sciendo

Journal of Official Statistics, Vol. 35, No. 3, 2019, pp. 531–576, http://dx.doi.org/10.2478/JOS-2019-0023

An Integrated Database to Measure Living Standards

Elena Dalla Chiara[1], *Martina Menon*[1], *and Federico Perali*[1]

This study generates an integrated database to measure living standards in Italy using propensity score matching. We follow the recommendations of the Commission on the Measurement of Economic Performance and Social Progress proposing that income, consumption of market goods and nonmarket activities, and wealth, rather than production, should be evaluated jointly in order to appropriately measure material welfare. Our integrated database is similar in design to the one built for the United States by the Levy Economics Institute to measure the multiple dimensions of well-being. In the United States, as is the case for Italy and most European countries, the state does not maintain a unified database to measure household economic well-being, and data sources about income and employment surveys and other surveys on wealth and the use of time have to be statistically matched. The measure of well-being is therefore the result of a multidimensional evaluation process no longer associated with a single indicator, as is usually the case when measuring gross domestic product. The estimation of individual and social welfare, multidimensional poverty and inequality does require an integrated living standard database where information about consumption, income, time use and subjective well-being are jointly available. With this objective in mind, we combine information available in four different surveys: the European Union Statistics on Income and Living Conditions Survey, the Household Budget Survey, the Time Use Survey, and the Household Conditions and Social Capital Survey. We perform three different statistical matching procedures to link the relevant dimensions of living standards contained in each survey and report both the statistical and economic tests carried out to evaluate the quality of the procedure at a high level of detail.

Key words: Propensity score; statistical matching; well-being; fused data; multidimensional poverty.

1. Introduction

In times of recession it is especially important to understand the multidimensional linkages among income, wealth and consumption and how costs and opportunities are distributed across social classes and territories. In France, the Fitoussi Commission (Stiglitz et al. 2010) set up by the French government to identify new tools to measure economic performance and social progress believes that it is now time to shift the attention from the measurement of economic production to the measurement of the well-being of people. To evaluate material welfare, the Commission proposes that income, consumption of both goods and time, and wealth, rather than production, should be evaluated jointly with the aim of broadening the measures traditionally used for family support, including the evaluation of non-market activities. Income or consumption alone cannot comprehensively

[1] University of Verona, Department of Economics, via Cantarane, 24, 37129, Verona, Italy. Emails: elena.dallachiara@univr.it, martina.menon@univr.it, and federico.perali@univr.it

© Statistics Sweden Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

describe a household's standard of living, although consumption inequality often mirrors income inequality (Attanasio et al. 2015). Consumption, defined by total household expenditure, including possibly an imputed income from housing, differs from income because a household can borrow or save, and it should better reflect long-term standard of living and lifetime resources (Slesnick 1993; Blundell and Preston 1995; Meyer and Sullivan 2011; Brewer and O'Dea 2012).

The measure of well-being is therefore the result of a multidimensional evaluation process no longer associated with a single indicator, as is usually the case when measuring gross domestic product. A person's standard of living depends on multidimensional circumstances such as health status, equal access to education, the ability to develop personal relationships, to enjoy a clean environment and to invest in activities creating social capital. The estimation of individual and social welfare, multidimensional poverty and inequality, which is especially important in light of the evaluation of the impact of Horizon 2020, requires an integrated living standards database where information about consumption, income, time use and subjective well-being are jointly available. Similarly, integrated information is necessary to properly model household production, male and female labor supply, the full cost of children, and fertility decisions accounting for the cost of time invested in child care (Caiumi and Perali 2015).

This integrated architecture is also appropriate for identifying the short and long-run actions guaranteeing the well-being of present and future generations as pursued, for example, by the ISTAT (Italian National Institute of Statistics) which, for the years 2013 and 2014, has produced a policy-relevant report on the Equitable and Sustainable Well-Being of Italians (ISTAT 2013, 2014). Integrated databases about living standards are also useful in epidemiological studies because they can serve as controls for case studies designed to capture all relevant quality-of-life dimensions in order to understand the causes of public health problems such as juvenile crime or public-health related aspects. The ecological framework, which is often used to explain why some groups in society are at a higher risk of exposure to public health problems while others are protected, views public "disease" as the outcome of interactions between many factors at four levels – the individual, the relationship, the community, and the societal (Krug et al. 2002).

An integrated database with a design similar to the one described in the present study has been built for the US by the Levy Economics Institute to measure the multiple dimensions of well-being. In the United States, as in Italy and most European countries, the state does not collect a unified database to measure household economic well-being. Hence data sources about income and employment surveys and other surveys on wealth and time use have to be statistically matched to form the Levy Institute Measure of Economic Well-being (LIMEW) database (Wolff and Zacharias 2003; Kum and Masterson 2010; Sharpe et al. 2011; Wolff et al. 2012).

The Living Standard Measurement Studies (LSMS) conducted by the World Bank in most developing countries, on the other hand, have been designed to capture all the dimensions affecting well-being and quality of life and, in most cases, do not need such a composite matching design. In a developing country context, it is more cost and time efficient to carry out an integrated survey rather than a survey specific to each relevant dimension, as is done in most developed countries where a higher level of statistical precision is required.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Our aim is to create an integrated data set to measure living standards that combines information available from different data sources using Italian data as an empirical example. Our main contribution to the literature is to evaluate both the statistical and economic robustness of the fused data. To this end, we show how to perform robust economic tests based on the fundamental Engel relationship verifying the viability of the fused database for economic analysis. We also illustrate the policy potential of the Italian integrated data set by presenting an excerpt of the results of a research measuring multidimensional poverty and of a causal investigation of juvenile crime in Italy. The matched data set contains information collected in four different surveys: the European Union Statistics on Income and Living Conditions survey (henceforth EUSILC), the Household Budget Survey carried out by the Italian National Statistical Institute (henceforth HBS), the Time Use Survey by the Italian National Statistical Institute (henceforth TUS), the Household Conditions and Social Capital survey of the International Center of Family Studies (henceforth CISF). We implement the statistical matching by using a propensity score approach (Rosenbaum and Rubin 1983; Caliendo and Kopeinig 2008). We also investigate uncertainty by calculating the Fréchet inequality for the contingency table associating income and expenditure classes, which is a special concern of the present analysis.

Our findings relate to Italian data. However, both the implementation method, which is rarely applied to the fusion of four data sets, and the evaluation method, adopting both statistical and economic tests of the quality of the matching, are of general interest. The matching performance is comparable with the matching results adopted by the Levy Institute (Kum and Masterson 2010; Masterson 2010, 2014; Wolff et al. 2012; Rios-Avila 2014, 2015, 2016; Albayrak and Masterson 2017) using mainly US and Canadian data, and with Eurostat (Leulescu and Agafitei 2013; Webber and Tonkin 2013). This evidence suggests that if the same method is applied to other EU countries, the performance is likely to be as statistically and economically robust.

