[47] Jacquet, L. and D. Van de Gaer (2011): "A comparison of optimal tax policies when compensation or responsibility matter", Journal of Public Economics, 95 (11–12), 1248–1262

[48] Jara, X.H. and E. Schokkaert (2017): "Putting measures of individual well-being to use for ex-ante policy evaluation", Journal of Economic Inequality, 15, 4, 421–440

[49] Kahneman D., A. Krueger, D.A. Schkade, N. Schwarz, A. Stone (2006): "Would You Be Happier If You Were Richer? A Focusing Illusion", Science, 312: 1908-1910

[50] Kitagawa, T., M. Nybom and J. Stuhler (2018): "Measurement error and rank correlations", Cemmap Working Paper CWP28/18

[51] Lise, J. and S. Seitz (2011): "Consumption Inequality and Intra-household Allocations", Review of Economic Studies, 78, 1, 328–355

[52] Loewenstein, G., O'Donoghue, T., & M. Rabin (2003): "Projection bias in predicting future utility", Quarterly Journal of Economics, 118; 1209–1248

[53] Ng, Y-K. (1997): "A case for happiness, cardinalism, and interpersonal comparison", Economic Journal, 107(445): 1848-1858

[54] Odermatt, R. and A. Stutzer (2015). "(Mis-)Predicted Subjective Well-Being Following Life Events", IZA DP No. 9252

[55] Ooghe E. & A. Peichl (2011): "Fair and efficient taxation under partial control: theory and evidence", Economic Journal

[56] Pazner, E., & D. Schmeidler (1978): "Egalitarian Equivalent Allocations: A New Concept of Economic Equity". Quarterly Journal of Economics, 92, 671-686.

[57] Peichl, A. and S. Siegloch (2012): "Accounting for labor demand effects in structural labor supply models", Labour Economics 19 (1), 129-138

[58] Pencavel, J. (1977): "Constant-Utility Index Numbers of Real Wages", The American Economic Review 67(2): 91–100.

[59] Perez-Truglia, R. (2015): "A Samuelsonian Validation Test for Happiness Data", Journal of Economic Psychology, 49, 74–83

[60] Petrongolo, B. (2004): "Gender Segregation in Employment Contracts", Journal of the European Economic Association, 2 (2/3, P&P), 331-345

[61] Ravallion, M., & Lokshin, M., (2001): "Identifying Welfare Effect from Subjective Questions", Economica, 68: 335-357

[62] Schokkaert, E., D. Van de Gaer, F. Vandenbroucke & R.I. Luttens (2004): "Responsibility sensitive egalitarianism and optimal linear income taxation", Mathematical Social Sciences, 48 151– 182

[63] Sen, A. 1985. Commodities and capabilities. Amsterdam: North-Holland

[64] Senik, C. (2005). "Income distribution and well-being: what can we learn from subjective data?", Journal of Economic Surveys 19: 43–63

[65] Stiglitz, J., Sen, A. & Fitoussi, J.-P. (2009): "Report by the Comission on the Measurement of Economic Performance and Social Progress", Technical Report.

[66] Stutzer, A., and B.S. Frey (2008): "Stress that doesn't pay: the commuting paradox", Scandinavian Journal of Economics, 110 (2), 339-366

[67] Thomson W. (1994): "Notions of equal, or equivalent, opportunities", Social Choice and Welfare 11: 137–156

[68] Thomson W. (2011): "Fair allocation rules"" in K.J. Arrow, A.K. Sen, and K. Suzumura (eds.), Handbook of Social Choice andWelfare, vol. 2, Amsterdam: North-Holland

[69] Trannoy, A. (2016): "Equality of Opportunity: A progress report", ECINEQ working paper

[70] van Praag, B.M.S., P. Frijters & A. Ferrer-i-Carbonell (2003). "The Anatomy of Subjective Well-Being", Journal of Economic Behavior and Organization, 51, 29-49

[71] van Soest, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach", The Journal of Human Resources, 30 (1), 63-88

[72] van Soest, A., M. Das & X. Gong (2002): "A structural labor supply model with non-parametric preferences", Journal of Econometrics, 107 (1-2), 345-374.

