

DISCUSSION PAPER SERIES

IZA DP No. 10908

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

Alpaslan Akay
Olivier B. Bargain
H. Xavier Jara

JULY 2017

DISCUSSION PAPER SERIES

IZA DP No. 10908

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

Alpaslan Akay

University of Gothenburg, IZA and LISER

H. Xavier Jara

University of Essex and ISER

Olivier B. Bargain

*Aix-Marseille University, CNRS, EHESS and
IZA*

JULY 2017

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

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

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 different ethical priors regarding work preferences, we compare the welfare rankings obtained with both methods. The correlation in welfare ranks is high in general and very high for the 60% of the population whose actual choices coincide with subjective well-being maximization. For the rest, most of the discrepancies seem to be explained by labor market constraints among the low skilled and underemployment among low-educated single mothers. Importantly from a Rawlsian perspective, the identification of the worst off depends on ethical views regarding responsibility for work preferences and the extent to which actual choices are constrained on the labor market.

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

Keywords: fair allocation, money metric, decision utility, experienced utility, labor supply, subjective well-being

Corresponding author:

Olivier Bargain
Aix-Marseille University
GREQAM
Av. Schuman
13100 Aix-en-Provence
France

E-mail: olivier.bargain@univ-amu.fr

* We are grateful to participants to seminars at the LSE, ISER, DIAL, LISER, AMSE, IZA, Sienna University and at the 2016 Royal Economic Society and HDCA conferences.

1 Introduction

The recent years have witnessed a resurgence of interest in the measurement of multidimensional welfare (see Stiglitz et al., 2009, and Fleurbaey, 2009). 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 the enlightening surveys of Senik, 2008, and Clark et al., 2008). Yet, this approach is fully welfarist in the sense that SWB is assumed to be a proxy for cardinal utility, which can lend itself to interpersonal comparison and aggregation. This assumption seems too strong for a large part of the economic profession.¹ At the same time, considerable progress has been made in the measurement of multi-dimensional welfare based on money metric utility. Notably, the ‘fair allocation’ theory suggest ways to construct welfare indices that only require information about ordinal (non-comparable) preferences – but nonetheless provide welfare metrics that are cardinal and comparable, just like ordinary income. Hence, these metrics can be used for distributional analyses and potentially for social welfare aggregation (Fleurbaey and Maniquet 2006, 2011).²

Crucially, to operationalize welfare comparisons based on ‘fair’ money metrics, it is necessary to retrieve ordinal preference heterogeneity from individual data. Several empirical implementations of the fair allocation theory have naturally relied on the ‘revealed preference’ approach.³ Other studies have originally suggested ways to infer ordinal preferences from SWB data.⁴ In

¹The most problematic aspect in terms of comparability is the fact that self-reported welfare reflects different levels of adaptation and aspiration. For instance, resilient poor may report high well-being levels while demanding rich may declare experiencing low satisfaction. In both cases, it cannot justify a policy redistributing to the latter or failing to address the poor conditions of the former (see Fleurbaey and Maniquet, 2014, Fleurbaey and Blanchet, 2013).

²With this approach, it is possible to rank individual situations when preferences differ while escaping from most of the standard criticisms about money metrics. Arrow’s impossibility is overcome by relaxing the Independence of Irrelevant Alternative and making use of all the information about an individual’s ordinal preferences. Interpersonal comparisons (dominance principle) are made in specific regions of the indifference set that are justified by precise ethical choices (Fleurbaey, 2008). This makes the choice of a reference set less arbitrary – or at least supported by explicit normative principles. In this paper, we shall rely on the *compensation principle* in the domain of income-leisure preferences. Reference wages/incomes will be defined according to specific priors that favor agents who are more or less work-averse.

³In particular, Bargain et al. (2013), Decoster and Haan, (2014) and Carpentier and Sapata (2016) have derived welfare metrics in the income-leisure domain. 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.

⁴Decancq 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

this paper, we ask *whether the way we elicit preference heterogeneity makes a difference for welfare analysis*. To the best of our knowledge, this question has never been addressed and our contribution is original in this respect. Nonetheless, this question is closely related to the literature comparing decision and experienced utility. It initiates from studies in behavioral economics and psychology, which have explored the possibility of basing economic appraisal on the measurement of experienced utility using small-scale experiments (Kahneman et al., 1997). More recently, the first explicit comparison has been suggested by Benjamin et al. (2012, 2014), who proxy experienced utility using SWB and decision utility using stated preferences in a tailor-made studies. Fleurbaey and Schwandt (2015) confront respondents with a broad range of life choices while Akay et al (2015) compare actual decisions in standard microdata with the SWB derived from these choices. These studies conclude to an *overall* congruence between decision and experienced utility and, when differences exist, provide relatively intuitive explanations for them.⁵ Our work further extends this question by asking whether revealed and subjective preferences lead to similar conclusions when used to rank people according to ethically-grounded money-metric welfare measures.

In the present study, we provide a first investigation that focuses on a bidimensional measure of welfare comprising income and non-market time (abusively called "leisure" in what follows). As in several of the aforementioned studies, we concentrate on a domain that is crucial for normative analyses. Indeed, this is the place where redistributive policies operate, as made clear in the long tradition of second best policy design and optimal taxation. The bulk of these literatures 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). Yet, new attempts to empirically measure welfare while accounting for preference heterogeneity are also welcome. We proceed here with the estimation of ordinal preferences that are consistent either with the labor supply decisions made by observed individuals in our British panel data or with the subjective experience they derive from these choices, as proxied by a combined

focusing on social progress and poverty respectively.

⁵Benjamin et al. (2012) show that most (but not all) individuals are able to predict their SWB at the moment of deciding about (hypothetical) job opportunities. Benjamin et al. (2014a) look at actual (rather than hypothetical) choices of residency, showing that SWB scores are correlated with the ranking of actual choices (even if the tradeoffs between aspects of residency tend to be different). Fleurbaey and Schwandt (2015) ask people if they can think of changes that would increase their SWB score. About 60% cannot think of an easy improvement, i.e. they feel as if they currently maximized SWB. Considering own income versus others' income, Clark et al. (2015) find similar relative concerns in happiness regressions and in hypothetical-choice experiments. Yet, more divergence is also found in recent studies based on job satisfaction (Ferrer-i-Carbonell et al., 2010), residential choice (Glaeser et al., 2016) or consumption (Perez-Truglia, 2015).

measure of income and leisure satisfactions. The former approach provides revealed preferences while the latter provides what we shall refer to as subjective preferences.

In Section 2, we present the data and a brief outline of the procedure used to elicit preference heterogeneity in both approaches. In Section 3, we define the welfare metrics and explain how to calculate them using estimated preferences. We focus on the ethical view of ‘responsibility-sensitive egalitarianism’ (Fleurbaey, 2008), taking up two polar cases whereby people are held minimally responsible (Rent metric) or maximally responsible (Wage metric) for their work aversion. Results are presented and discussed in Section 4. For each metric, we first compare the welfare distributions of equivalent rents/wages obtained with revealed versus subjective preferences. We then characterize the reranking from using one rather than the other type of preferences. The correlation in welfare orderings is high in around 60% of the population, for whom actual choices are consistent with SWB maximization. For the rest, most of the reranking is due to the low-skilled, who seem to work less than optimally, and from single mothers, who face large costs of work. Finally, we find a fairly large overlap between the groups identified as the worst-off individuals. Discrepancies crucially depend on ethical views regarding responsibility for work preferences and the extent to which work hours are constrained on the labor market. We end the paper with a thorough discussion about the implications of using particular combinations of ethical views and preference types, subsequent recommendations and new questions for future research.

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 literatures, yet with a special care of making functional forms similar in both approaches. 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 forms are slightly 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 welfare analysis at the individual level in couple households would require the estimation of the intrahousehold decision process, which is beyond the scope of our work. 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 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 person 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 4,560 person \times year observations.

Income and Leisure Time. The key variables for our analysis are working time and disposable income. Weekly working hours reported in the data are denoted h_{it} , for individual i in year t . Disposable income, denoted y_{it} , is calculated as $y_{it} = \psi_t(g_{it}, \mu_{it}, \zeta_{it})$, using reported gross earnings $w_{it}h_{it}$ (hourly wage rates $w_{it} \times$ work hours h_{it}), unearned income μ_{it} and individual characteristics ζ_{it} . Function ψ_t represents the aggregation of all incomes and the imputation of taxes and benefits, using numerical simulations of tax-benefit rules of each period $t = 1, \dots, T$. 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.).

We shall also use household and individual characteristics to model preference heterogeneity in our labor supply and SWB estimations. These variables include gender and age, being single, widowed or divorced, health status (categories very good health to very poor health), educational level (elementary school, high school or university), being a native or immigrant, ethnicity (simplified to white or non-white origin), number of household members (mainly children or elderly dependents), a dummy for the presence of young children (aged 0 to 2), living in London and personality traits (the so-called ‘big five’, on 1-4 scales: conscientiousness, neuroticism, openness, extraversion and agreeableness).

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 in 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 goods. Interestingly, however, the data contains 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” 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 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 $S_{it} = \gamma^y S_{it}^y + \gamma^l S_{it}^l + e_{it}$. We then use the predicted value, $V_{it}^E = \hat{\gamma}^y S_{it}^y + \hat{\gamma}^l S_{it}^l$, as our “income-leisure concentrated” SWB measure (our baseline). Note that we have also experimented alternative measures of V_{it}^E , namely nonlinear estimations for our ‘concentrated’ satisfaction, heterogeneity in coefficients γ , or just using general life satisfaction in place of the ‘concentrated’ measure, as discussed later.

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.1.⁸ Denote τ the maximum time available for work (or alternative activities), so that leisure is written $l_{it} = \tau - h_{it}$ for individual i in year t . The deterministic function of income and leisure that defines ordinal preferences over these two dimensions is written $u_{it}^m(y_{it}, l_{it})$, with $m = E$ (experienced utility) or D (decision utility). Estimations rely on the identity:

$$V_{it}^m = u_{it}^m(y_{it}, l_{it}) + \eta_{it}^m, \tag{1}$$

with a box-cox specification for function u_{it}^m , the parameters of which are allowed to vary with the characteristics of individual i at time t , namely dummies for gender, age above 40, higher education, presence of children aged 0 to 2, living in London, non-white ethnic origin,

⁶The data also contains information on mental health (the index from the General Health Questionnaire, GHQ-12) and answers to the happiness question. These alternative measures of SWB lead to relatively similar results regarding the estimation of ordinal income-leisure preferences (see Akay et al., 2015).

⁷Importantly, 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.

⁸See also the companion paper Akay et al. (2015), where estimations are used for a comparison of indifference curves between approaches (overall and by subgroups of observed characteristics).

migrant, above-average conscientiousness and above-average neuroticism.⁹ This specification is presented in detail and justified in Appendix A1. While the specification is common to both approaches, estimation methods and the assumptions underlying residuals η_{it}^m , $m = D, E$, are necessarily specific, as we now explain.

Estimation of Subjective Preferences. For preferences elicited from SWB measures, u_{it}^E , the main information used to estimate equation (1) is the concentrated income-leisure satisfaction index V_{it}^E assumed to proxy ‘experienced utility’, i.e. 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 in well-being responses. This comprises observable characteristics z_{it} corresponding to the usual determinants of well-being (cf. Clark et al., 2008),¹⁰ 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), and help to clean SWB measures from individual effects that prevent interpersonal comparison (see Fleurbaey and Blanchet, 2013). Utility V_{it}^E is treated as continuous and the model is estimated by maximum likelihood (to address the nonlinearity of the box-cox specification). Note that terms z_{it} and α_i , assumed additively separable from income-leisure preferences, aim to clean SWB levels from individual subjectivities (see Decancq et al, 2015).

Estimation of Revealed Preferences. For preferences revealed from individual choices, u_{it}^D , the required information is the labor supply choice, deemed optimal for individual i at time t . We adopt modern techniques that address the presence of nonlinear taxation in the budget constraint by discretizing work options (e.g., Blundell et al, 2000, van Soest, 1995). Precisely, agents are assumed to face (y_{ijt}, l_{ijt}) pairs, $j = 1, \dots, J$, and choose the one maximizing utility, so

⁹Among the personality traits, these two 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.

¹⁰Observed 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.

that ‘decision utility’ $V_{ijt}^D = u_{it}^D(y_{ijt}, l_{ijt}) + \eta_{ijt}^D$ must be evaluated only at each of the J options.¹¹ As usual in this literature, the random component η_{ijt}^D 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 maximum likelihood estimations.

3 Defining and Retrieving Welfare Metrics

Once estimations are performed, estimated utility functions $\widehat{u}_{it}^m(y_{it}, l_{it})$ are used to derive money metrics for $m = D$ and E in the ways explained hereafter. Money metrics are calculated for each individual i and period t in the data using her/his own heterogenous preferences \widehat{u}_{it}^m and assuming she/he reaches the utility levels $\bar{u}_{it}^m = \widehat{u}_i^m(y_i^{obs}, l_i^{obs})$, $m = D, E$, obtained at actual choices $(y_{it}^{obs}, l_{it}^{obs})$.