This assertion does not imply that this work is exempted from limitations. In absence of auxiliary information, the present application is developed under the conditional independence assumption. We studied the inferential consequences of this assumption by analyzing the uncertainty associated with the lack of joint information about the variables of interest. Another important limitation relates to the matching of complex sample surveys. This aspect is particularly exacerbated when the final integrated database is obtained after more than two linkages. Because of the potential accumulation of sources of imprecision as more surveys are fused mixing data from different clusters and strata, the reliability of the results may be affected. This is a relevant issue that, in our view, deserves greater research attention.

The rest of the article is organized as follows Section 2 describes the methodology to implement statistical matching using the propensity score approach. The single data sets are delineated in Section 3. Section 4 provides a detailed description of the three statistical match procedures and analyzes both the statistical and economic robustness of the outcomes. Section 5 illustrates an empirical application about the measurement of multidimensional poverty in Italy that exploits the fused living standard database. Section 6 summarizes the main findings and draws conclusions that could be useful for future. The supplemental material consists of Tables A1–A18 and Figures A1–A8.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

2. The Statistical Matching Method

Statistical matching techniques enable the integration of two or more data sources that refer to the same target population and share a common set of variables. Matching combines information observed in a donor data set, which can also be considered the control group, with units of a recipient data set, which can be considered as the treatment group, with missing values for those variables. The donor data set is the database that contains the extra information and normally includes the largest number of observations. In practice, statistical matching can be seen as a method of variable imputation from a donor to a recipient survey (Rubin and Schenker 1986; D'Orazio et al. 2006a; Kum and Masterson 2010; Tedeschi and Pisano 2013; Donatiello et al. 2014).

Let A and B be two independent samples of size n_A and n_B respectively, drawn from the same population. Variables Y are observed only in A, while variables Z are observed only in B. A set of variables X are collected in both samples and are correlated with both Y and Z. The main goal of statistical matching is to estimate the joint distribution of (Y, Z, X) or at least on the pairs of target variables that are not observed jointly (Y, Z). The relation between these common variables and the specific variables observed only in one of the data sets is used to impute from a donor data set A information on Y in the recipient data set B for similar units and a synthetic dataset is generated with complete information on X, Y and Z representative of the population of interest.

Statistical matching methods can be classified into three broad categories: non-parametric methods such as the constrained or unconstrained hot deck method; regression-based parametric methods; and mixed methods. Hot deck imputation involves replacing missing values with values from a donor unit similar in terms of common characteristics. A hot deck application is random when the donor is selected randomly from a donor pool. The constrained hot deck method ensures that each record in the donor file is used only once to impute the non-observed variables in the recipient file using values really existing in the donor file. Mixed methods involve a combination of parametric and non-parametric techniques in a two-stage process such as the predictive mean matching imputation method or the propensity score matching.

This study adopts the latter approach. Statistical matching is a delicate exercise because of the dimensionality problem related to the high number of shared covariates, the number of possible values of categorical variables, and the presence of continuous variables that can reflect many different values. The propensity score is one possible balancing score that deals with the high dimensionality of the procedure reducing the problem to one-dimension. There are other attractive ways to deal with the dimensionality problem, such as the predictive mean matching (PMM) also when integrated in hot deck matching schemes (Kum and Masterson 2010; Leulescu and Agafitei 2013). The hot deck matching tends to break down when the sample size is small or the set of selecting variables is large, because the pool of potential donors is limited and robust matches are rare (Mittag 2013). Andridge and Little (2010) contend that very little is known about the theoretical properties of hot deck procedures. On the other hand, because the hot deck is a nonparametric technique, it is less exposed to model misspecification.

Rosenbaum and Rubin (1983) proposed the use of *balancing scores* applied to the most relevant observed common variables. The balancing score $b(X)$ is a function of the

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

observed covariates X such that the conditional distribution of X given $b(X)$ is independent ($\perp$) of assignment in the treatment (D) $D \perp X|b(X)$. Originally, this technique was introduced to estimate causal effects between treated and control groups in non-randomized experiments.

The propensity score is estimated using a logistic or probit regression specified on the selected set of covariates that are common to all questionnaires, and its estimated score can be considered a synthetic indicator of the shared variables used in this function. The propensity score is the conditional probability of assignment to a particular treatment conditional on a set of observed covariates $p(X) = prob(D = 1|X)$, where D is an indicator equal to *1* if an observation refers to the treated group and *0* otherwise.

For a statistically robust application of the propensity score, the assumptions normally made when implementing a statistical matching procedure can be stated in a randomized trial context (Rosenbaum and Rubin 1983):

Conditional independence: given a set of common covariates that are not affected by treatment, the potential outcomes are independent of treatment assignment

$$D \perp Y_0, Y_1|X => D \perp Y_0, Y_1|p(X).$$

Common support: observations with the same covariate values have a positive probability of being both in treated and untreated

$$Y_0, Y_1 \perp D|X.$$

The conditional independence assumption asserts that the outcome in the control group is independent of the treatment D conditional on the selected set of covariates. In the early statistical matching implementations, it was frequent to assume the independence of the never jointly observed variables Y and Z given the set of common variables X, $f(x,y,z) = f_{Y|X}(y|x)f_{Z|X}(z|x)f_X(x)$ where $f_{Y|X}$ is the conditional density function of Y given X, $f_{Z|X}$ is the conditional density function of Z given X and f_X is the marginal density function of X (D'Orazio et al. 2006a). Conditional independence rarely holds in practice. In a statistical matching context where only A and B are available it is not possible to test the conditional independence assumption. Modern applications exploit, when possible, relevant information from an auxiliary data source to overcome the conditional independence assumption (Donatiello et al. 2014) and evaluate the uncertainty associated with the lack of joint information about the variables of interest (Conti et al. 2017).

The common support requirement states that the distribution of observed covariates is as similar as possible in both groups. This assumption ensures that there is an overlap in the characteristics of treated and untreated observations sufficient to have potential matches in the untreated group.

Note that when using the terms treated and control in the context of statistical matching rather than a randomized trial context, we refer to the treated group as the recipient data set and to the control group as the donor data set. This analogy says that the treated group is the recipient of the treatment, that is, the additional information coming from the control (donor) data set that donates information (treats) the recipient. In a multiple matching

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

exercise, as it is in our application, there are multiple donor data sets contributing information to the single recipient data set.