[73] Viitanen, T. (2005): "Cost of Childcare and Female Employment in the UK", Labour, 19 (Special Issue): 149-179

[74] Wichert, L. & W. Pohlmeier (2010): "Female Labor Force Participation and the Big Five", ZEW discussion paper 10-003

A Appendix

A.1 Models Specification

Specification of the Utility Functions. Both experienced and decision utilities are specified according to the box-cox form:

$$u_{it}^m\left(y_{it}, l_{it}\right) = \beta_y^m \left(\frac{y_{it}^{\lambda_y^m} - 1}{\lambda_y^m}\right) + \beta_l^m(x_{it}, \phi_i) \left(\frac{l_{it}^{\lambda_l^m} - 1}{\lambda_l^m}\right), \quad m = D, E$$

Used in recent welfare analyses (Decoster and Haan, 2014, and Bargain et al., 2013), box-cox utility allows easily checking that preferences are well-behaved, which facilitates the derivation of ordinal preferences (i.e., indifference curves) and the calculation of welfare metrics. Monotonicity and concavity conditions on consumption and leisure are satisfied if, respectively, β's are positive and λ's are in a range between 0 and 1. We check ex post that both conditions are fulfilled empirically for all our observations. More flexible forms could be used but tangency conditions are necessary for calculating welfare metrics.[31]

Preference heterogeneity across individuals is introduced as follows. Parameters on leisure vary linearly with taste shifters x_{it} and a normally distributed random term ϕ_i:

$$\beta_l^m(x_{it}, \phi_i) = \beta_{l0}^m + \beta_{l1}^{m\prime} x_{it} + \beta_{l2}^m \phi_i, \quad m = D, E$$

Vector x_{it} includes the following binary characteristics: male, age above 40, higher education, presence of children aged 0 to 2, living in London, non-white ethnic origin, migrant, above-average conscientiousness and above-average neuroticism. Unobserved preferences ϕ_i are dealt with using simulated maximum likelihood.

Budget Constraints. In both approaches, disposable income is computed according to the budget constraint $y_{it} = \psi_t(w_{it}h_{it}, \mu_{it}, \zeta_{it})$. Function ψ_t aggregates gross earnings $w_{it}h_{it}$ and unearned income μ_{it} into net income y_{it}, adding taxes and withdrawing benefits that depend on these income levels (benefit means-tested, tax brackets, etc.) and on individual characteristics ζ_{it} (tax credits or benefits being a function of family composition, for instance). It is approximated by numerical simulations using the tax-benefit rules of each period $t = 1, \ldots, T$. In the same way, we also predict $(y_{ijt}, \tau - h_{ijt})$ pairs for the $j = 1, \ldots J$ potential choices used in the labor supply model. To do so, we first estimate an Heckman-corrected wage equation (instrument is non-labor income and the presence of children aged 0-2) in order to predict wage rates w_{it} (wages are unobserved for non-workers). Then we numerically compute disposable

[31] An assessment of the box-cox functional form for labor supply behavior, compared to more flexible specifications, is suggested by Dagsvik and Strøm (2006).

income $y_{ijt} = \psi_t(w_{it}h_{ijt}, \mu_{it}, \zeta_{it})$ for the J discrete labor supply values of h_{ijt} (see Bargain et al., 2014).

Identification and Limitations. The econometric identification of ordinal preferences requires some discussions (see also Akay et al., 2015). For labor supply models, a well-known difficulty pertains to the role of omitted preference shifters that may affect both wage rates and work preferences. For instance, hard-working types may work a lot also because they tend to have higher wage rates, i.e. an underestimation of preferences for leisure. For the SWB model, a similar type of bias may exist. For instance, actual heterogeneity in work preferences may be correlated with other unobserved determinants of well-being. If hard workers are more likely to experience positive shocks to SWB, then the bias goes in the same direction as for labor supply, i.e. work aversion is underestimated. However, the bias could go the other way. We suggest three strategies to reduce these concerns. First, we account for individual heterogeneity x_{it} – notably relevant personality traits, i.e. conscientiousness and neuroticism – in work preferences and, for the SWB equation, in the separately additive term z_{it}. Second, we account for random preference-for-leisure parameters (note, however, that ϕ_i is normally distributed and, hence, cannot capture the true distribution of omitted variables). Finally, and most importantly, we use spatial and temporal variation in factors that affect the *net* wages. As used in the labor supply literature, it corresponds to spatial variation in tax-benefit rules (Hoynes, 1996) and time variation in these rules over 1996-2005 (i.e., tax-benefit reforms, as in Blundell et al., 1998). The period covered in our data includes quite much variation in tax-benefit rules to improve identification (see a detailed account in Akay et al., 2015). These approaches are the best we can do in the present setting. Nonetheless, one can never exclude that remaining biases cause some of the observed differences between preferences derived from both methods.