3.1 Overall Principles

We use welfare metrics as suggested in the growing literature on fair allocation (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 (and a rejection of Arrow’s independence axiom). It means that all the information about an individual’s ordinal preferences, represented by indifference curves, is taken into account. In our case, individuals are assumed to choose a bundle (y_i, l_i) resulting from the classic utility maximization problem:

$$(y_i, l_i) = \max [u_i(y_i, l_i) \mid y_i \leq \psi(w_i(\tau - l_i), \mu_i, \zeta_i), l_i \leq \tau]$$

with tax-transfer rules $\psi(\cdot)$ determining nonlinear budget sets $y_i \leq \psi(w_i h_i, \mu_i, \zeta_i)$. The challenge of the fair allocation theory is thus to define equality when individuals have heterogeneous preferences u_i over the multiple dimensions of a good life (i.e., in our simple two-dimensional case, when indifference curves cross in the (y, l) space).

In a relatively general formulation of the equivalence approach (Thomson, 1994), equivalent situations take the form of a collection of nested sets $(B_r)_{\lambda \in \mathbb{R}_+}$, 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

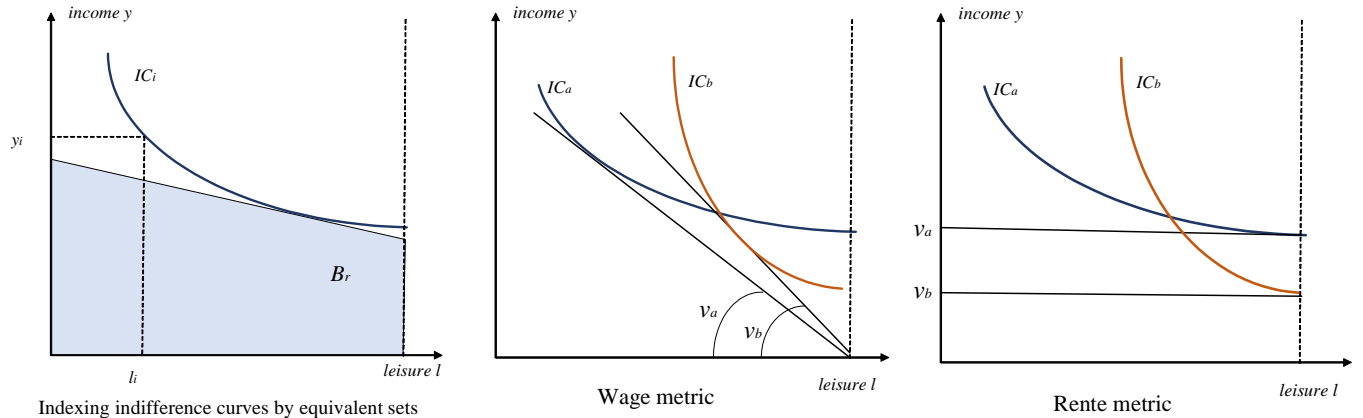
¹¹We use a relatively thin discretization with $J = 7$ options corresponding to weekly work hours from 0 (inactivity) to 60 (overtime), with a step of 10 hours. For each option j , we specify decision utility as a function of (discrete) leisure l_{ijt} and income y_{ijt} . The latter is simulated as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

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 $y \leq \tilde{w}_i l + \tilde{\mu}_i$, with virtual wage and nonlabor income \tilde{w}_i and $\tilde{\mu}_i$, so that the associated indirect utility function is:

$$v_i(\tilde{w}_i, \tilde{\mu}_i) = \max[u_i(y_i, l_i) | y_i \leq \tilde{w}_i(\tau - l_i) + \tilde{\mu}_i].$$

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, i.e. it can be chosen according to explicit fairness criteria. Then, the second principle of *restricted dominance* can apply. It confines the dominance principle – i.e. a better bundle in all dimensions always reflects a better situation – to the reference set. In other words, it allows interpersonal comparison on a subset defined according to some ethical priors, which we now make explicit.¹²

Figure 1: Welfare Metrics: Graphical Representation



3.2 Welfare Metrics

Definitions. In our setting, we consider two cases based on the evaluation of individual situations according to hypothetical, linear budget constraints, as indicated above. Ethically, they

¹²Decancq et al. (2015) show that with restricted dominance and nonpaternalism, it is possible to define a reference set in which we can project the situations of individuals a and b and compare them according to the dominance principle.

both give priority to the *compensation principle*: inequalities arising from endowed circumstances (like innate ability), but not due to responsibility factors (like preferences), should be removed.

The first one is a *Wage metric*: the reference parameter r will be nonlabor income $\tilde{\mu}$ and it will be set equally to 0 for all, so that 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 second graph of Figure 1, it is simply the slope of the tangent through the origin at the actual indifference curve. 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).

The second is a *Rent metric*: r will be the wage rate \tilde{w} , set to 0 for all, so the metric will consist of the nonlabor income level μ_i^* that allow individuals to reach their current utility level. As shown in the third graph of Figure 1, it is the vertical intercept of the actual indifference curve in the case of well-behaved preferences. 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).

Normative Interpretations. When equalizing external resources (either wage or unearned income), individuals will work and earn at their convenience only, i.e. individual responsibility characteristics, like preferences, are unchanged. However, our two measures embody different ethical priors, in the realm of the compensation principle, on how to weigh people with different preferences. In fact, the Wage and Rent metrics have been chosen because they represent two polar cases. With the Wage metric, a situation where all individuals have the same wage leads to laissez-faire as the best possible allocation, i.e. remaining inequalities are solely due to differences in preference and are legitimate. Thus, the Wage metric is implicitly interpreted as holding people maximally responsible for their work aversion. On the second graph of Figure 1, individual b (work averse) is deemed better off than a (hard worker), so b -to- a redistribution is justified.

In contrast, on the third graph, interpersonal comparison is conducted in a counterfactual where inequalities from productivity differences are ignored (all productivities are set to zero). A situation with the same unearned income for all would lead to equal welfare for all: 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. As a result, in our example, this metric supports redistribution towards the work averse b .¹³

¹³Our welfare metrics more generally belong to the domain of *responsibility-sensitive egalitarianism*, an ap-

Calculating Welfare Metrics. To bring theory to practice, let us first formally define the Wage and Rent metrics as:

$$\begin{aligned}\nu_i^W(u, \mu^r = 0) &= \min_{\tilde{w}_i} [\tilde{w}_i | v_i(\tilde{w}_i, \mu^r = 0) \geq u_i] \\ \nu_i^R(u, w^r = 0) &= \min_{\tilde{\mu}_i} [\tilde{\mu}_i | v_i(w^r = 0, \tilde{\mu}_i) \geq u_i]\end{aligned}$$

respectively (we drop time to simplify notations). The empirical application first consists in evaluating the utility levels \bar{u}_i^E and \bar{u}_i^D obtained at actual choices (y_i^{obs}, l_i^{obs}) , which are ‘decision utility’-maximizing. Then we retrieve individual indifference curves for each observation i in the data. They correspond to implicit functions of y and l defined as $\bar{u}_i^m = \hat{u}_i^m(y, l)$, for $m = E, D$, using estimated utility functions \hat{u}_i^m . Finally, metrics are 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 $j = 1, \dots, J$ as used for the labor supply estimation). The Wage metrics 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 (technical details on these procedures are provided in Bargain et al. 2013).

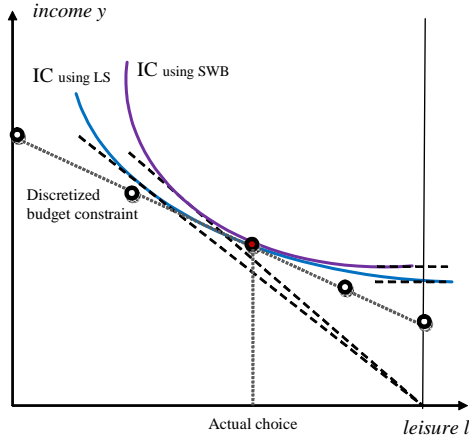
Characterizing Situations when Revealed and Subjective Preferences Diverge. In Figure 2, we characterize four possible situations where revealed and subjective preferences differ. In graph 1, the labor supply choice occurs on a relatively flat portion of the budget constraint, and welfare evaluations diverge only a little when using revealed rather than subjective preferences (both for the Rent and Wage metrics). In graph 2, differences are larger, and so is the difference in welfare evaluation (both for the Rent and Wage metrics). Yet, the contrast between revealed and subjective preferences still does not lead to different optimal choices. People indeed concentrate at certain work hours because of kinks in the nonlinear budget curve as represented in graph 2 (for instance due to the working tax credit in the UK) or to institutional constraints.

In graphs 3 and 4, revealed and subjective preferences are sufficiently different, and the budget constraint on a flat enough portion, so that actual work hours diverge from what would maximize SWB. That is, actual hours are "too high" ("too low") in graph 4 (graph 3), i.e. working

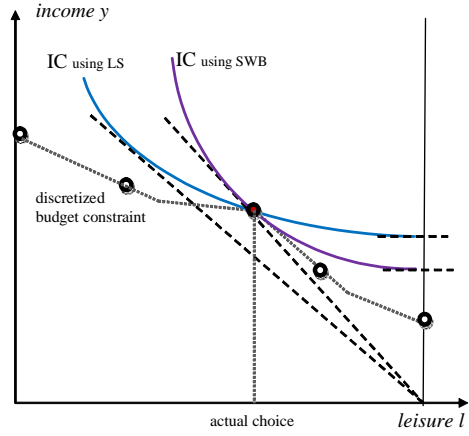
proach that helps to rank individuals when their outcomes differ because of differences both in endowed circumstances and in individual preferences. This ethical approach keeps individuals responsible for the latter but not for the former (Fleurbaey 2008), which can be done along two principles. The first one is our *compensation principle*, that seeks to eliminate inequalities due to nonresponsibility factors (like innate ability), as in the tradition of second-best optimal policy design. The second, the *principle of liberal reward*, deems legitimate the inequalities due to individual preferences. These principles are logically independent from each other but can nonetheless be in conflict. The Wage (Rent) criterion corresponds to the situation where they are most compatible (most in conflict).

overtime (being inactive) is rationalized as low (high) work aversion by the revealed preference approach while SWB-maximization would imply lower work hours (positive work hours, i.e. participation to the labor market). These two situations have different implications for welfare evaluations. In graph 3, overworking individuals are deemed worse (better) off according to revealed (subjective) preferences under the Wage metric. The reverse is true under the Rent metric. In graph 4, quite symmetrically, inactive people are deemed better (worse) off according to revealed (subjective) preferences under the Wage metric, but there is no difference under the Rent metric given the use of the no-work situation as the implicit reference point in this case. This characterization will prove useful hereafter.

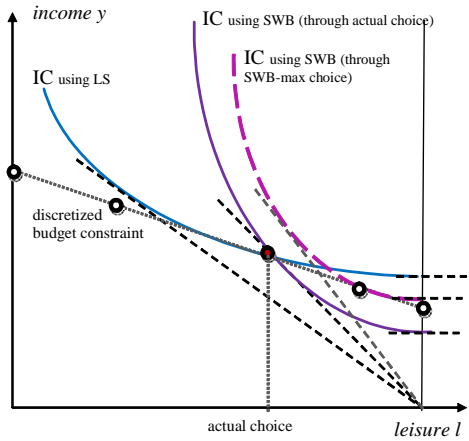
Figure 2: Possible Situations Explaining Reranking



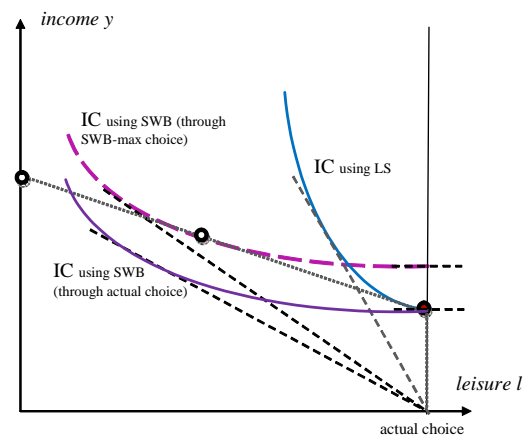
Graph 1: Different preferences, same optimal choice



Graph 2: different preferences, same optimal choice (kink)



Graph 3: different preferences, different optimal choices ("overwork")



Graph 4: different preferences, different optimal choice ('underwork')