Another relevant assumption underlying the implementation of a statistical matching procedure is that the processes generating the missing data is missing completely at random (MCAR). There is no systematic relationship between the propensity of missing values and any data, either observed or missing, because missingness is induced by the sampling design (D'Orazio et al. 2006a). In general, ignorability assumes that missing data can be considered as occurring effectively at random, so that the effects of the unobserved, possibly confounding, factors and missing data can be ignored. Strong ignorability (Rosenbaum and Rubin 1983) presumes that the conditional independence assumption holds and that there is common support, or overlap, between the data sets. In most cases, it is difficult to validate the ignorability assumption because statistical matching suffers from the identification problem concerning the association of the variables never jointly observed, given that the variables common to both data sets cannot be estimated from the observed data. This is a general problem that affects all statistical matching procedures, not just the propensity score. The validity of a matching technique concerning the preservation of the true association of the variables never jointly observed depends on the explanatory power of the common variables (Rässler 2002, 2004; Kiesl and Rässler 2009). Given these common variables, the variables not jointly observed can be more or less independent after statistical matching.

For every variable specific to each data set to be fused, the marginal joint cumulative distribution function is bounded by the Fréchet inequality (D'Orazio et al. 2006a,b, 2009, 2017; Kiesl and Rässler 2009; Conti et al. 2012; Conti et al. 2017). The range of these bounds may be used to evaluate the data fusion procedure, although the bounds may not represent a sufficiently stringent interval to be useful in all practical situations. In general, the higher the explanatory power of the common variables and the narrower the bounds of the association, the more reliable are the matching results at all interesting levels of validity. In any event, it is important to recognize that, from the observed data, we are not able to uniquely recover the underlying joint distribution that could have generated the data because of the range indeterminacy.

In Subsection 4.1, we investigate uncertainty stemming from the identification problem associated with the lack of joint information on the variables of interest by calculating the Fréchet inequality for the contingency table associating income and expenditure classes. This is an especially important economic relation not only for the estimation of short-term savings but also for the related measures of well-being, poverty and inequality (Donatiello et al. 2014; Conti et al. 2016; Conti et al. 2017). The distance between bounds is affected by the number of classes and by the elements included in the set of matching variables. Shorter intervals decrease uncertainty and as a consequence increase trust in the conditional independence assumption. It is in this sense that the analysis of uncertainty can be viewed as a measure of the relevance of the conditional independence assumption and the overall quality of the procedure, and as a specification tool for selecting the most appropriate set of matching variables.

The assumption of conditional independence is especially untenable in the case of consumption and income, although conditional independence seems to be an innocent assumption when the matching variables include a reliable proxy for income as auxiliary

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

information (Singh et al. 1990; Donatiello et al. 2014; Conti et al. 2016; Conti et al. 2017). In the HBS survey, information about aggregate household incomes is recorded in large intervals as it is stated by respondents, while in the EUSILC database it is constructed with a much higher level of detail on all different types of income earned by all household members. Though affected by large measurement errors, it maintains a high correlation with income. Thus, it may serve as reliable auxiliary information (Singh et al. 1993; Coli et al. 2005; Donatiello et al. 2014). Because the income section of the HBS is not available to users that do not belong to ISTAT, we imputed income at the individual level using information from EUSILC and then summed individual incomes to determine household income. As predictors included in the multiple imputation procedure using the predictive mean matching method, we used the variables region, family type, age, gender, education level, occupational status, job, part-time or full-time worker, and the distinction between dependent or self-employed worker. Predicted income was then used as a matching variable and included in the specification of the logistic model estimating the propensity score, where it performed with high explanatory power.

2.1. Implementation of the Statistical Matching Method

We now describe in sequence the steps adopted to implement our statistical matching procedure.

1. **Harmonization of the data sets**. The first step of the matching procedure harmonizes the common variables across data sets by comparing and adjusting the definitions and classifications to make them homogeneous. We also need to choose the best set of "matching variables" observed in both data sets that have a significant relationship with the variables of interest. A correct selection of variables controls for differences within groups because the selected variables need to be independent of the group assignment, thus affecting the outcome but not the exposure. The model specification involves a trade-off between the common support condition and the plausibility of the conditional independence assumption. A parsimonious specification may not affect common support, but may affect the plausibility of conditional independence, while a full specification may give rise to a support problem by affecting the common support condition (Black and Smith 2004; Caliendo and Kopeinig 2008). The main purpose of the propensity score estimate is to balance all covariates, not to define the best selection into groups (Augurzky and Schmidt 2001).
2. **Compare the distribution of X**. To inspect whether the common variables are independent of sample selection, we compare the marginal and joint distribution in the recipient and donor group by testing the similarity in distribution and calculating the between groups distance using both the absolute difference and Cramer's V test (Sisto 2006; Masterson 2010; Leulescu and Agafitei 2013). Distributions can be also compared using the Hellinger distance. In our context, this measure is always coherent and consistent with Cramer's V test, which is our selected test. Both the Hellinger distance and Cramer's V assume values between 0 and 1. A value close to 0 means that the relationship between the two distributions is weak. For Cramer's V test, the acceptance threshold of weak relationship is 0.15. Before matching, the

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

common set of variables may have statistically different distributions, but after the implementation of the propensity score matching procedure, the common set of variables should be balanced within the strata.

3. **Estimate the selected statistical matching method** (Propensity Score Matching). The matching variables are then used to estimate the propensity score value. The set of matching variables is specific to each pair-wise matching that we describe in the next sections.
4. **Validate the propensity score procedure** by a) computing balancing tests, and b) checking the overlap and region of common support between the two groups. As summarized by Lee (2013), to validate the result of the selected propensity score specification, four balancing tests are recommended: i) standardized differences proposed by Rosenbaum and Rubin (1985) for evaluating the bias reduction due to the success of the matching procedure, and consequently analysis of the distance in marginal distributions of the common variables; ii) t-tests to evaluate the equality of each covariate mean between the recipient and donor groups (Rosenbaum and Rubin 1985); iii) stratification test for testing the mean differences within strata of the propensity score (Dehejia and Wahba 1999, 2002); iv) Hotelling test or F-test to verify the joint equality of covariate means between the reciepient and donor groups (Smith and Todd 2005).

The standardized difference was computed as the percentage of the ratio between the difference of sample means in the recipient and donor subsamples and the square root of the average of sample variance in both groups. Following Rosenbaum and Rubin (1985) a standardized difference is "large" if it is greater than 20. We also computed a t-test to verify if the mean of each common variable between the recipient and the donor database is not statistically different before and after the matching.