A.2 Models Estimation Results

Table A.1: Models' Estimates

Coefficients	Labor Supply	Income-Leisure Concentrated Satisfaction
Lambda income	0.143***	0.491***
	(0.0227)	(0.0833)
Lambda leisure	0.416***	0.444*
	(0.0154)	(0.239)
Income	3.100***	0.0499***
	(0.225)	(0.0146)
Leisure	1.816***	0.0567*
	(0.0632)	(0.0312)
x male	-0.472***	0.0284
	(0.0410)	(0.0304)
x over 40	0.0549	0.0394*
	(0.0385)	(0.0204)
x higher education	-0.277***	0.0178
	(0.0465)	(0.0242)
x young kid	1.748***	0.0332
	(0.187)	(0.0881)
x london	0.157**	0.163***
	(0.0766)	(0.0608)
x non-white origin	-0.238	-0.0867
	(0.169)	(0.134)
x migrant	0.257*	-0.0146
	(0.135)	(0.0862)
x conscientious	-0.211***	-0.0480**
	(0.0398)	(0.0236)
x neurotic	-0.00541	-0.00601
	(0.0385)	(0.0179)
Log-likelihood	-7128.00	-4604.01
#Obs	4,570	4,570

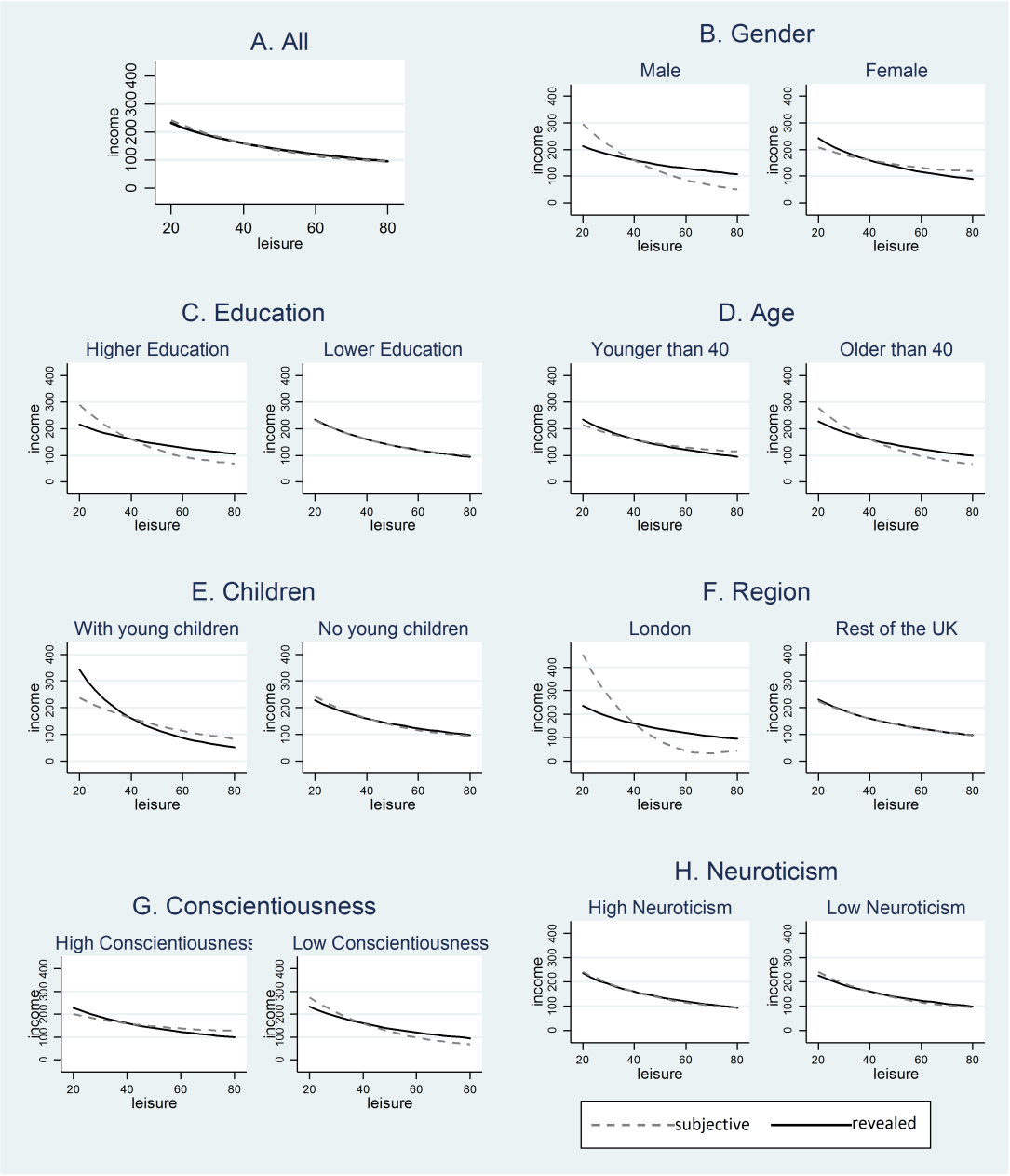
Box cox specifications. Subjective well-being equations (concentrated satisfaction) also include additively separable controls (same variables as in leisure interaction terms plus age squared, family size, health status, home ownership, all personality traits, region and year dummies). *, **, *** indicate 1%, 5% and 10% significance levels. Standard errors in parenthesis.