4 Results

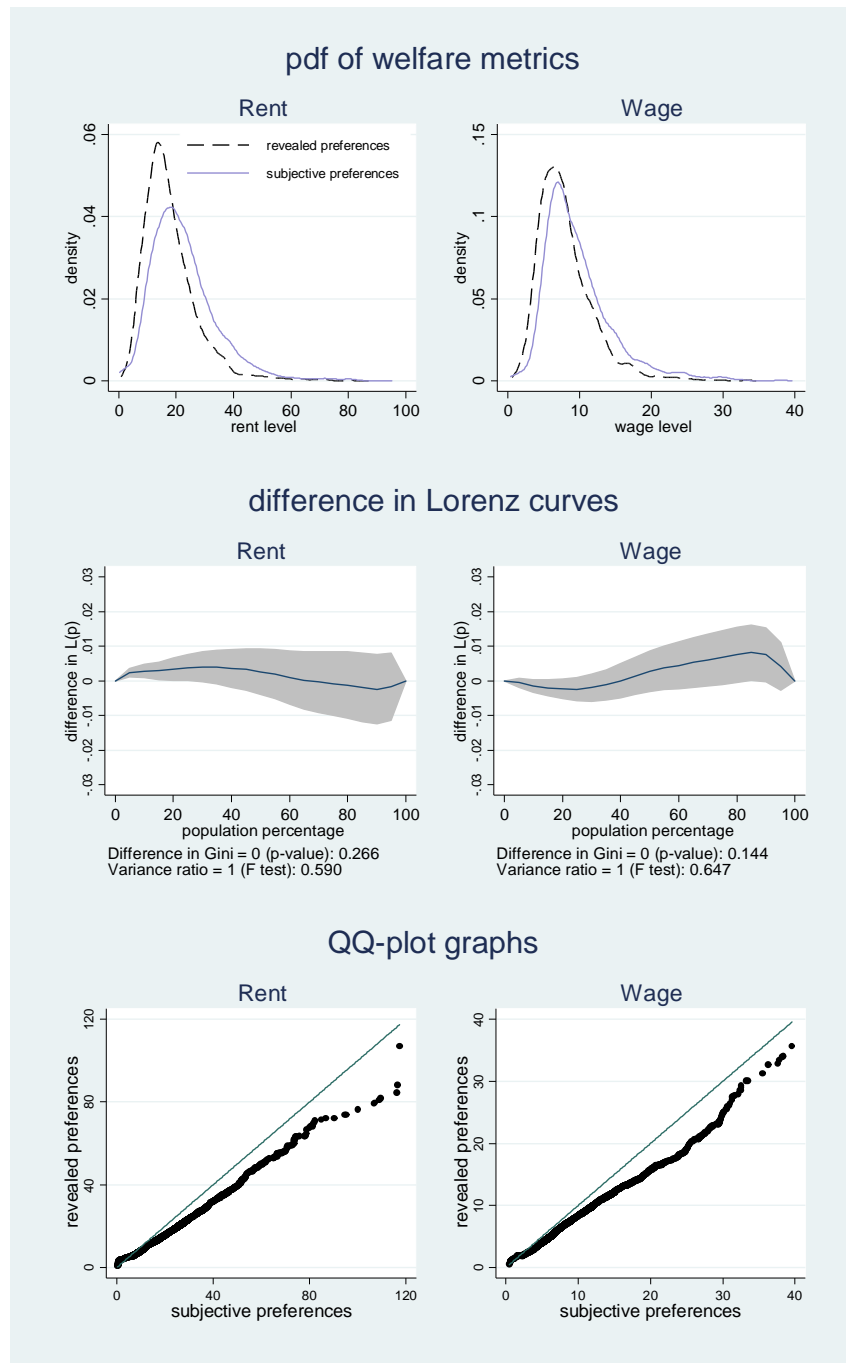
We present the results in three steps, bearing in mind our initial objective of assessing whether the way we capture preference heterogeneity makes a difference for distributional welfare analysis. First, we compare the distributions obtained with revealed versus subjective preferences as it would be done in a standard inequality analysis (yet, with a broader welfare concept than income). Then, our main contribution consists in the direct confrontation of the two orderings, investigating the possible explanatory channels for reranking. Finally, we focus on a comparison when characterizing the worst off.

4.1 Welfare Inequality Analysis

We start with very basic comparisons of the welfare distributions obtained when evaluating individual welfare at actual choices using revealed versus subjective preferences. The upper panel of Figure 3 shows the densities of the two distributions for the Rent metric (left) or the Wage metric (right). In both cases, kernel distributions look rather similar and log-normal. The choice-based welfare levels (dashed lines) are slightly more concentrated while SWB-based welfare levels (solid purple) appear a bit more right-skewed. An alternative representation could be the c.d.f. or, with the domain of the random variable in $[0, 1]$, the Lorenz curve. Thus, in the intermediary graph of Figure 3, we show the difference in Lorenz curves obtained with revealed versus subjective preferences. As indicated in the test statistics below the graphs, we cannot reject at conventional levels that the difference in Gini coefficients is null, nor that the variance ratios are equal to one. Hence, the overall welfare dispersion with the two types of preference measure is similar. Yet, inequality possibly comes from different segments of the distribution. Indeed, on the graphs, the Lorenz gap crosses the zero line – i.e. the Lorenz curves cross – at 40% (60%) of the distribution with the Rent (Wage) metric. In both cases, however, the distance between the Lorenz curves is almost never significant along the cumulated distribution.

To directly compare the distributions to one another, we suggest quantile-quantile plots in the last panel of Figure 3. They compare two distributions by plotting their quantiles 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 $\pounds 30$ per week and wages below $\pounds 10$ per hour, which correspond to the vast majority of observations (i.e. the lower 80 – 90% of the distribution). Some differences appear above these levels – higher quantiles are more disperse according to subjective preferences, which is consistent with the density graphs. While these results are encouraging, a more precise characterization requires to check the degree of reranking that takes place when using one type of preference rather than the other, which we now investigate in detail.

Figure 3: Comparing Welfare Distributions based on Revealed vs. Subjective Preferences

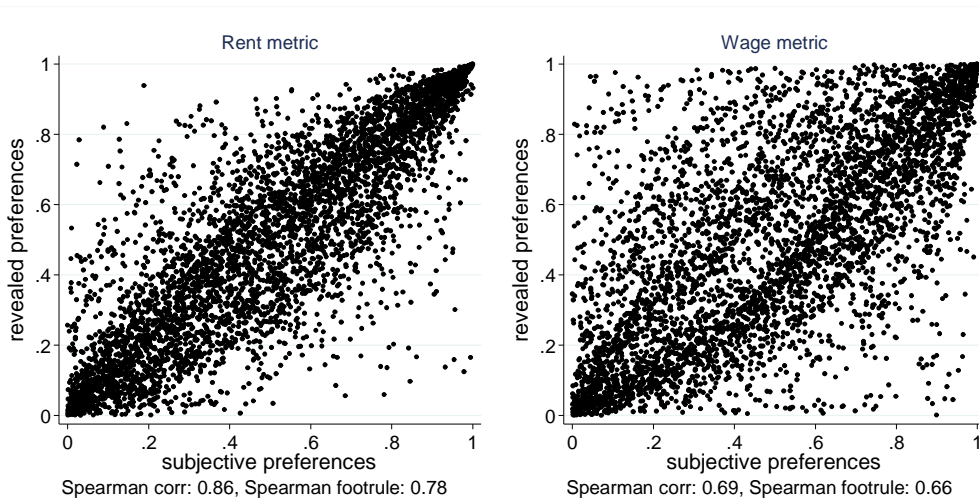


4.2 Analyzing Reranking: General Characterization

We move to our core results whereby we directly compare the ranks of each observation according to revealed versus subjective preferences, for either the Rent or the Wage metric. Below each graph that follows, we indicate two summary indices of the overall correlation between the two distributions, which decrease with the distance between ranks for each observation. The first, the Spearman rank correlation, is a function of the sum of squared distances between ranks. The second, the Spearman footrule, is the sum of absolute distances between ranks.

Overall Reranking. For the overall sample, results are presented in Figure 4, for the Rent and Wage metrics separately. It turns out that the Spearman correlation and footrule measures are relatively high, especially with the Rent metric. A relatively basic test of whether the two distributions are similar, in each case, 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 can reject that the two distributions are similar in the case of the Wage metric (p-value of .01) but not for the Rent metric (p-value of .38).

Figure 4: Rank Correlation of Welfare Metrics: Whole Sample



Note: for either Rent or Wage metrics, 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. Preferences are modelled using box-cox utility functions with preference heterogeneity (male, age, education, presence of young children, London, non-white, migrant, conscientious, neurotic).

Sensitivity Checks. In the Appendix A.2, we provide some sensitivity checks for these basic results. First, we see that results are relatively similar when using alternative measures of SWB to proxy V_{it}^E . Figure A.1 shows that this is particularly true when adding heterogeneity in the relative weights on income and leisure satisfactions in our concentrated measure. Admittedly, the dispersion increases when using the general life satisfaction measure in place of our concentrated measure, which is expected given that the former covers all the dimensions that shape subjective satisfaction of one’s life and hence adds considerable noise to our welfare characterizations. Nonetheless, the Spearman correlation remains substantial in this case as well. Finally, recall that we pool several years of a panel – a choice mainly driven by the attempt to get precise estimates. Yet, we wonder if having the same persons several times in the observations of Figure 4 has some influence on the results. Figure A.2 reports welfare comparisons when using the time average welfare level for each person in our sample. The picture is very similar to the baseline.

Interpretations. How can we interpret reranking? There are multiple factors that could explain the dissonance between revealed and subjective preferences. First of all, the former may not be authentically "revealed" preferences if observed income-leisure choices are constrained (for instance by restrictions on the labor market). This could be a conjecture to explain less dispersion with the Rent metric, if much of the reranking corresponds to situations of constrained inactivity. Let us illustrate this point using the graph 4 of Figure 2 above. There, inactivity is rationalized by high aversion to work in "revealed" preferences while subjective preferences may actually point to higher tastes for work. These differences possibly lead to a lot of reranking under the Wage metrics. Yet they are simply ignored with the Rent approach for people at zero hour – the shape of indifference curves at positive work hours does not matter.

As discussed in length in Akay et al. (2015), divergences between revealed and subjective preferences may also come from different types of optimization ‘errors’ (for instance, Kahneman et al., 2006, study how ‘focusing illusions’ give too much importance to income compared to other aspects of a good life).¹⁴ They may also stem from the pursuit of other goals than short-term well-being.¹⁵ An important aspect is that these differences could concern the overall

¹⁴Several 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).

¹⁵There is also a grey zone between constraints and alternative life objectives, containing aspects like moral obligations. Working ‘too much’ (resp. ‘too little’) can be due to the obligation of bringing money to the family (of staying at home to care for children). Alternative objectives may also relate to intertemporal optimization (ex: people work harder today to secure comfortable old days). Clearly, it is not possible to disentangle these different factors 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).

balance between income and leisure, reflecting systematic discrepancies between the approaches. In Akay et al. (2015), we have shown that it is not the case. As discussed in Appendix A.3, indifference curves from revealed versus subjective preferences do overlap on average (graph A of Appendix Figure A.3). However, when looking at different groups (graphs B-H), some differences tend to appear, the distributional consequences of which are now explored.

4.3 Analyzing Reranking: Sub-Groups

Reranking within Groups. In order to investigate whether reranking is mainly driven by certain aspects of preference heterogeneity, we now compare the ranks based on revealed versus subjective preferences when looking at broad groups defined by gender, age, education or personality traits. Results in Appendix Figure B.1 show within-group reranking using group-specific ranks (i.e. ranks redefined among observations of the same group, for instance among males). It turns out that the extent of reranking, i.e. the thickness of the plotted area, varies across groups. It is particularly small in some cases, for instance among men (for them, the Spearman correlation is as high as .91 with the Rent metric). It becomes much larger among women, the young or the low-educated when evaluated with the Wage metric (the Spearman correlation is still large but goes down to .60 – .70).

To check the contribution of each group to the *overall* reranking, we plot population ranks (as used in Figure 4), rather than group-specific ranks, in Appendix Figure B.2. Again, there is consistently more dispersion with the Wage metrics for groups like women, the above-40 and low-educated (compared to, for instance, men or the under-40). Yet, some groups show specific asymmetrical patterns, which are intuitively explained by the empirical indifference curves of Appendix Figure A.3. In particular for gender (graph B in that Figure), "revealed" indifference curves of men are relatively flat, which is the way the labor supply model rationalizes relatively high working hours for men. In contrast, their "subjective" indifference curves are steeper (the subjective experience of working long hours should imply relatively larger compensation). Consistently, for the Rent metric, men are deemed relatively better off according to their labor supply, i.e. ranked higher in the overall distribution, than according to their subjective preferences – hence they tend to be concentrated above the 45° line in Figure B.2. By construction, the opposite is true with the Wage metric. A relatively symmetrical picture is observed for women in both Figure A.3 and B.2, which is consistent with the fact that their work hours are lower on average – and possibly "too" low according to subjective preferences. While gender offers the clearest example of these patterns, similar interpretations apply to other dimensions – for instance, the highly educated or above-40 show trends that are similar to men's.

Deviations. As discussed above, it is likely that reranking occurs more often when revealed and subjective preferences diverge in a way that leads to ‘suboptimal’ choices with respect to SWB-maximization. To investigate this point, let us call ‘deviation’ the absolute distance between actual and SWB-maximizing work hours. As discussed above, deviations may stem from bounded rationality, from the pursuit of alternative objectives or simply from labor market constraints. Figure C.1 in the Appendix presents the correlation graphs by ‘deviation’ groups, using group-specific ranks. Clearly, reranking is larger in the group characterized by deviations above 10 hours per week.¹⁶ It is consistent with the previous analysis according to which ‘sub-optimal’ choices reflect large discrepancies between revealed and subjective preferences (graphs 3 and 4 in Figure 2). In Figure C.2, we focus on the group showing deviations larger than 10 hours per week. We represent first those who "underwork" while being inactive (as in graph 4 of Figure 2), those who "underwork" in part-time activities, and finally those who "overwork" (as in graph 3 of Figure 2). Inactivity seems to be the situation generating the highest dispersion under the Wage metric (the Spearman correlation goes as low as .42) while it is relatively harmless for the Rent metric comparison (using one type of preference or the other is neutral in this case, as discussed above).¹⁷

Subgroup Contributions. We finally combine these different dimensions to better characterize the channels of reranking. We consider subgroups defined by a combination of gender, education and age levels. We denote M (F) for male (female), H (L) for high (low) education, and Y (O) for below (above) 40 years of age. Among young low-educated women (F-L-Y), we also differentiate between those with children (F-L-Y-C) and those without (F-L-Y-N). We then report the individual contribution of each group to total reranking on Figure 5 for the Rent metric and Figure 6 for the Wage metric. Groups enter the picture by increasing order of their contribution to total deviation, from M-H-Y (representing 5% of the sample but contributing to 3% of total deviation) to F-L-Y (representing 23% of the sample but contributing to 30% of total deviation). In each Figure, we add one group at a time – its contribution is plotted in red – while neutralizing reranking for the groups not yet entered (their observations are on the 45° line). We report the overall Spearman correlation and footrule at each step. For instance, the inclusion of the 7th group F-L-O in Figure 5 decreases Spearman correlation from .93 to .88, i.e. a marginal contribution of .05. With both metrics, we observe a very high correlation in most subgroups but a faster drop in the Spearman measures when introducing F-L-O and

¹⁶Notice than 10 hours is the step used for hour discretization in our labor supply model, so that deviations below 10 hours may just be due to the approximation induced by this (standard) modelling choice.