The stratification test was developed in two steps. In the first phase the observations were divided into strata. To determine the number of strata, the estimated propensity score was split into ranges provided that its mean within each stratum was not statistically different in the recipient and donor group. In the second step, for each stratum a t-test was performed to test whether the common covariates presented the same distribution in both groups (Dehejia and Wahba 2002; Caliendo and Kopeinig 2008; Garrido et al. 2014). If the t-test is rejected in even only one stratum, then the propensity score model is not well specified and the specification should be corrected until there are no significant differences between the two groups and the conditional independence assumption is more likely to hold (Caliendo and Kopeinig 2008; Lee 2013).

The Hotelling test is used to jointly test the equality of the means in all covariates used in propensity score specification, between the recipient and donor data set. If the null hypothesis is rejected, there is no balance in covariates between the two data sets. This test is adopted in multivariate tests of hypotheses and it is the generalization on the t-test used in univariate problems.

To assess whether the characteristics observed in the recipient group are also observed in the donor group, it is important to verify the overlap of the region of common support of the propensity score value between these two groups (Lechner 2008). This investigation is crucial because the lack of common support may lead

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

to biased results since the donor group may not be sufficiently similar to the recipient one. A graphical analysis of the density distributions of the propensity score in the recipient and donor group permits a visual inspection of the range and shape of the propensity score distributions (Caliendo and Kopeing 2008). The estimated propensity score is then used to match each individual in the recipient group to an individual in the donor group.

5. **Choose the matching algorithm**. Rodgers (1984) distinguishes between the constrained and unconstrained algorithm types. An unconstrained method imposes no restrictions on the number of times a donor unit may be imputed because it takes simple random samples with replacement. It has the advantage of permitting the closest possible match to each record at the cost of increasing the sample variance of the estimators (Rodgers 1984; Rässler 2002; Kum and Masterson 2010). The distributions of the imputed variables are therefore more likely to represent the empirical marginal or conditional distributions of the selected sample, rather than the ones observed in the original donor file. Despite this disadvantage, unconstrained matching is still the method most frequently used (Rodgers 1984; Kum and Masterson 2010). On the other hand, a disadvantage of the constrained method is that the average distance between the recipient and donor values of the matching variables is plausibly larger, and sometimes unacceptably larger, than in the unconstrained case because matching is implemented without replacement. It is important to remark that the use of sampling weights make sure that donor records can be matched to more than one recipient and vice versa. From a practical point of view, constrained matching is computationally more demanding than unconstrained matching.

   The main matching algorithms are nearest neighbor, caliper and radius, stratification and interval, kernel and local linear, and weighting (Chen and Shao 2000; Caliendo and Kopeinig 2008; Kum and Masterson 2010). The choice in regard to performing a matching with or without replacement and the number of comparison units involves a trade-off between bias and variance. The two aspects are inter-related because, for example, a matching with replacement and a smaller number of comparison units reduces both the bias and the precision (Dehejia and Wahba 2002). All methods yield similar results with large samples, while the trade-off between bias and variance is mainly relevant for small samples. As a result, there is not a better matching algorithm, but its choice should be evaluated case-by-case on the data structure (Caliendo and Kopeinig 2008). We compared different matching algorithms. Our preferred choice was the nearest neighbor algorithm with replacement and one comparison unit because it was the most effective algorithm in preserving the distribution of the donor data set as it is described in Subsection 4.1. For each individual of the recipient database we selected the individual in the donor database with the closest distance in terms of propensity score. The matching algorithm imputed the missing values of the recipient sample using the information from the donor sample.

6. **Assess the statistical matching quality** by a) inspecting distributions, b) analyzing the trend of the imputed variables by the set of X covariates comparing the ratio of mean and median in the two groups, and c) performing uncertainty analysis by

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

computing Fréchet Bounds between the variables of interest. Rässler (2002) describes four levels of validity to evaluate a matching procedure: preserving individual values, preserving joint distributions, preserving correlation structures, and preserving marginal distributions. In most cases, only the last level, which establishes a minimum validity requirement, can be verified, although recent literature shows that both the preservation of the joint distribution and the preservation of the correlation structure can be evaluated (Conti et al. 2016; Conti et al. 2017). Statistical matching can be considered successful if the marginal and the joint distribution of the covariates and the imputed information show similar trends in the original and the synthetic databases. We assessed the matching procedure by both inspecting the distributions of the extra information in the two databases, and comparing the distribution of the imputed covariates by the set of common variables used in the propensity function, computing the ratio of mean and the ratio of median (Kum and Masterson 2010; Webber and Tonkin 2013). The ratio of mean (median) is calculated as the ratio between the mean (median) of the recipient data set and the mean (median) of the donor data set. To demonstrate whether the two groups are different in the means or medians, we consider the distance of the ratio from 100, being the value that represents the perfect similarity in the means or medians of the two groups. There is no defined threshold to establish if the imputed information in the two samples can be considered comparable, but the closer the ratio is to 100, the greater the similarity of the extra information.

Further, as part of the statistical evaluation, it is important to deal with the source of indeterminacy stemming from the conditional independence assumption and improve the overall quality of the procedure by exploring the degree of uncertainty associated with the matching results, as we did in our empirical application, and possibly exploiting auxiliary information when available, or introducing meaningful logical constraints (D'Orazio et al. 2006b; Conti et al. 2016; Conti et al. 2017).

7. **Assess the economic matching quality** using Engel curves, poverty and inequality analysis.

3. Data Sets Description

In the following section, we briefly describe the four surveys used in this work. Subsequently, we analyze the characteristics and properties of each statistical matching performed.

The implementation of the Stiglitz et al. (2010) proposal to measure well-being in a comprehensive manner based on an extended notion of income that accounts for the value of private and public consumption, working and nonworking time, financial and social assets, requires the integration of several sources of information about households. We now describe the data sets related to the consumption, income and wealth, time use and social dimensions that we combined to construct a multidimensional measure of economic well-being representative of the Italian population. This objective requires adopting a matching procedure that is careful to preserve at least the marginal distribution of the main

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

economic and social variables of interest, paying especial attention to the varying sampling designs of each data set.

3.1. European Union Statistics on the Income and Living Conditions Survey (EUSILC): The Recipient Survey

EUSILC is an annual statistical survey that gathers comparable cross-sectional and longitudinal data for the EU Member States. In Italy, the National Statistical Institute (ISTAT) conducts the survey. The EUSILC sample is drawn with a two-stage sampling design where primary units are municipalities and secondary units are households. A sample of 760 municipalities is selected, according to a conditional Poisson design with inclusion probabilities proportional to demographic sizes within strata. From each selected municipality, households are drawn by simple random sampling. We use the 2010 sample of 19,147 households corresponding to 47,551 individuals. The sampled households are selected with a rotational design where a fraction of the sample of the previous survey is dropped and replaced with a new sample of equal size maintaining the same representativeness of the whole population. The survey collects information on incomes, wealth and living conditions at both the household and individual levels. EUSILC also gives detailed information on socio-demographic characteristics, housing conditions, health and education, employment status, economic activity and other firm-specific attributes.