B Online Appendix

B.1 Indifference Curves by Broad Groups

We derive indifference curves in the income-leisure space for every individual in our sample. For the whole population (graph A) or within each group (graphs B-H), we average individual indifference curves through a common point set at 40 hours of leisure and $\overline{y}(40)$ (the sample mean net income at this leisure level).

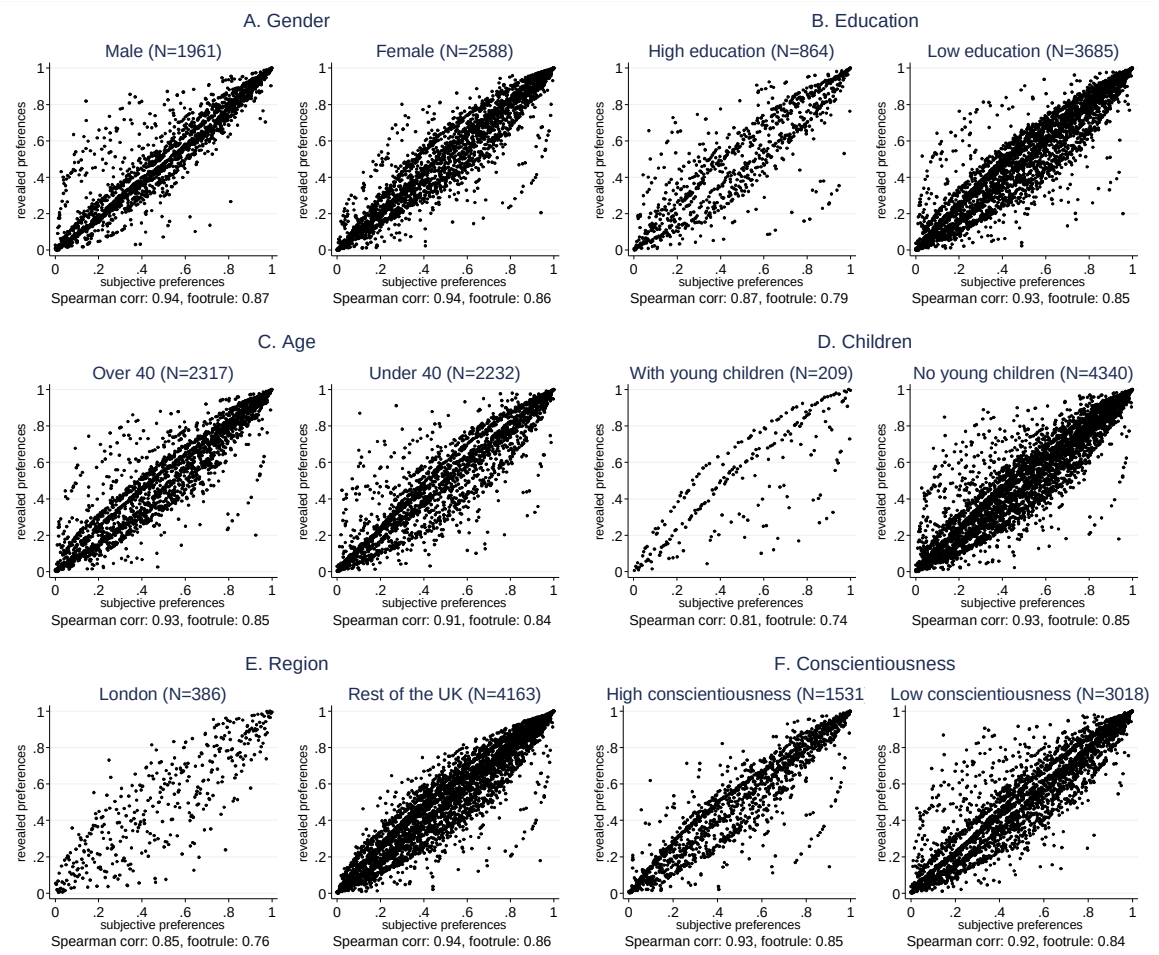
Figure B.1: Indifference Curves with Revealed vs. Subjective Preferences