¹⁷We do not expect more dispersion with the Wage metric than with the Rent metric when considering other cases. Consistently, we see that "underwork" at part-time gives similar dispersion with both metrics while "overwork" actually shows more dispersion with the Rent metric.

F-L-Y-C (the contribution of M-L-O also seems important with the Wage metric). There is no reason for these groups to be affected by limited rationality or alternative life goals more than the rest of the population. Instead, this decomposition conveys that most of the reranking is attached to groups that potentially suffer from specific labor market constraints related to low skills and possible other dimensions (gender).

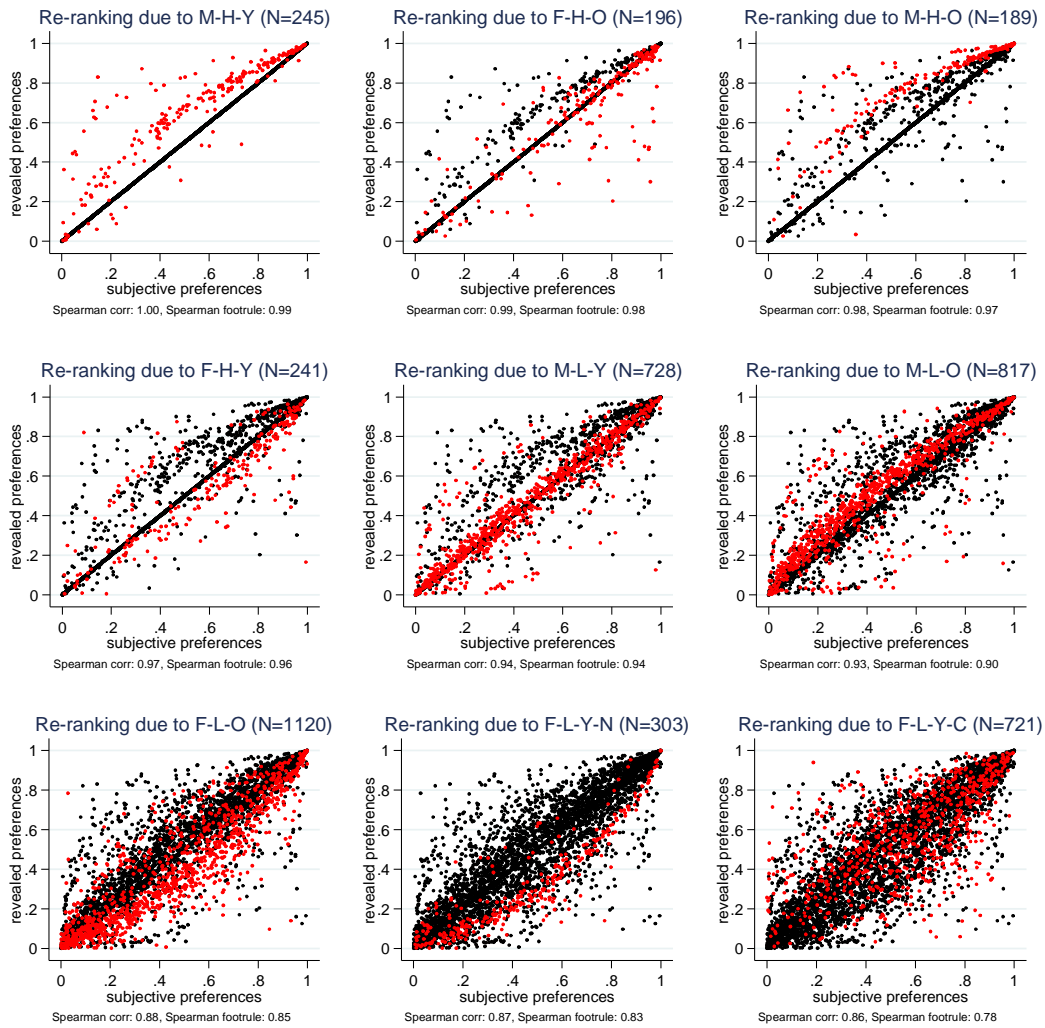
These results are summarized in Table 1. The first two columns present the distribution by groups in the sample and the proportion of large deviations (>10 hours/week) in each group (it is 40% in the overall population). The third column shows the distribution of contributors to the total number of large deviations.¹⁸ Three groups previously mentioned, M-L-O, F-L-O and F-L-Y-C, account for 74% of large deviations. The next four columns, summing to 1, show the breakdown by type of deviation in the population characterized by large deviation. The role of underwork and overwork is relatively balanced overall (and among low-skilled older women) while the contribution of single mothers is clearly associated with underemployment, and particularly with complete inactivity. The last four columns of Table 1 first report, for each metric, the total reranking (in *italic*), calculated as one minus the Spearman footrule. Below, we report the contribution of each subgroup to total reranking. It is simply extracted from Figures 5 and 6 as the marginal change in Spearman footrule associated with adding this group (for instance, a marginal contribution of .05 for group F-L-O under the Rent metric, as exemplified before), expressed in percentage of total reranking. Note that this approach yields a perfect decomposition of total reranking, which is *not sensitive to the order of decomposition* (and is actually identical to one where we extract the contribution of each group at a time by neutralizing *all* other groups).¹⁹

Results confirm that reranking is closely associated to large deviations (the correlation between the contribution to the former and the contribution to the latter is .98 for the Rent and .92 for the Wage, as indicated in the lower part of the table). Among the main contributors to reranking, it turns out that low-educated men and women, especially older, are characterized by overwork. While overwork may be related to ‘focusing illusion’ or high aspirations (cf. Akay et al., 2015), a more likely set of explanations pertains to the role of labor markets. In particular, low-skilled workers may avail of less flexible jobs (no possibility of part-time work, for instance), tighter budget constraints and/or genuinely higher work aversion (due to the bad quality of jobs they can find). Older low-skilled women are also substantial contributors through underwork,

¹⁸As in the previous figures, the groups are ordered by contribution to total deviation. A similar order is obtained when using the proportion of large deviation or – unreported – the average level of absolute deviation, with the exception of M-L-Y (the latter has the lowest rate of high deviation but is a relatively large group).

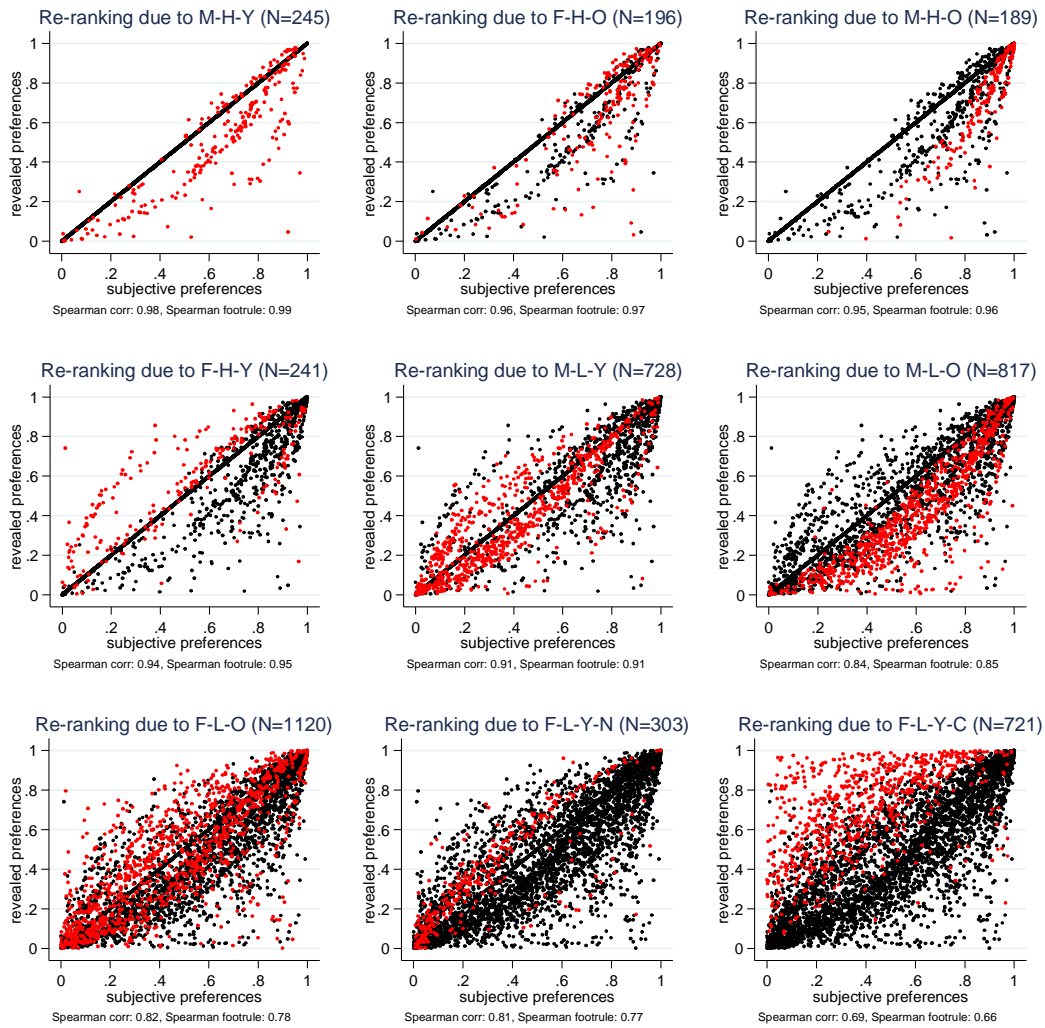
¹⁹This is not the case with the Spearman correlation, which is sensitive to the order of decomposition. Shapley values could be calculated but it would require to compute contributions for the 362,880 permutations of the 9 subgroups.

Figure 5: Re-ranking by Cumulative Demographic Subgroups (Rent Metric)



Note: the graphs compare Rent metric ranks obtained using revealed versus subjective preferences. Observations are grouped by subgroups defined by M (F) for male (female), H (L) for high (low) education, and Y (O) for below (above) 40 years of age. In the first graph, the contribution of the first group is assessed by imposing no-reranking among other groups (welfare rank under subjective preference is set equal to the one under revealed preference, i.e. these groups are on the 45° line). In the following graphs, the actual contribution of each other group is consecutively added.

Figure 6: Re-ranking by Cumulative Demographic Subgroups (Wage Metric)



Note: the graphs compare Wage metric ranks obtained using revealed versus subjective preferences. Observations are grouped by subgroups defined by M (F) for male (female), H (L) for high (low) education, and Y (O) for below (above) 40 years of age. In the first graph, the contribution of the first group is assessed by imposing no-re-ranking among other groups (welfare rank under subjective preference is set equal to the one under revealed preference, i.e. these groups are on the 45° line). In the following graphs, the actual contribution of each other group is consecutively added.

which may be related to gender discrimination (see Petrongolo, 2004) and rationing out of the labor market due to depreciated skills after long-term unemployment. The largest contributor is the group of low-skilled single mother, who cumulate the difficulties described above with low work incentives due to childcare costs.²⁰ As expected, their contribution to reranking is largest with the Wage metric since the Rent metric does not generate reranking in the case of unemployment, as previously discussed. The last rows of the table indicate that contribution to reranking is particularly correlated with contribution to underwork (more than with overwork) and, as expected, that this is even more so with the Wage metric. This result is mainly due to single mothers. More generally, without this group, the Spearman footrule would be as high as .83 with the Rent and .77 with the Wage, which are the levels obtained in the population with zero deviation (first graph of Figure C.1).