3.2. Household Budget Survey (HBS)

The ISTAT consumption survey collects detailed information on household expenditure on goods and consumer services in diverse categories, such as foodstuffs, clothing, housing, transport, education, health and holidays. Expenditures in the HBS are classified using the United Nations' five-digit Classification of Individual Consumption According to Purpose (COICOP) classifications. The main aim of this survey is to analyze and evaluate the trend in household expenditure in relation to the socio-demographic characteristics of family members. We used the data collected in 2009. The HBS sample is drawn with a two-stage sampling design. The primary sampling units are municipalities. They amount to around 470 selected among two groups according to a conditional Poisson design with inclusion probabilities proportional to demographic sizes within strata. From each selected municipality, households are drawn by simple random sampling. The sample is composed of 23,005 households.

3.3. Time Use Survey (TUS)

The TUS records the time employed in daily activities by each household member. The respondent keeps a diary reporting the main activity undertaken, any other activity taking place at the same time, and the places in which the activities are carried out. Each family, selected according to a random procedure, compiles a diary for either one day of the week, Saturday or Sunday according to the day of the visit. To implement the matching procedure, we first imputed the time spent on each activity for those days that the household member did not have to fill in the diary. The TUS also reports on socio-demographic characteristics, education, economic activity, housing, and health conditions.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

The sampling design is implemented in two stages. The first stage units are municipalities (508) and the second stage units are households. The interviewees are each family member aged three or over. The 2008–2009 cross-sectional wave interviewed 18,250 households and 44,606 individuals.

3.4. Household Conditions and Social Capital Survey (CISF)

The survey on household conditions and social capital was designed by the International Center of Family Studies (CISF) in 2009 with the aim of describing the well-being of Italian families and their stock of social capital. The survey was carried out through telephone interviews by COESIS. It collects household level data about socio-demographic characteristics, income and overall economic condition, and a detailed set of questions on social capital and relational well-being. The sampling design is stratified by geographic areas and family types. The sample includes 4,017 households and has both national and macro-regional representativeness. Unlike the others, the CISF survey is not scheduled with regular frequency. It is the only survey not implemented by the Italian National Statistical Institute (ISTAT) included in our integrated data set.

In general, multi-stage cluster and stratified sampling are two distinctive features of complex surveys such as those used in this statistical matching exercise. As a consequence, observations cannot be assumed to be independent and do not have equal probability of being selected, as is the case of simple surveys. Observations that are from the same cluster or strata are likely to be more similar to each other. Ignoring the sampling design may introduce serious bias in both the imputation method and the outcome models. Several authors (D'Orazio et al. 2006a; Ridgeway et al. 2015; Conti et al. 2016; Austin et al. 2018) have analysed how to account properly for complex designs and the different survey weights when implementing a statistical matching procedure, placing especial emphasis on Renssen's two-step procedure (Renssen 1998) based on calibration of the weights and Rubin's (Rubin 1986) file concatenation.

In our analysis, we minimized the adding complexity of different survey designs by selecting the three main surveys to be matched (EUSILC, HBS, and TUS) from the same statistical institute. Instead of integrating the income information from EUSILC, we could have selected the Survey on Household Income and Wealth (SHIW) that is conducted by Banca d'Italia every two years. The SHIW survey, which is part of the Household Finance and Consumption Survey of the European Central Bank, is also conducted in two stages. Municipalities with more than 40,000 inhabitants are all included in the sample, while smaller primary units are selected using a probability sampling scheme proportional to size. Secondary sampling units are then selected by simple random sampling. On average, the sample comprises about 8,000 households (20,000 individuals) distributed across around 300 Italian municipalities. The SHIW size of both primary and secondary units is about 1/3 of the HBS size. Conti et al. (2016, Table 2) show that the estimated proportions of households, conditional on two main design variables such as macro-region and household size are not significantly different between EUSILC and HBS. In the context of the present application, this is also the case for all the ISTAT data bases EUSILC, HBS and TUS. There are no significant differences also for the CISF database, which is not produced by ISTAT. Therefore, with

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

the intent of not adding complexity, we preferred EUSILC to SHIW even though we recognize that SHIW is interesting for the higher value of the information on the value of assets, debts and regular savings with respect to EUSILC. Our choice was also due to the consideration that ISTAT is actively committed to improving the *ex-ante* harmonization of the EUSILC, HBS, and TUS social surveys and in complementing the wealth dimension as part of the revision process under development within the new European Framework Regulation on Social Statistics. Moreover, in recent years, the Italian version of EUSILC has consistently made use of registered data that cross-verify the income data collected through surveys using available social security and tax records.

As part of our specification strategy of the propensity score regression, in the set of matching variables we included some relevant variables of the sampling design such as regions and household characteristics. According to Kum and Masterson (2010), the propensity score matching method's dimensionality reduction is effective in minimizing the potential bias that may stem from the complex designs of the fused data sets.

Figure 1 illustrates how consumption, time use, and social capital donor data sets have been linked to the income and wealth survey. The donor data sets include the extra information missing in the recipient database. The recipient data set contains the most detailed and accurate information about common variables gathered in all surveys. Combining these relevant dimensions of well-being yields a "new" database, to which we refer as the Italian Integrated Living Standard survey (IILS).

To respect the temporal correspondence between income and related variables, we used the 2010 cross-sectional wave for the EUSILC survey because the information on income refers to the previous reference period. We used the 2009 cross-sectional wave for the HBS and the 2008–2009 wave for the TUS.

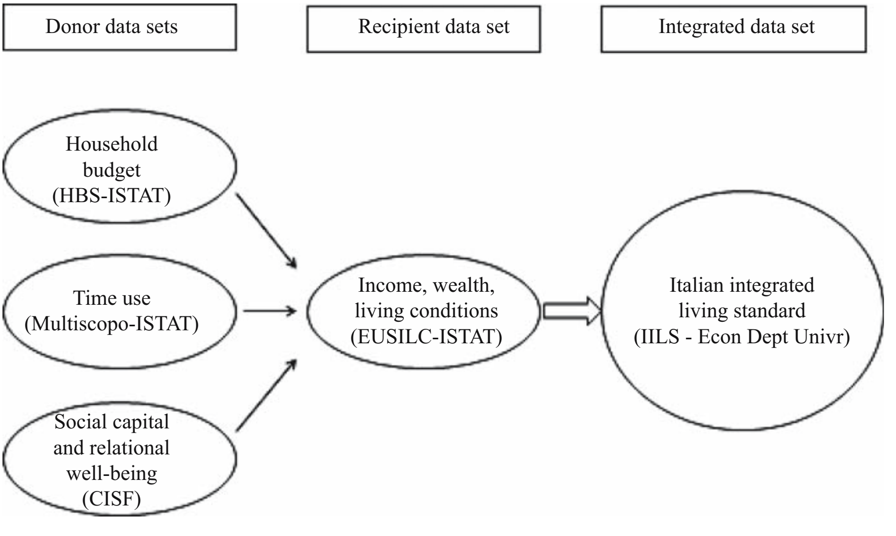


Fig. 1. The data sets used to create the integrated database.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

In the next section, we describe the features of each one-to-one matching implemented following the sequential representation of Figure 1 and evaluate the statistical and economic quality of the linking procedure.