Note: solid (dash) lines indicated indifference curves for revealed (subjective) preferences. Indifference curve representations on these graphs are obtained using estimated parameters of the income-leisure utility functions, overall or for particular groups (e.g. women), and averaging individual indifference curves drawn through a common point, defined as $(\overline{y}(40), 40)$. On the graphs, weekly leisure points range from 20 to 80 hours, corresponding to weekly work hours from 60 (overtime) to 0 (inactivity).

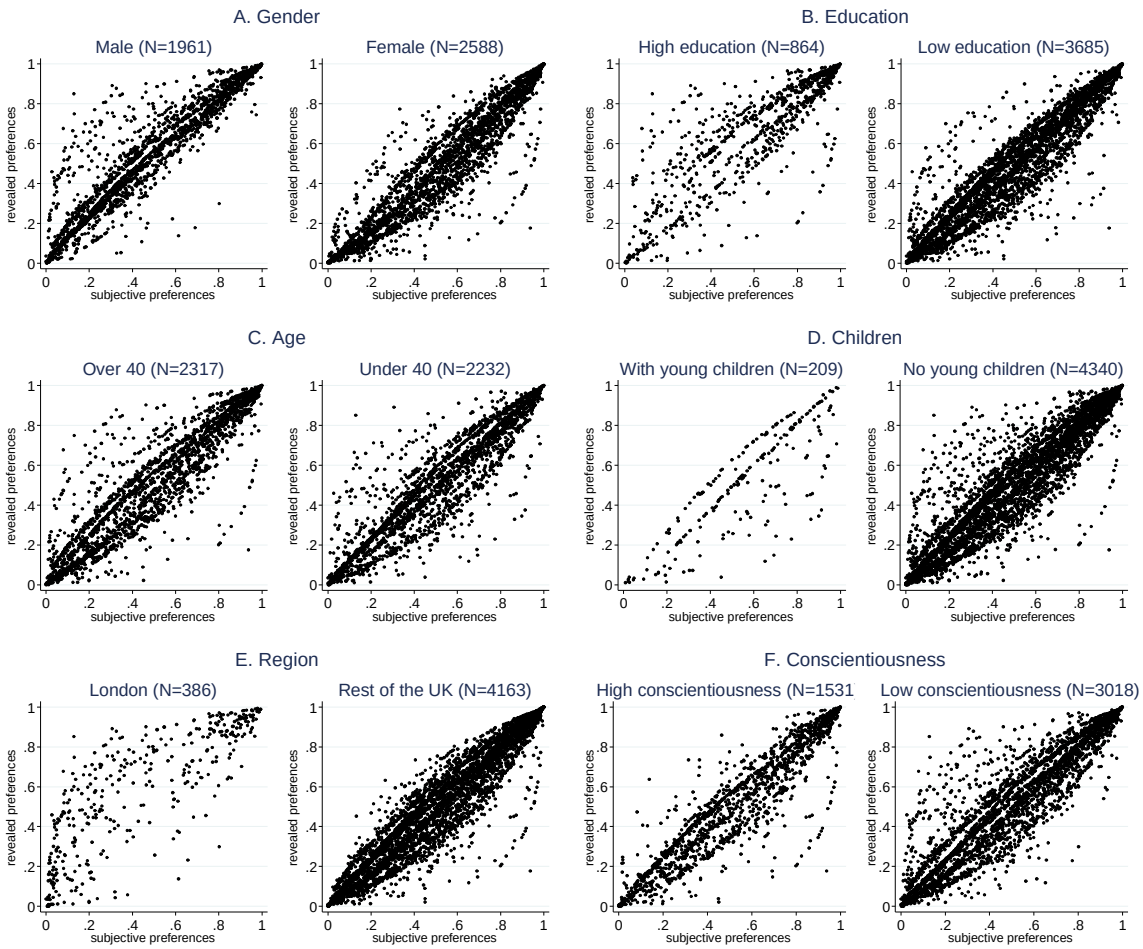
B.2 Reranking by Broad Groups

Figure B.2: Welfare Rank Correlation (Rent Metric) by Groups, using Group-specific Ranks