Table 1: Re-ranking by Socio-Demographic Sub-Groups

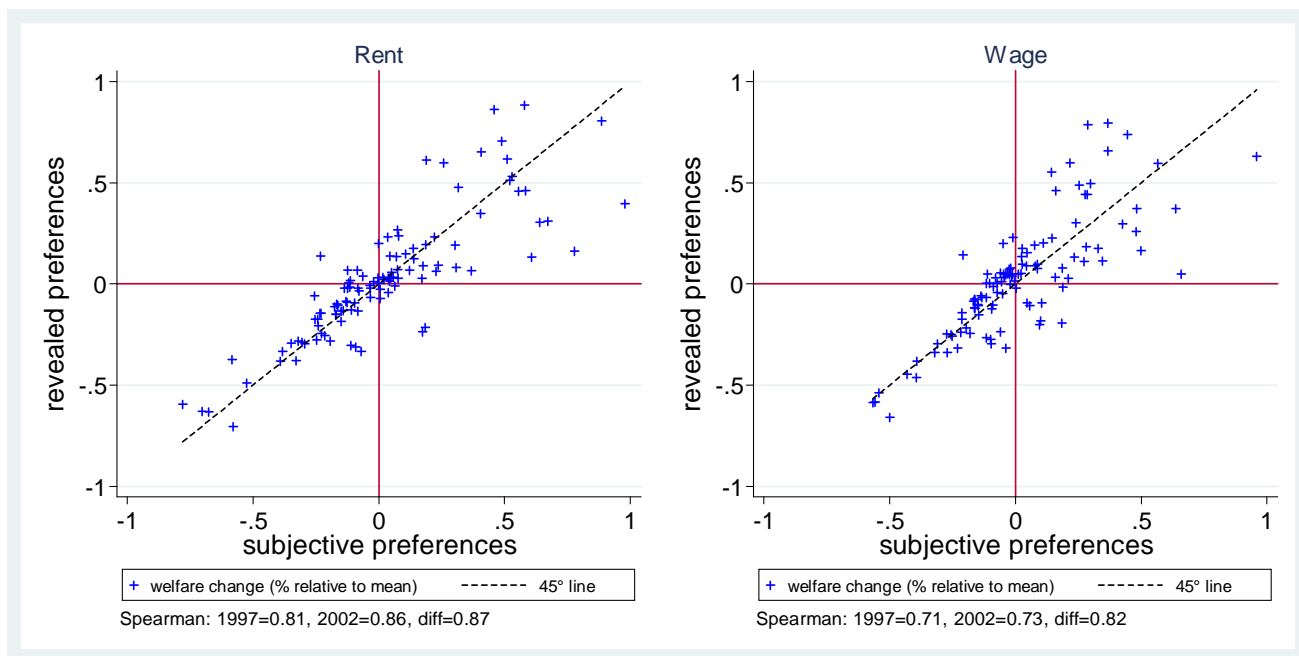
Subgroup	% of Total Sample	% Large Deviations (>10)	Contribution to Large Deviations (>10)				Contribution to overall reranking (Spearman footrule)		
			Overall	Under-work (zero)	Under-work (part-time)	Under-work (others)	Over-work	Rent	Wage
Total		40%	100%	breakdown:				<i>0.22</i>	<i>0.34</i>
				31%	16%	2%	50%		
M-H-Y	5%	25%	3%	0%	0%	0%	3%	5%	3%
F-H-O	4%	33%	4%	0.5%	1%	0%	2%	5%	6%
M-H-O	4%	34%	4%	0%	0%	0%	3%	5%	3%
F-H-Y	5%	43%	6%	0.4%	1%	2%	2%	5%	3%
M-L-Y	16%	18%	7%	2%	1%	0%	5%	9%	12%
M-L-O	18%	43%	20%	3%	0%	0%	16%	18%	18%
F-L-O	25%	45%	28%	8%	6%	0%	14%	23%	21%
F-L-Y-N	7%	21%	4%	1%	1%	0%	2%	9%	3%
F-L-Y-C	16%	64%	26%	16%	7%	0%	2%	23%	32%
Correlation between contribution to reranking and contribution to...									
...large deviations								0.98	0.92
...large deviations (underwork, 0 hours)								0.85	0.94
...large deviations (whole underwork)								0.83	0.91
...large deviations (overwork)								0.63	0.42
Observations are grouped by combined socio-demographic characteristics denoted by: M (F) for male (female), L (H) for low (high) education, Y (O) for younger than 40 (older), N (C) for no children (children). Groups are ranked according to the % of large deviations (large absolute distance between actual and SWB-maximizing hours). The last 2 columns indicate the level of welfare reranking calculated as 1 minus the Spearman footrule (in italic), and the contribution of each subgroup (obtained for each group by imposing no-reranking among all subsequent groups and extracting the marginal effect of this group compared to the previous row).									

²⁰The 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).

4.4 Additional Results

Time Changes. Given the time dimension of our data, we can isolate a balanced panel of observation corresponding to a certain period of important tax-benefit changes, for instance the first term of the New Labour (1997-2001). We calculate time change in welfare metrics relative to the mean change, i.e. we identify the relative winners and losers. Results in Figure 7 show that the Spearman correlation tends to increase over time, possibly due to a decrease in labor market frictions and an improvement in the situation of single mothers (see Bargain, 2012). Interestingly, the correlation of the welfare change (indicated by "diff" on the graph) is higher than the correlation in levels at any year, especially for the Wage metric. That revealed and subjective preferences tend to give more similar conclusions when looking at changes over time is perhaps encouraging regarding their (joint) use for the analysis of tax-benefit reforms.

Figure 7: Time Change in Welfare Metrics (1997-2001)



Worst-off: General Considerations. We finally check the profile of the most deprived in our sample, using the different welfare measures. 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. 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.

Worst-off: Overlap across Metric and Preference Measures. We define the most deprived as the bottom quintile of the welfare metric distributions.²¹ We first study the degree of overlap across the groups identified as worst off with the different measures. Results are reported in Table 2. Figures in bold focus on the impact of using revealed or subjective preferences for each metric. Consistently with the previous analysis, we find a larger degree of overlap with the Rent metric (71%) than with the Wage metric (58%). We also report the overlap with the income-poor. It is largest for the Rent metric and revealed preferences. Indeed, the income-poor are typically unemployed while the value of leisure is the lowest with the Rent metric (minimal responsibility to those "revealed" as highly work averse). Other figures indicate that ethical principles underlying the Rent and the Wage metrics play a more limited role in the definition of the worst off when using revealed preferences (*italic*) rather than subjective preferences (underlined). In the former case, 82.8% of the worst-off are common to both welfare criteria, which is very similar to what is found by Carpantier and Sapata (2016).

Table 2: Overlap of the Most Deprived across Measures

	Income	Rent, Rev. Pref.	Wage, Rev. Pref.	Rent, Subj. Pref.
Rent, Rev. Pref.	0.612			
Wage, Rev. Pref.	0.554	<i>0.828</i>		
Rent, Subj. Pref.	0.565	0.713	0.732	
Wage, Subj. Pref.	0.570	0.603	0.578	<u>0.552</u>

Cells report the % of overlap of the worst off (bottom welfare quintile) between two measures.

²¹Note that this is different from standard poverty analyses that rely on poverty lines, the definition of which would add another degree of arbitrariness to our characterization. Focusing on the bottom 20% of the well-being distributions, as we do, allows comparing a group of the same size across the different welfare measures.

Profile of the Worst Off. Table 3 suggests a portrait of the worst off according to each welfare measure. We report the mean characteristics of the bottom quintile for income alone (as a benchmark) as well as Rent and Wage metrics (using revealed or subjective preferences). The first two rows concern mean income and working time. The four welfare metrics giving a non-zero value to leisure consistently identify the worst off as people with larger income but lower "leisure". Income-leisure satisfaction (SWB) is also slightly larger with the metrics than when using income alone. This indicates that the income-poor are more often inactive and, despite enjoying more non-market time, report lower satisfaction from their income-leisure bundle. Considering the other characteristics in Table 3, we notice that the worst off according to income alone are more likely to be women, low educated and single parents, which is consistent with low labor market outcomes in this population – low wages but also lower work hours that are not valued in the income measure.

Table 3: Characteristics of the Most Deprived by Metric

	Income	Rent			Wage		
		Rev. pref.	Subj. pref.	Δ	Rev. pref.	Subj. pref.	Δ
Disp. Income	117.2 (36.32)	146.7 (71.47)	155.3 (69.34)	*	150.8 (70.66)	155.9 (60.91)	
Hours	13.7 (16.96)	29.5 (18.54)	29.5 (17.93)		31.7 (17.97)	26.1 (19.11)	***
SWB	4.57 (0.84)	4.69 (0.81)	4.68 (0.84)		4.75 (0.82)	4.65 (0.83)	*
Male	0.28 (0.45)	0.37 (0.48)	0.48 (0.50)	***	0.54 (0.50)	0.31 (0.46)	***
Over 40	0.52 (0.50)	0.59 (0.49)	0.58 (0.49)		0.57 (0.50)	0.35 (0.48)	***
High Education	0.05 (0.22)	0.07 (0.26)	0.11 (0.31)	*	0.10 (0.30)	0.09 (0.28)	
Child 0-2	0.38 (0.49)	0.11 (0.31)	0.12 (0.33)		0.05 (0.22)	0.26 (0.44)	***
London	0.06 (0.23)	0.07 (0.26)	0.14 (0.35)	***	0.05 (0.22)	0.01 (0.10)	***
Non-white	0.01 (0.09)	0.01 (0.09)	0.01 (0.10)		0.01 (0.07)	0.00 (0.06)	
Migrant	0.02 (0.14)	0.02 (0.14)	0.02 (0.15)		0.02 (0.13)	0.01 (0.09)	
Conscientious	0.28 (0.45)	0.29 (0.45)	0.27 (0.44)		0.34 (0.47)	0.45 (0.50)	***
Neurotic	0.53 (0.50)	0.52 (0.50)	0.48 (0.50)		0.48 (0.50)	0.58 (0.49)	***

Notes: income is in pounds per week, hours are weekly, SWB is a weighted average of financial and leisure time satisfactions on 1-7 scales. 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.

Strikingly, the gender and education composition of the worst off varies dramatically across

welfare metrics and preference measures. First, consider the *Rent* metric. The worst-off group is more often composed of women, low-educated workers or non-Londoners when using revealed preferences, while the reverse trend is observed with subjective preferences. The previous reasoning applies: These groups are more often characterized by under-work, which is rationalized as high work aversion by revealed preferences and, hence, higher compensation (and a more frequent classification among the poor) under the *Rent* metric. Things are somewhat reversed with the *Wage* metric: with revealed preferences, work-loving "preferences" of males, highly educated or Londoners (see Figure A.3) lead to higher compensation for these groups and their more frequent allocation to the worst-off group than when using subjective preferences (consistently, the welfare poor show higher work hours when characterized by revealed preferences). Inversely, unemployed women with children are deemed responsible for their "work aversion" according to revealed preferences while they disproportionally make up the poor group with subjective preferences. For both metrics, we indicate statistically significant differences in the characteristics of the worst-off between revealed and subjective preferences. As expected, the picture of the worst-off is more contrasted across preference types when using the *Wage* metric.

5 Summary and Concluding Discussion

The literature tends to show that for standard decisions in life (like work choices), there is an overall congruence between decision and experienced utility (Benjamin et al., 2012, Akay et al., 2015, Fleurbaey and Schwandt, 2015). Yet, there may be distributional implications of using revealed preferences rather than subjective well-being. Focusing on income-leisure preferences, our study suggests a first investigation of whether the way we assess preference heterogeneity matters for welfare analysis. Ordinal preferences are elicited using either actual labor supply choices (revealed preferences) or subjective well-being levels consistent with these choices (subjective preferences). Estimations are used to derive money metrics based on a 'fair allocation' approach in which the compensation principle prevails. We retain two polar cases whereby workers are held minimally or maximally responsible for work aversion (*Rent* and *Wage* metrics). We find that rank correlation is high – and very high in groups whose actual decisions are well in line with SWB-maximization. Most of the welfare reranking seems to be associated with the underemployment of single mothers and the 'suboptimal' work hours for low-educated workers, possibly facing constraints on the labor market. The identification of the worst off also depends on ethical views regarding responsibility for work preferences and the extent to which labor market constraints affect the low-skills and those with high costs of work due to childcare.

These conclusions call for further work. We believe that the most urgent question relates to

whether underemployment is involuntary – and the extent to which it must be treated as a non-responsibility factor. This old question should be addressed anew, both econometrically and normatively. 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 apparently idle workers (possibly long-term unemployed who are discouraged from searching or single mothers facing high costs of work). There are many possible ways to "clean" our estimates from these aspects, yet they all rely on specific assumptions.²² The main difficulty pertains to the fragile identification of what stems from preferences, what is due to actual rationing (e.g. discrimination, productivity below minimum wage, frictional unemployment, discouragement or access to low-quality jobs only, etc.) and what is due to false perceptions about the choice set. The same difficulty applies to the modelling of work costs (like childcare costs) because it is difficult to identify them non-parametrically from preferences (see van Soest et al., 2002). Also, we are not aware of attempts to identify work costs in SWB estimations.

Beyond econometric challenges, it is certainly necessary to extend the normative treatment of these questions. A very conservative view may entail for instance that there is no such thing as involuntary unemployment as one could always create his own (possibly informal) job or take up any job (even if of lower quality or not matching one's skill). Inversely, unemployment may be thought of as covering a very broad set of unchosen situations and be deemed a non-voluntary outcome. The various factors listed above, that possibly explain underemployment, should lead to specific ethical characterizations. Dynamic aspects should also be considered. 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 stigma which will take time to be rubbed out."