4. Results of the Statistical Matching Procedures

We implemented three different statistical matching procedures using the EUSILC data set as the recipient sample because this survey includes the most detailed information regarding socio-demographic characteristics, household conditions, occupational status, income, wealth, health and education.

In sequence, the first linking procedure performed the data fusion between the EUSILC and HBS data set to impute the information related to household consumption. The second statistical matching associated the information about household time use with the EUSILC data set. The third matching filled in the missing values of the EUSILC data for social capital, family relationships and family well-being, using the CISF survey.

The three data fusions were implemented using the method outlined in Section 2. For illustrative convenience, we report the EUSILC-HBS match only. For this matching, we describe a) the alignment of common variables, b) their frequency distributions, c) the standardized differences and t balancing tests, d) the distribution of propensity score, e) the distribution of the extra information imputed with the propensity score procedure, in the original and matched data sets, f) their ratio of mean and median by covariates, and g) the investigation of uncertainty constructing the Fréchet bound, and implement an economic evaluation of the statistical procedure. The results of the EUSILC-TUS and EUSILC-CISF matching procedure are reported in the supplemental material.

4.1. *Data Fusion Between the EUSILC and HBS*

The EUSILC database does not record data about family consumption that is typically collected in household budget surveys. As shown in Figure 1, we add household consumption to the former survey. We aggregate detailed household expenditures into nine categories: cereals; meat, fish and dairy products; fruit and vegetables; other food products; clothing; housing; transport and communication; recreation and education; health and hygiene.

The two basic conditions for implementation of the statistical matching are satisfied. Both samples refer to the same target population and share a set of covariates related to socio-demographic characteristics, household characteristics and working status conditions. The common variables are defined in the same way in both surveys. Table 1 documents how we harmonized and aggregated the variables of major interest to achieve the same alignment omitting trivial reclassifications.

The adopted propensity score specification that satisfied the balance property includes: region of residence (five dummies coded as North-West, North-East, Center, South, Islands), a dummy variable to indicate the presence in the family of children between 0–5 years old, and between 6–14 years old, a dummy variable to denote the presence in the family of at least one self-employed worker, a single-parent dummy, home-ownership (dummy variable equal to 1 if the household head is a home-owner), average family

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 1. Alignment of common variables in EUSILC – HBS match.

Variable	EUSILC	HBS	Harmonized variable
Education	*istr_c* 1 = unqualified, illiterate 2 = unqualified, can read and write 3 = primary school 4 = first grade secondary school 5 = second grade secondary school (2–3 years) 6 = second grade secondary school (4–5 years) 7 = certificate post-A levels 8 = bachelor's degree or master's degree 9 = superior graduate school 10 = Ph.D.	*titstu* 1 = Ph.D or superior graduate school 2 = master's degree 3 = bachelor's degree 4 = second grade secondary school (4–5 years) 5 = second grade secondary school (2–3 years) 6 = first grade secondary school 7 = primary school 8 = unqualified	1 = unqualified 2 = primary school 3 = first grade secondary school 4 = second grade secondary school (2–3 years) 5 = second grade secondary school (4–5 years) or certificate post-A levels 6 = bachelor's degree or master's degree 7 = superior graduate school or Ph.D.
Status in employment	*pl040* 1 = self-employed with employees 2 = self-employed without employees 3 = employee 4 = family worker	*posprof* 1 = executive 2 = manager 3 = clerk 4 = intermediate categories 5 = foreman 6 = other employee 7 = trainee 8 = homeworker 9 = military force (armed force) 10 = entrepreneur 11 = self-employed 12 = independent contractor	1 = self-employed 0 = other

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 1. Continued.

Variable	EUSILC	HBS	Harmonized variable
		13 = partner of cooperatives 14 = assistant 15 = project worker 16 = occasional contractor (from code 1 to code 9 employee, from code 10 to code 16 self-employed)	
Tenure status of the house	*hh020* 1 = owner 2 = tenant or subtenant paying rent at prevailing or market rate 3 = accommodation is rented at a reduced rate (lower price that the market price) 4 = accommodation is provided free	*tipoccup* 1 = tenant or subtenant paying rate 2 = owner 3 = accommodation is in usufruct 4 = accommodation is provided free by relatives or friends	1 = owner 0 = other

Note: The name of the variables in each survey are indicated in italics.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

education (five dummies coded as Primary, Middle, Middle-High, High, University) and total disposable household income.

Table 2 shows the frequency distribution of the variables used in the propensity score specification. Geographical area shows the largest absolute differences. The value of Cramer's V test supports the hypothesis that the common variables are independent of the group assignment. Therefore, considering a threshold of 0.15 associated with a weak

Table 2. Comparison between frequency distributions for some common variables.

	EUSILC	HBS	Absolute difference	Cramer's V*
Geographical area				0.094
North-West	23.03	23.58	0.55	
North-East	24.04	21.15	2.89	
Center	22.97	17.62	5.35	
South	21.36	26.61	5.25	
Islands	8.60	11.04	2.44	
Children 0–5 years old				0.020
No	88.53	89.75	1.22	
Yes	11.47	10.25	1.22	
Children 6–14 years old				0.011
No	84.01	83.18	0.83	
Yes	15.99	16.82	0.83	
Self-employed				0.005
No	80.51	80.15	0.36	
Yes	19.49	19.85	0.36	
Single-parent				0.025
No	91.41	92.78	1.37	
Yes	8.59	7.22	1.37	
Homeownership				0.009
No	25.50	24.72	0.78	
Yes	74.50	75.28	0.78	
Average family education				0.036
Primary	26.95	26.83	0.12	
Middle	24.28	27.24	2.96	
Middle-High	19.16	18.15	1.01	
High	23.16	21.48	1.68	
University	6.44	6.29	0.15	
Household income				0.025
1st quintile	19.51	20.41	0.90	
2nd quintile	19.48	20.44	0.96	
3rd quintile	20.17	19.85	0.32	
4th quintile	19.85	20.13	0.28	
5th quintile	20.99	19.17	1.82	

*The acceptance threshold of a weak relationship is 0.15.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 3. Test for standardized differences and t-test on the equality of means.