Note: for the Rent metric, the graph compares welfare ranks with revealed versus subjective preferences, i.e. income-leisure ordinal preferences from actual choices versus from SWB experienced at these choices. Observations are grouped by demographic type, using group-specific ranks.

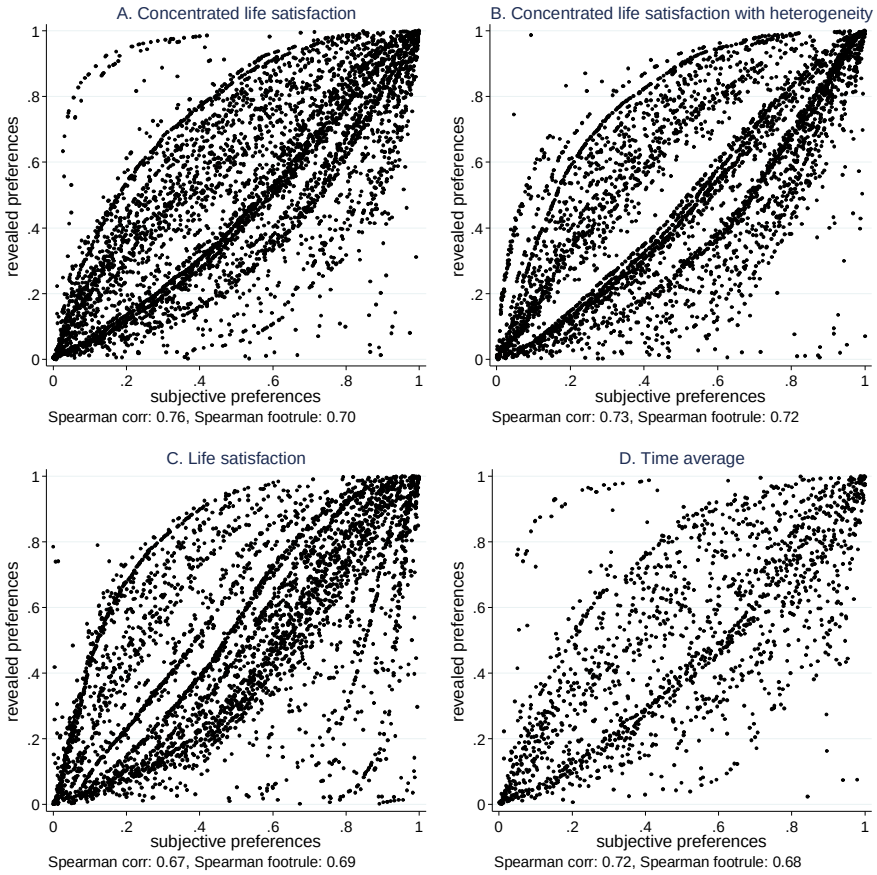
Figure B.3: Welfare Rank Correlation (Rent Metric) by Groups, using Overall Ranks



Note: for the Rent metric, the graph compares welfare ranks with revealed versus subjective preferences, i.e. income-leisure ordinal preferences from actual choices versus from SWB experienced at these choices. Observations are grouped by demographic type, using overall ranks.