Our study has contributed to show that measurement is intricately related to these normative questions. In particular, the choice of ethical priors may not be independent from the type of preferences we use and the information contained by these preferences. For instance, if underwork is constrained, leading to wrong inference about what actual hours reveal, then the *Rent metric*, that holds people minimally responsibly for their "revealed" preferences, seems more appropriate. 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 (zero in the *Rent metric*).²³

²²Many studies have explicitly accounted for labor market rationing, for instance by modelling the probability of involuntary unemployment (e.g., Haan and Uhlenhorff, 2013), the demand-side of the labor market (Peichl and Siegloch, 2012) or the distribution of job opportunities (see a modern account in Befy et al., 2016).

²³Fleurbaey and Maniquet (2006) suggest that involuntary unemployment (resp. constrained part-time jobs)

If the *Wage metric* is chosen, we have shown that the type of preference measure matters a lot for the unemployed and the underemployed – and hence for a Rawlsian objectives of helping the worse off. In particular, if combined to the revealed preference approach, the Wage metric gives maximal responsibility for underemployment (and maximal value to non-market time) to those with little or no work. More generally, the Laissez-Faire principle underlying the Wage metric becomes unacceptable if individual preferences are not fully respectable, and notably if they reflect external factors leading to underemployment (but also overemployment: moral obligation regarding financial support to the extended family, workaholism due to social pressure, etc.). Thus, it might be wiser to rely on subjective preferences – at least on the regularities they embed – to construct welfare measures under the Wage metric.²⁴ SWB measures have at least the merit to reveal that staying at home causes emotional distress (Clark and Oswald, 1994), establishing flatter and more acceptable indifference curves than choice-based preferences.

In fact, it might be possible to suggest more fuzzy assessments of the correspondence between "authentic" preferences (Fleurbaey and Schokkaert, 2013) and observed choices. Individuals may not maximize happiness alone but include it in a grand utility function along with other arguments (Glaeser et al., 2016). In this case, recent advances that incorporate insights from behavioral economics into welfare measurement may be transposed to the normative approach, in the line of Fleurbaey and Schokkaert (2013). Further work should explore whether the interval between revealed and subjective preferences can be helpful to define incomplete preference relation (in the vein of Bernheim and Rangel, 2009) and whether the latter can still be used for distributional judgments on the basis of partial orderings. Another path for further work could combine normative work with experimental data. The degree of labor market constraint is hard to measure and to define in nonexperimental data. 'Deviation' as we measured them could be seen as a way to evaluate frictions on the labor market, at least by subgroups as defined in our analysis (or along other set of heterogeneity factors). Experiments could help to control for different factors explaining why someone does not pick up a job or is underemployed – with

may be viewed as nullifying (resp. reducing) the agents' earning ability, which is what the Rent metric is doing. The authors add: "the worry that greater work aversion may be explained by disadvantages can partly be addressed [...] by selecting a low reference wage rate in the construction of the utility index. However, addressing these issues completely and satisfactorily requires adding the relevant features into the model, and, for applications, finding estimates of the distribution of characteristics in the relevant population." Few optimal tax applications address this issue (see the review in Fleurbaey and Maniquet, 2014). An exception is Luttens and Ooghe (2007), who assume that the productivity of the workers deemed involuntary unemployed is zero or below the minimum wage and that their work preferences can be taken in the neighbourhood of those of the working or deviate partly from them.

²⁴Yet, it raises new questions. Why should subjective indifference curves be used at actual work hours, if choice is constrained? And what is the interpretation of the (virtual) welfare evaluation at the SWB-maximizing hours (something that we have not done here but that is suggested by dashed indifference curves in Figure 2)?

different degrees of responsibility being attached to these factors.

Finally, further work should also attempt to use welfare metrics for aggregation in a social welfare function. After all, equivalent incomes and wages are interpersonally comparable. A well-known issue is that equivalent measures are not necessarily concave in income and, hence, may induce antiegalitarian policy implications (see Blackorby and Donaldson, 1988). This is why we have focused on distributional analyses and welfare ranks. It is the relevant perspective for the implementation of (progressive) redistribution, which requires a ranking of all individuals in a society. Note however that 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 occur only when indifference curves change shape when utility increases, in a way that makes the violation of the principle not so shocking.²⁵ Social welfare aggregation could be pursued more systematically as suggested in Bosmans et al. (2017).

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.

²⁵We thank Marc Fleurbaey for making this point.

- [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
- [10] Blackorby, C. & Donaldson, D. (1988): "Money Metric Utility: A Harmless Normalization", *Journal of Economic Theory* 46: 120–129
- [11] 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
- [12] Blundell, R., Duncan, A. & Meghir, C. (1998): "Estimating labor supply responses using tax reforms", *Econometrica* 66(4), 827–861
- [13] Boadway, R. (2012): "Review of 'A Theory of Fairness and Social Welfare' by Fleurbaey and Maniquet". *Journal of Economic Literature*, 50(2), 517–521
- [14] Bosmans, K., K. Decancq and E. Ooghe (2017): "Who's afraid of aggregating money metrics?", mimeo
- [15] Boyce, C.J. (2010): "Understanding Fixed Effects in Human Wellbeing", *Journal of Economic Psychology*, 31: 1-16.
- [16] Carpentier, J.F. and C. Sapata (2016): "Empirical Welfare Analysis: When Preferences Matter", *Social Choice and Welfare Economics*, 46(3), 521-542
- [17] Clark, A. and A. Oswald (1994): "Unhappiness and Unemployment", *Economic Journal*, 104, 424, 648-659
- [18] 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
- [19] Clark, A., C. Senik, & K. Yamada (2015): "When Experienced and Decision Utility Concur: The Case of Income Comparisons", IZA DP No. 9189
- [20] Decancq, K., M. Fleurbaey and E. Schokkaert (2014): "Happiness, equivalent incomes, and respect for individual preferences", forthcoming in *Economica*
- [21] 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
- [22] 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

- [23] Decoster, A. & P. Haan (2015): "Empirical welfare analysis with preference heterogeneity", *International Tax and Public Finance*, 22 (2), 224-251
- [24] 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
- [25] 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
- [26] 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
- [27] Fleurbaey, M. (2008), *Fairness, Responsibility and Welfare*, Oxford University Press.
- [28] Fleurbaey, M. (2009): "Beyond GDP: The Quest for a Measure of Social Welfare", *Journal of Economic Literature* 47: 1029–1075.
- [29] Fleurbaey M. & F. Maniquet (2006): "Fair Income Tax", *Review of Economic Studies*, 73(1), 55-83
- [30] 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
- [31] Fleurbaey M. & F. Maniquet (2011), *A Theory of Fairness and Social Welfare*, Cambridge University Press
- [32] Fleurbaey M. & F. Maniquet (2014): "Optimal Taxation Theory and Principles of Fairness", Core Discussion Paper.
- [33] Fleurbaey, M., & Blanchet, D. (2013), *Beyond GDP, Measuring Welfare and Assessing Sustainability*, Oxford University Press.
- [34] Fleurbaey, M. & H. Schwandt (2015): "Do People Seek to Maximize Their Subjective Well-Being?", IZA DP No. 9450
- [35] Fleurbaey, M. and E. Schokkaert (2013): "Behavioral welfare economics and redistribution", *American Economic Journal: Microeconomics*, 5(3):180-205, 201
- [36] Frijters, P. (2000). "Do individuals try to maximise satisfaction with life as a whole" *Journal of Economic Psychology*, 21, 281-304.
- [37] 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

- [38] Glaeser, E.L., J.D. Gottlieb & O. Ziv (2016): "Unhappy cities", *Journal of Labor Economics*, 34(S2), S129 - S182
- [39] Haan, P. & A. Uhlendorff (2013): "Intertemporal Labor Supply and Involuntary Unemployment", *Empirical Economics*, 44, 2, 661-683
- [40] 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
- [41] Kahneman, D., P. Wakker, & R. Sarin (1997): "Back to Bentham? Explorations of Experienced Utility", *Quarterly Journal of Economics*, 112(2): 375-405
- [42] Loewenstein, G., O'Donoghue, T., & M. Rabin (2003): "Projection bias in predicting future utility", *Quarterly Journal of Economics*, 118; 1209–1248
- [43] Luttens, R.I. & E. Ooghe (2007): "Is It Fair to 'Make Work Pay'?", *Economica*, 74, 296, 599-626
- [44] Odermatt, R. and A. Stutzer (2015). "(Mis-)Predicted Subjective Well-Being Following Life Events", IZA DP No. 9252
- [45] Ooghe E. & A. Peichl (2011): "Fair and efficient taxation under partial control: theory and evidence", *Economic Journal*
- [46] Pazner, E., & D. Schmeidler (1978): "Egalitarian Equivalent Allocations: A New Concept of Economic Equity". *Quarterly Journal of Economics*, 92, 671-686.
- [47] Peichl, A. and S. Siegloch (2012): "Accounting for labor demand effects in structural labor supply models", *Labour Economics* 19 (1), 129-138
- [48] Pencavel, J. (1977): "Constant-Utility Index Numbers of Real Wages", *The American Economic Review* 67(2): 91–100.
- [49] Perez-Truglia, R. (2015): "A Samuelsonian Validation Test for Happiness Data", *Journal of Economic Psychology*, 49, 74–83
- [50] Petrongolo, B. (2004): "Gender Segregation in Employment Contracts", *Journal of the European Economic Association*, 2 (2/3, P&P), 331-345
- [51] Ravallion, M., & Lokshin, M., (2001): "Identifying Welfare Effect from Subjective Questions", *Economica*, 68: 335-357
- [52] 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

- [53] Senik, C. (2005). "Income distribution and well-being: what can we learn from subjective data?", *Journal of Economic Surveys* 19: 43–63
- [54] Stiglitz, J., Sen, A. & Fitoussi, J.-P. (2009): "Report by the Commission on the Measurement of Economic Performance and Social Progress", Technical Report.
- [55] Thomson W. (1994): "Notions of equal, or equivalent, opportunities", *Social Choice and Welfare* 11: 137–156
- [56] Thomson W. (2011): "Fair allocation rules" in K.J. Arrow, A.K. Sen, and K. Suzumura (eds.), *Handbook of Social Choice and Welfare*, vol. 2, Amsterdam: North-Holland
- [57] Trannoy, A. (2016): "Equality of Opportunity: A progress report", ECINEQ working paper
- [58] 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
- [59] van Soest, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach", *The Journal of Human Resources*, 30 (1), 63-88
- [60] 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.
- [61] Viitanen, T. (2005): "Cost of Childcare and Female Employment in the UK", *Labour*, 19 (Special Issue): 149-179
- [62] Wichert, L. & W. Pohlmeier (2010): "Female Labor Force Participation and the Big Five", ZEW discussion paper 10-003

A Appendix A

A.1 Model Specification

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

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

Used in recent welfare analyses (Decoster and Haan, 2014, and Bargain et al., 2013), box-cox utility allows imposing or easily checking that preferences are well-behaved, which facilitates the derivation of ordinal preferences (i.e., indifference curves). The paper attempts to retrieve preference heterogeneity across individuals, so that parameters on leisure and income terms vary linearly with taste shifters x_{it} and possibly a normally distributed random term ϕ_i , dealt with using simulated maximum likelihood. That is, we specify $\beta_l^m(x_{it}, \phi_i) = \beta_{l0}^m + \beta_{l1}^{m'} x_{it} + \beta_{l2}^m \phi_i$ and $\beta_y^m(x_{it}) = \beta_{y0}^m + \beta_{y1}^{m'} x_{it}$ for $m = D, E$. 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.

Budget Constraints. In both approaches, disposable income is computed according to the budget constraint $y_{it} = \psi_t(w_{it}h_{it}, \mu_{it}, \zeta_{it})$ and depends on gross earnings $w_{it}h_{it}$, unearned income μ_{it} and a set of individual characteristics ζ_{it} . All these inputs are transformed by function ψ_t into net income y_{it} , i.e. this function aggregates labor and non-labor income, imputes taxes and imputes benefits. It is approximated by numerical simulations using the tax-benefit rules of each period $t = 1, \dots, T$. In the same way, we also predict $(y_{ijt}, \tau - h_{ijt})$ pairs for the $j = 1, \dots, J$ 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 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. In Akay et al. (2015), we discuss in more detail the econometric identification of ordinal preferences and summarize here the argument. For labor supply models, the difficulty pertains to the role of unobservables that affect both wages and preferences (ex: being a hard-working type). As in the bulk of the literature, identification is obtained by exploiting exogenous variation in *net* wages stemming from spatial variation in tax-benefit rules (as in 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 sufficient variation in tax-benefit rules to

guarantee identification (see Akay et al., 2015). For the SWB model, the potential bias pertains to essential heterogeneity, for instance if actual heterogeneity in work "preferences" (x_{it}, ϕ_i) is correlated with other unobserved determinants of well-being ϵ_{it} . We cannot completely rule out this correlation but put forwards two arguments. First, we account for some of the individual heterogeneity (conscientiousness and neuroticism) in both work preference parameters x_{it} and separately additive well-being terms z_{it} . Second, as in the case of labor supply decisions, the potential role of omitted variables can be addressed using exogenous variation stemming from policy reforms. Precisely, the same person may not make the same labor supply choice at two points in time because she faces different socio-fiscal regimes.