Variable	Test for Standardized differences		T-test	
	Standardized difference before matching	Standardized difference after matching	P-value before matching	P-value after matching
Geographical area				
North-West	−1.30	0.00	0.1820	0.9750
North-East	6.90	−1.70	0.0000	0.1650
Center	13.30	2.00	0.0000	0.0880
South	−12.30	−0.20	0.0000	0.8980
Islands	−8.20	−0.20	0.0000	0.8510
Children 0–5 years old	3.90	−2.80	0.0000	0.0180
Children 6–14 years old	−2.30	−1.50	0.0220	0.1990
Self-employed	−0.90	−1.00	0.3560	0.3900
Single-parent	5.10	−0.80	0.0000	0.4930
Homeownership	−1.80	0.40	0.0640	0.7180
Average family education				
Primary	0.30	4.10	0.7790	0.0010
Middle	−6.80	−1.40	0.0000	0.2500
Middle-High	2.60	−1.80	0.0080	0.1240
High	4.00	−1.50	0.0000	0.2030
University	0.60	0.60	0.5280	0.6030
Household income	7.60	−1.60	0.0000	0.1850

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

relationship, we can conclude that all these variables are independent of the groups. This conclusion is generally supported by the evidence presented in Table 3. Before matching, all standardized differences between recipient and donor groups were less than 20%, indicating that the two data sets are similar. The magnitude of these differences decreased after matching, becoming very close to zero. We use the test of standardized differences to illustrate the reduction in bias that can be attributed to matching on common variables (Rosenbaum and Rubin 1985; Lee 2013). Table 3 also shows the p-values of the t-test to compare the means of the common variables. As pointed out by Rosenbaum and Rubin (1985) and Caliendo and Koepinig (2008), it is reasonable to expect differences before the matching execution. After matching, the covariates should be balanced in both groups and hence no significant differences should be found, as is the case in Table 3. In general, the balance in covariates is less likely to be achieved by covariates that do not significantly impact the outcome (Garrido et al. 2014). Before matching, there are many covariates that do not have the same proportion, but after matching the proportions in the recipient and donor groups become equal. The sole exception is represented by the "Primary" category of education, which is balanced before matching but after matching does not show the same mean in the two samples. The Hotelling test also confirms that the covariates are balanced between the two groups. The null hypothesis of joint equality of the means is not rejected (Table 4).

These statements are supported by the evidence presented in Table 5. Conditioning on the propensity score, all variables are balanced within the two samples. The upper part of the table shows t-test values verifying whether the density distributions of the propensity score are equal in the two selected samples within each stratum. The lower part shows the

Table 4. Hotelling test after matching.

Variable		Mean of HBS	Mean of EUSILC
Geographical area			
North-West		0.230	0.230
North-East		0.248	0.240
Center		0.221	0.230
South		0.214	0.214
Islands		0.087	0.086
Children 0–5 years old		0.124	0.115
Children 6–14 years old		0.166	0.160
Self-employed		0.199	0.195
Single-parent		0.088	0.086
Homeownership		0.743	0.745
Average family education			
Primary		0.251	0.269
Middle		0.249	0.243
Middle-High		0.199	0.192
High		0.238	0.232
University		0.063	0.064
Household income		3.197	3.158
Hotelling p-value	0.069		

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 5. Test statistic of t-test in each stratum.

	Stratum										
	2	3	4	5	6	7	8	9	10	11	12
Balance of propensity score distribution across recipient and donor groups	−0.077	1.368	1.332	−0.076	−1.218	−2.238	−0.654	−0.425	−1.980	−0.593	−1.381
Balance of covariates across recipient and donor groups											
Geographical area – ref. cat. "North-West"											
North-East	−1.287	−1.200	0.747	−1.346	0.336	0.681	2.501	0.279	1.127	−0.586	−0.775
Center	–	–	−0.984	0.576	−0.330	−0.104	−2.117	−0.254	−0.719	−0.840	−0.775
South	−0.987	2.089	−2.197	0.194	−1.116	−0.561	−2.189	–	−0.873	–	–
Islands	1.471	−1.586	−0.961	−0.169	−0.731	−0.589	0.033	–	−0.873	−0.622	–
Children between 0–5 years old	0.591	−0.539	−0.717	−0.429	0.036	0.884	0.483	−1.016	1.029	1.066	−0.775
Children between 6–14 years old	0.842	0.284	0.746	0.081	−0.563	0.813	−1.892	−1.766	0.331	−1.568	−0.775
Self-employed	1.210	0.907	0.343	1.842	−1.040	−0.768	−0.226	−2.346	−0.877	−1.031	–
Single-parent	−0.279	2.350	0.287	0.627	−1.750	−0.062	−0.050	0.357	−0.100	−0.048	–
Homeownership	1.121	−1.716	2.048	0.118	−1.077	2.482	−1.434	−0.905	−1.465	−2.269	–
Average family education -ref. cat. "Primary"											
Middle	−0.253	0.823	1.416	0.226	−0.529	1.156	0.550	−0.329	0.480	−1.090	–
Middle-High	−2.494	1.061	−0.595	1.186	0.733	−0.961	1.547	−0.014	−0.117	−0.992	−0.775
High	0.685	1.555	−1.714	1.661	−1.370	−0.612	0.042	0.601	−1.080	0.079	−1.549
University	−0.007	−2.082	−0.495	0.456	0.388	0.774	−0.636	0.276	1.533	0.798	–
Household income	2.326	1.124	−0.092	2.256	−0.981	0.050	−0.217	−1.203	−1.518	−0.508	−0.793

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

t-test values carried out to determine whether the common covariates have the same distributions in the two data sets. The first stratum is not shown because the propensity score takes values higher than the first quintile into which the sample was initially divided. Considering a 0.01 significance level, the propensity score and the common covariates have the same distribution in the two samples.

We also performed a preliminary test to investigate the region of common support of the propensity score value. As shown in Figure 2, the estimated propensity score takes values in a similar range and displays comparable density distributions. Therefore, the observations have the same probability of belonging to the recipient or the donor group.