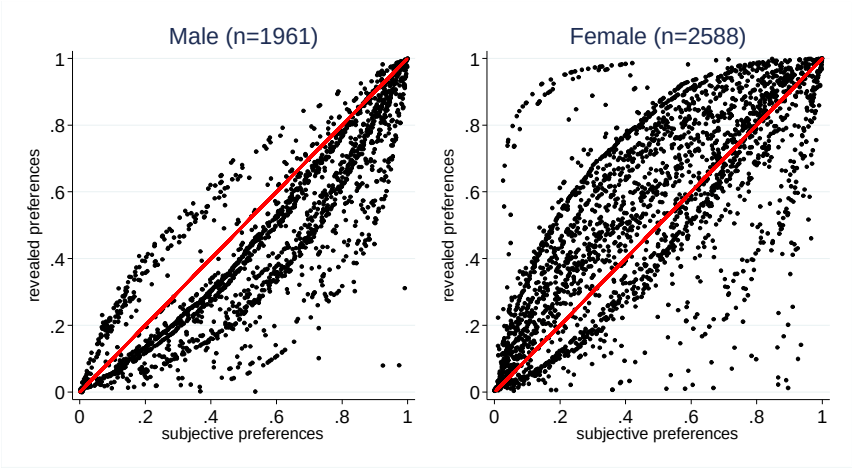
B.3 Results for the Wage Metric

Figure B.4: Welfare Rank Correlation (Wage Metric): Whole Sample



Note: for the Wage metrics, these graphs compare the welfare ranks obtained with revealed vs. subjective preferences, i.e. income-leisure ordinal preferences from actual choices vs. from the SWB experienced at these choices. Preferences are modelled using box-cox utility functions with preference heterogeneity (male, age, education, presence of young children, London, non-white, migrant, conscientious, neurotic).

Figure B.5: Welfare Rank Correlation (Wage Metric) by Gender, using Overall Ranks



Note: for the Wage metric, the graph compares welfare ranks with revealed versus subjective preferences, i.e. income-leisure ordinal preferences from actual choices versus from SWB experienced at these choices. Observations are grouped by gender type, using overall ranks.

Table B.1: Characteristics of the Most Deprived (Wage Metric)

	Income	Wage		
		Rev. pref.	Subj. pref.	Δ
Disposable income	117.4	152.49	156.73	
	(36.46)	(71.57)	(60.16)	
Worked hours	13.5	30.46	27.02	***
	(16.67)	(17.85)	(18.37)	
Income-leisure satisfaction	4.57	4.73	4.68	
	(0.84)	(0.83)	(0.84)	
Male	0.27	0.51	0.33	***
	(0.45)	(0.50)	(0.47)	
Over 40	0.52	0.56	0.41	***
	(0.50)	(0.50)	(0.49)	
High Education	0.05	0.11	0.09	
	(0.22)	(0.31)	(0.28)	
Child 0-2	0.38	0.08	0.23	***
	(0.49)	(0.28)	(0.42)	
London	0.06	0.05	0.01	***
	(0.23)	(0.22)	(0.10)	
Non-white	0.01	0.00	0.00	
	(0.09)	(0.07)	(0.06)	
Migrant	0.02	0.02	0.02	
	(0.14)	(0.13)	(0.13)	
Conscientious	0.28	0.34	0.52	***
	(0.45)	(0.47)	(0.50)	
Neurotic	0.53	0.49	0.56	**
	(0.50)	(0.50)	(0.50)	

Notes: income is in pounds per week, hours are weekly, satisfaction is a weighted average of financial and leisure satisfactions on a 1-7 scale. Standard deviations in brackets. Δ: *, **, *** indicates significant difference in mean characteristics of the worst-off between revealed and subjective preferences at the 10%, 5%, 1% significance levels respectively.