A.2 Sensitivity Checks

Alternative Subjective Well-Being Measures. Our baseline results are obtained using a concentrated measure of income-leisure satisfaction $V_{it}^E = \hat{\gamma}^y S_{it}^y + \hat{\gamma}^l S_{it}^l$ (graph A in Figure A.1). First, we have tried more flexible specifications than the linear form, namely the addition of interaction terms between S_{it}^y and S_{it}^l (the coefficient of which proved insignificant) and/or quadratic terms. These variants hardly change the results (detailed outcomes unreported but available from the authors). Second, we introduce some heterogeneity, i.e. we write $V_{it}^E = \hat{\gamma}_{it}^y S_{it}^y + \hat{\gamma}_{it}^l S_{it}^l$ with $\gamma_{it}^y = \gamma^{y0} + x'_{it} \gamma^{y1}$ and $\gamma_{it}^l = \gamma^{l0} + x'_{it} \gamma^{l1}$ (the set of demographics x_i is the same as preference shifters in the model). Again, results are hardly affected, as seen in the graph B of Figure A.1. Third, we use overall life satisfaction, i.e. $V_{it}^E = S_{it}$. Given that the latter carries much more noise than the concentrated measure, results tend to show a little more dispersion. This is especially the case for the Rent metric and as summarized by the Spearman rank correlation – see the graph C in Figure A.1.

Time-Collapsing Panel Observations. The main results display welfare levels for each observations in our panel. Estimations were conducted on pooled years in order to make our estimates as precise as possible and, also, because identification of the empirical model relies on time-variation in socio-fiscal rules, as explained in Appendix A.1 When collapsing observations in time-average welfare levels, in Figure A.2, we obtain relatively similar plots as in the baseline.

A.3 Indifference Curves by Subgroups

We derive indifference curves in the income-leisure space for every individuals 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 $\bar{y}(40)$ (the sample mean net income at this leisure level). Results in Figure A.3 show black solid curves for the

Figure A.1: Reranking for different Subjective Well-Being Measures

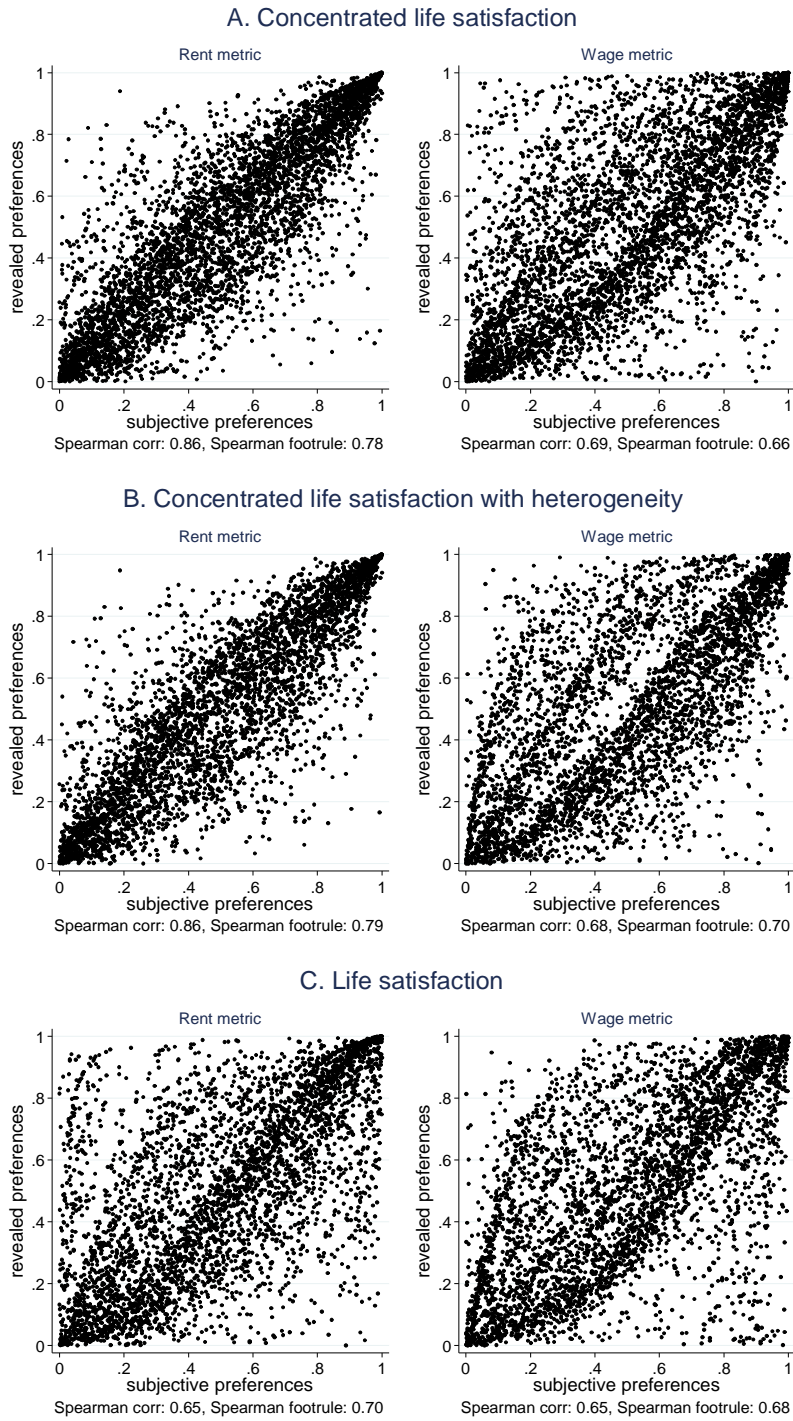
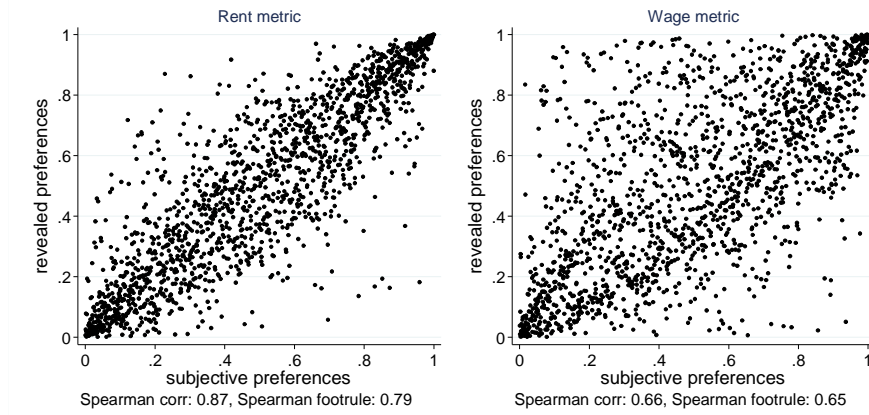
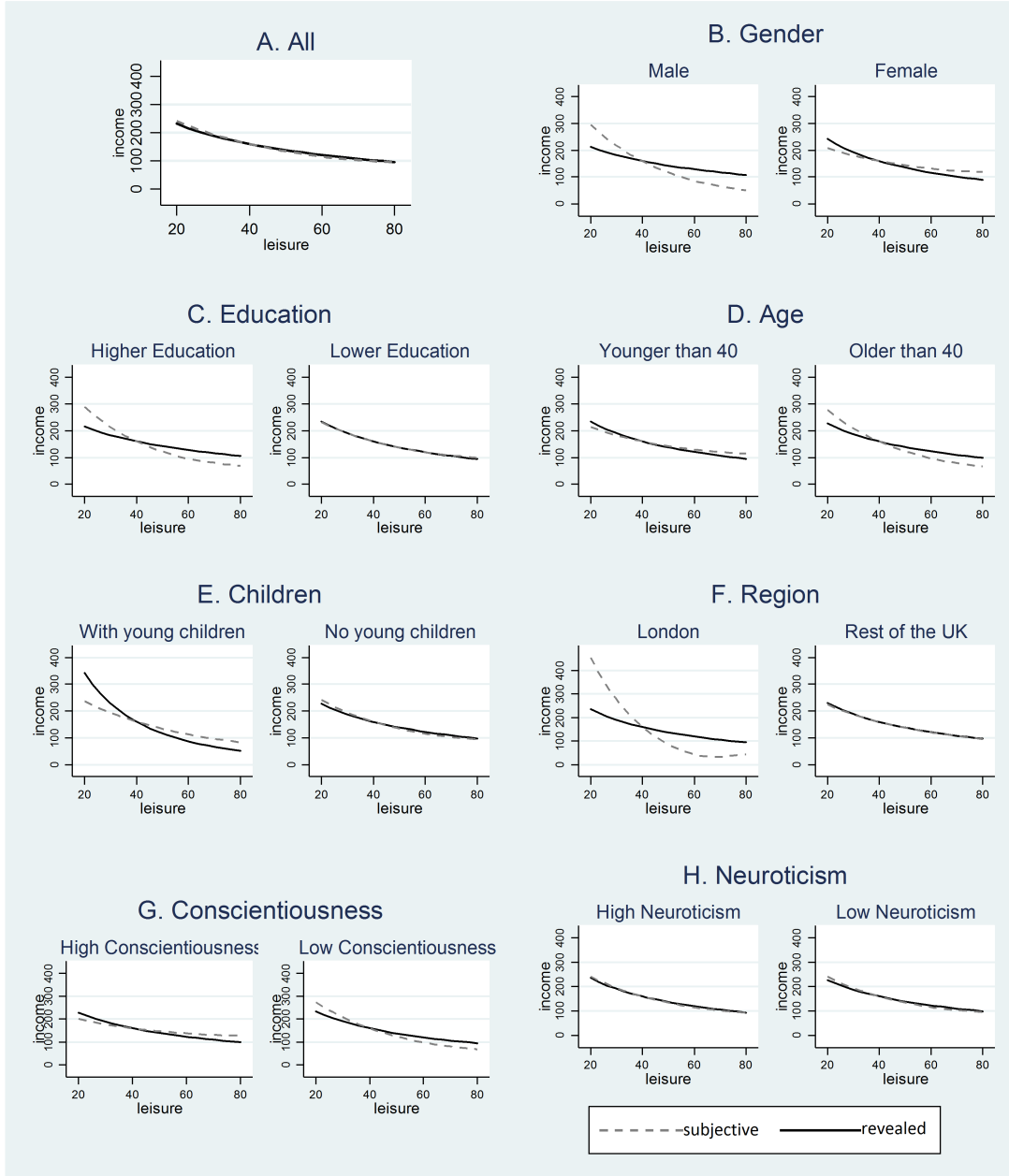


Figure A.2: Reranking when using Time Average



indifference curves derived from the labor supply model while the gray dashed curves represent those consistent with the subjective experience. Weekly leisure points range from 20 to 80 hours, corresponding to weekly work hours from 60 (overtime) to 0 (inactivity). Indifference curves with revealed versus subjective preferences seem to overlap quite well overall, i.e. the ordinal preferences rationalizing actual choices do not differ from those implicit in SWB information *on average*, while some differences appear when looking at specific groups of the population (see Akay et al., 2015).

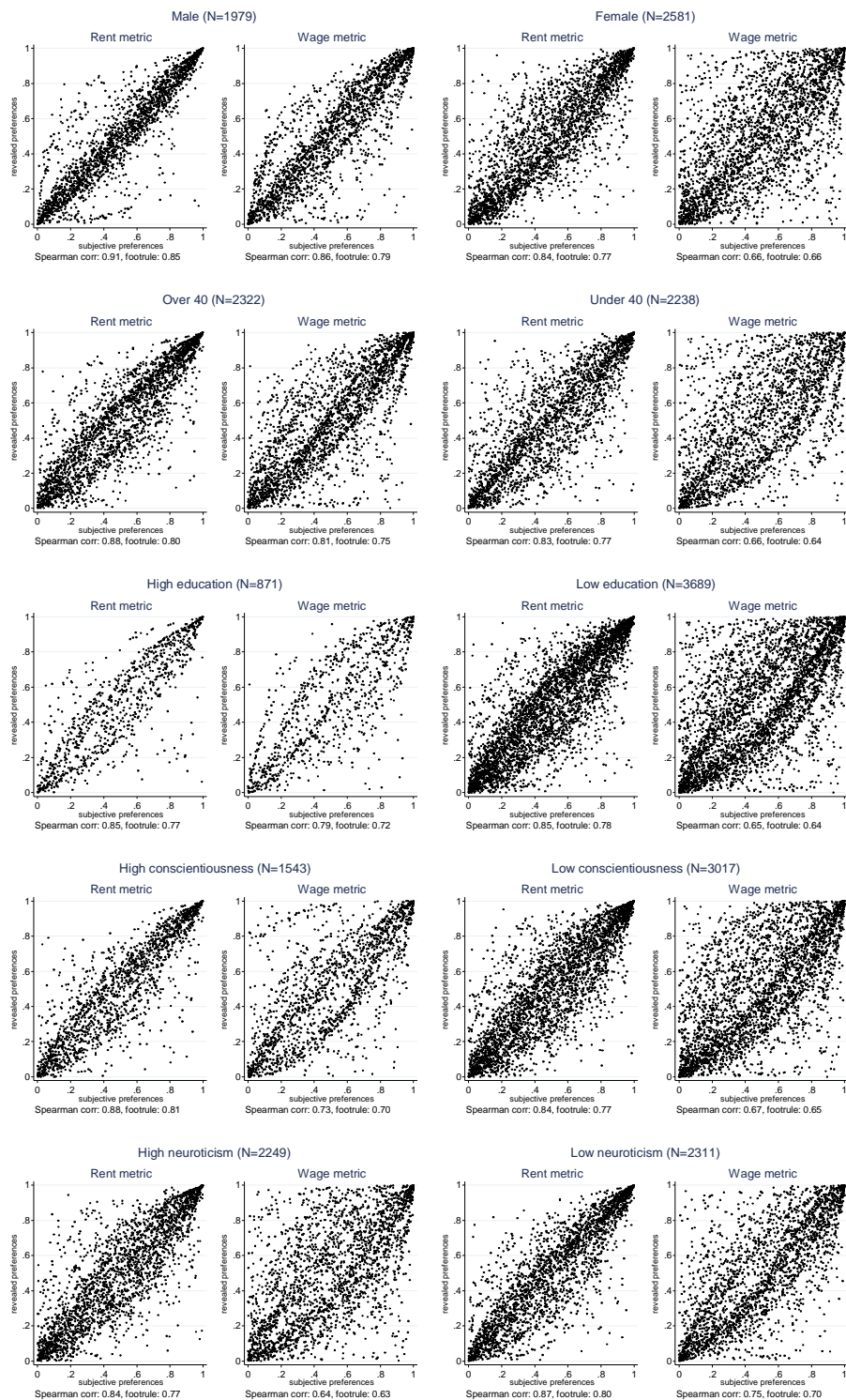
Figure A.3: Indifference Curves with Revealed vs. Subjective Preferences



Note: Indifference Curves (ICs) are obtained using estimated parameters of income-leisure utility functions, estimated using either income-leisure satisfaction ('subjective') or labor supply ('revealed'). We use box-cox utility functions with preference heterogeneity (male, age, education, presence of young kid, London, non-white, migrant, conscientious, neurotic). These variables as well as additional controls (age squared, family size, health status, home ownership, all personality traits, region and year dummies) enter the SWB equation as additively separable controls (hence, not affecting the calculation of ICs). Graphs are obtained by averaging all individual ICs (using either SWB or Utility) drawn through a common point, defined as $(\bar{y}(40), 40)$.

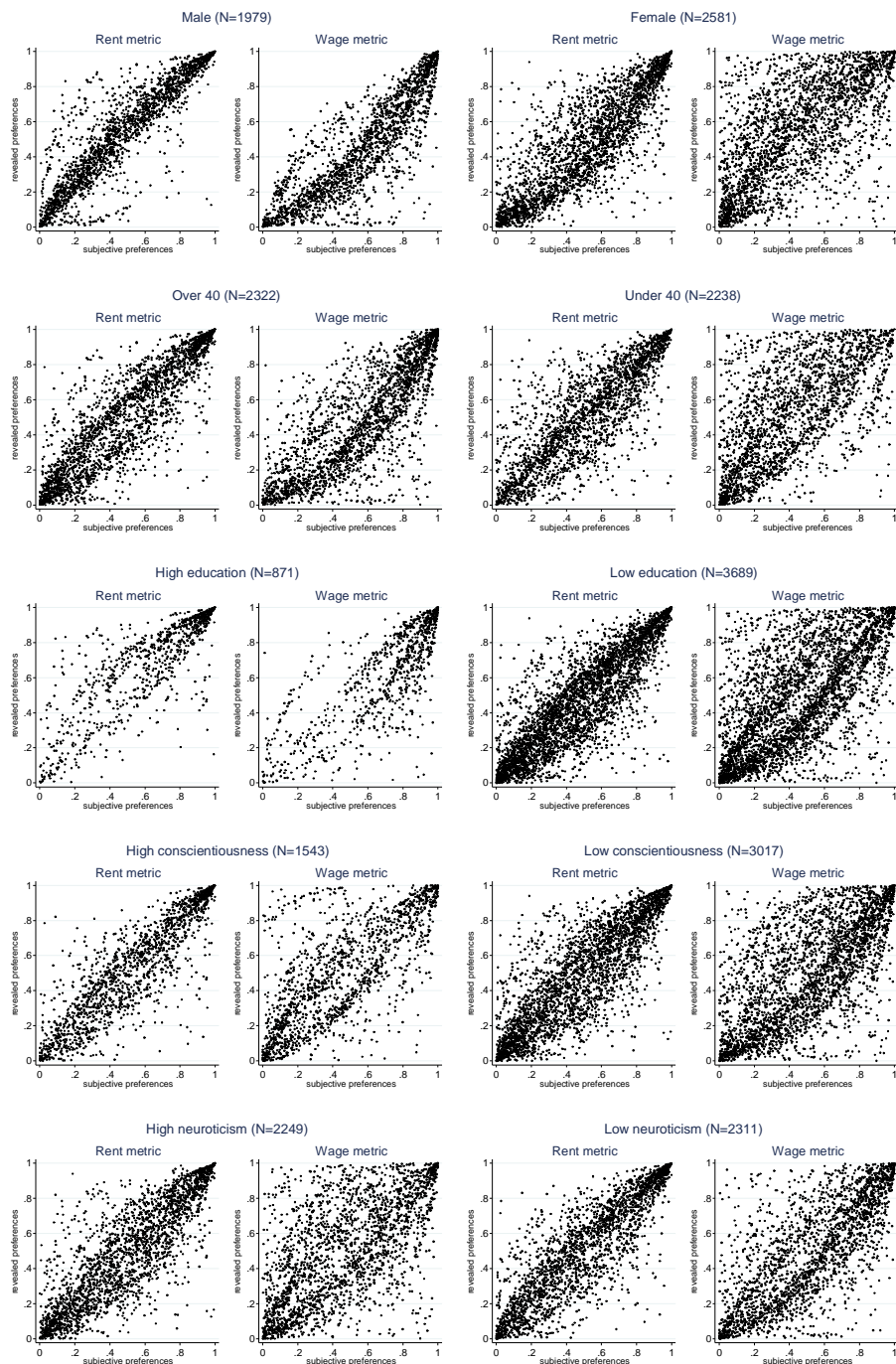
B Appendix B: Reranking by Socio-Demographic Groups

Figure B.1: Rank Correlation of Welfare Metrics by Groups (using Group-specific Ranks)



Note: for either Rent or Wage metrics, 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.

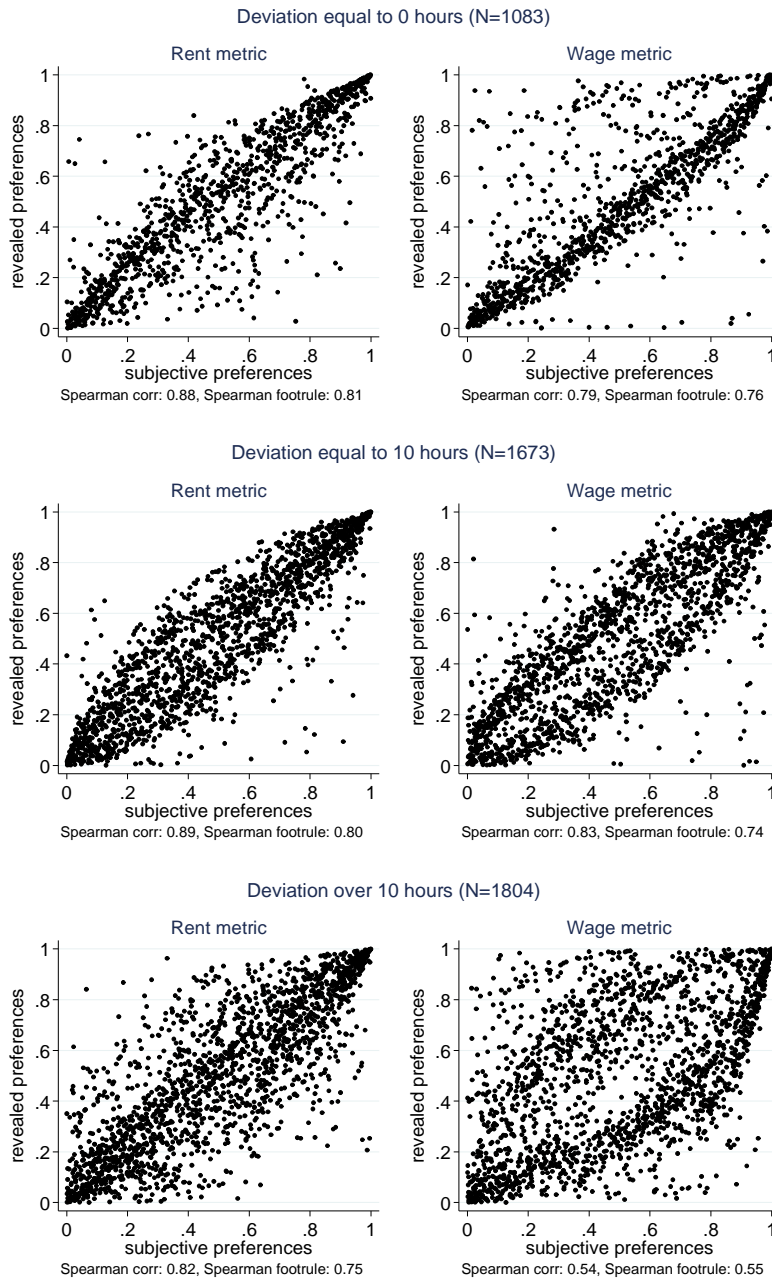
Figure B.2: Rank Correlation of Welfare Metrics by Groups (using Overall Ranks)



Note: for either Rent or Wage metrics, 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.

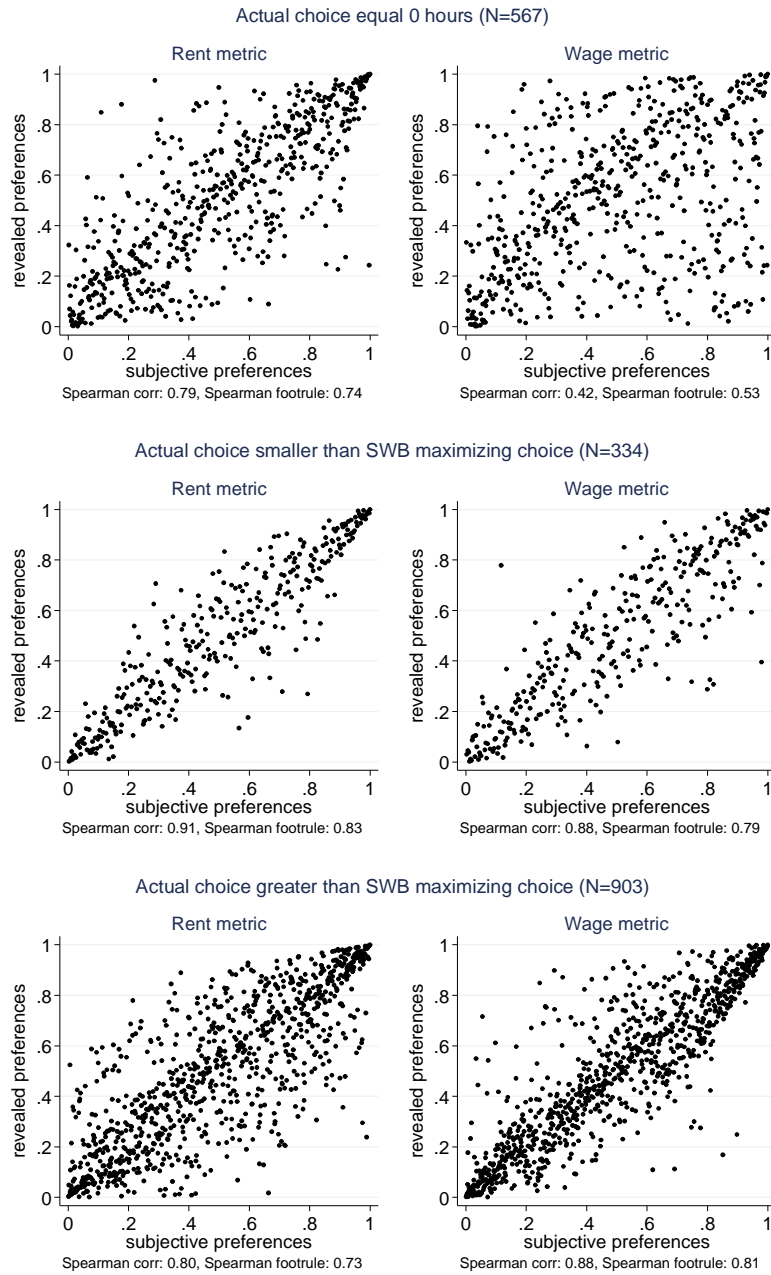
C Reranking by Deviation Level & Type

Figure C.1: Re-ranking by Deviation Level (Group-specific Ranks)



Note: the graphs compare Rent or Wage metric ranks obtained for all observation using revealed versus subjective preferences. The comparison is carried out for different levels of absolute deviation between actual and SWB-maximizing work hours, using the group-specific welfare ranks.

Figure C.2: Re-ranking by Type of High Deviation (Group-specific Ranks)



Note: the graphs compare Rent or Wage metric ranks obtained for all observation using revealed versus subjective preferences. We focus on observations in the group with deviations (the absolute distance between actual and SWB-maximizing work hours) that are larger than 10 hours. The comparison is carried out for three types of high deviations, using the group-specific welfare ranks, namely those who underwork while being inactive (first graph), those who underwork while being part-time (second graph), those who overwork (third graph).