In addition, we implemented a comparative analysis of different matching algorithms such as radius, caliper, Epanechnikov and Gaussian kernels, nearest neighbor with and without replacement and multiple comparison units. The results of all algorithms are consistent in mean but they differ in distribution. Extra information imputed using radius, caliper and both kernel matching algorithms produce mean and median values that are similar to the same statistics of the original distribution, but standard deviations are significantly smaller compared to the original variables. On the other hand, the distribution of imputed values generated using the nearest neighbor algorithm is the most similar to the donor's distribution with and without replacement, and with different comparison units. Table 6 reports these results for the three main consumption categories: cereals, protein foods such as meat, fish and dairy products, and clothing. When adopting one comparison unit, there are no significant differences between distributions with and without replacement. As the number of comparison units increase, differences become more marked, especially in terms of standard deviations. In light of these results, for our matching exercise we adopt the nearest neighbor algorithm with replacement and one comparison unit.

To verify the matching quality, we analyzed the distribution of the extra information transferred from the donor to the recipient. We tested whether the extra information in the matched data set preserves the same distribution as the original data set. We also compared the distributions of the covariates used in the propensity score specification by computing

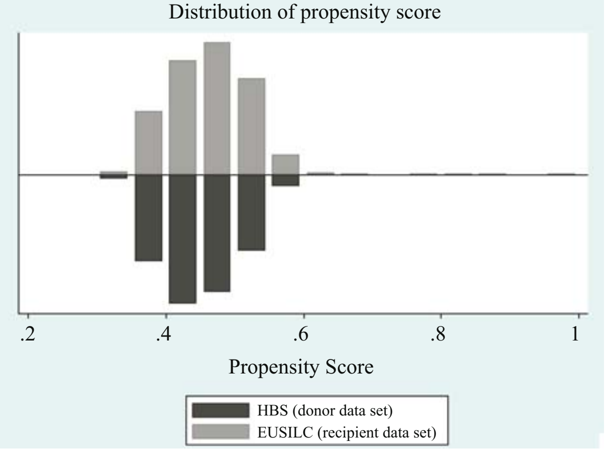


Fig. 2. Distribution of propensity score across recipient and donor data sets.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 6. Comparison of statistic distributions using different nearest neighbor matching algorithms (in euros).

		Nearest neighbor					
			With replacement and (*1,....,5*) comparison units				
	Donor database	Without replacement and 1 comparison unit	1	2	3	4	5
Cereals							
Mean	69.2	69.72	70.6	70.37	70.25	70.03	70.18
Median	59.28	59.69	60.33	64.91	66.44	66.93	67.95
Std. Dev.	46.1	46.43	46.96	33.26	27.19	23.45	20.92
Min	1.37	1.37	1.7	4.58	12.25	11.29	15.55
Max	469	469	469	291.97	222.27	203.22	194.14
Meat, fish and dairy products							
Mean	213.41	213.33	216.38	215.57	215.46	214.95	215.22
Median	181.53	181.46	184.52	196.71	203.8	205.1	207.5
Std. Dev.	142.5	147.08	147.93	106.23	86.77	75.36	68.18
Min	2.39	2.39	2.64	8.28	33.42	33.62	45.43
Max	1522.19	1522.19	1522.19	956.02	736.78	611.66	576.99
Clothing							
Mean	149.78	151.5	154.59	153.96	153.32	153.1	153.1
Median	142.53	144.88	147.98	151.36	152.02	151.52	151.82
Std. Dev.	76.03	76.04	78.96	56.47	47.25	41.87	38.12
Min	31.78	31.78	33.1	40.41	41.46	44.42	46.8
Max	2563.31	2563.31	2563.31	1349.77	977.84	747.98	636.32

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

the ratio of mean and the ratio of median. The ratio of median is not reported for the other two data fusions because the meaning of their imputed variables does not fit well since most of their values are concentrated in a single point of mass. Median is an indicator more robust for skewed distributions, but in this context mean is the more appropriate tool with which to evaluate the quality of the matching. A less accurate imputation can preserve the same central tendency between the two databases, but when imputed values are very different from those recorded in the donor data set, it is more difficult to preserve the average value since the mean is largely influenced by outliers.

The distributions of all categories of expenditure are very close to each other, showing that the matching procedure reproduced the same distribution as the original data set. For illustrative purposes, in Figures 3 and 4, we report only the distributions of the four main categories of expenditure and in Figure 5 we report the total household expenditure without disaggregations. This evidence is not sufficient to characterize the quality of the matching outcome completely. It is also necessary to inspect the marginal distribution of imputed variables by variables used to estimate the propensity score value and to compute the matching algorithm. Tables 7–8 and Tables A1–A3 in Supplemental material report the means and the medians of the extra information in the integrated and donor data sets and their ratio by the covariates used to estimate these values. These results show that the synthetic database well preserves the marginal empirical distribution of the common variables in the donor data set. Consequently, the original and matched groups are statistically similar. The lowest income category records the highest difference in mean and median between the two samples. Other discrepancies arise in the presence of children 0–5

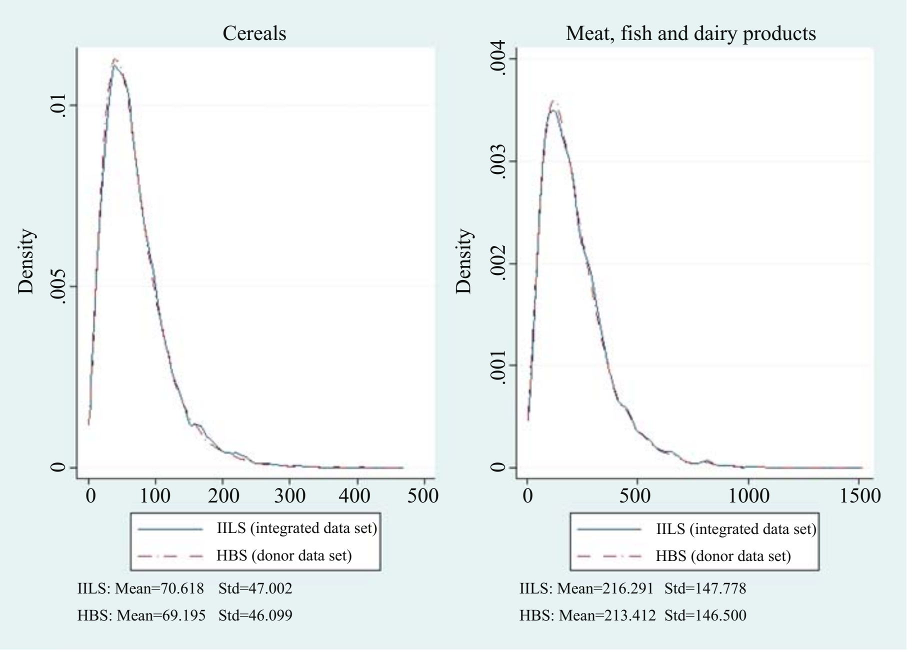


Fig. 3. Distribution of expenditure for "Cereals" and "Meat, fish and dairy products" in integrated and donor data sets.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

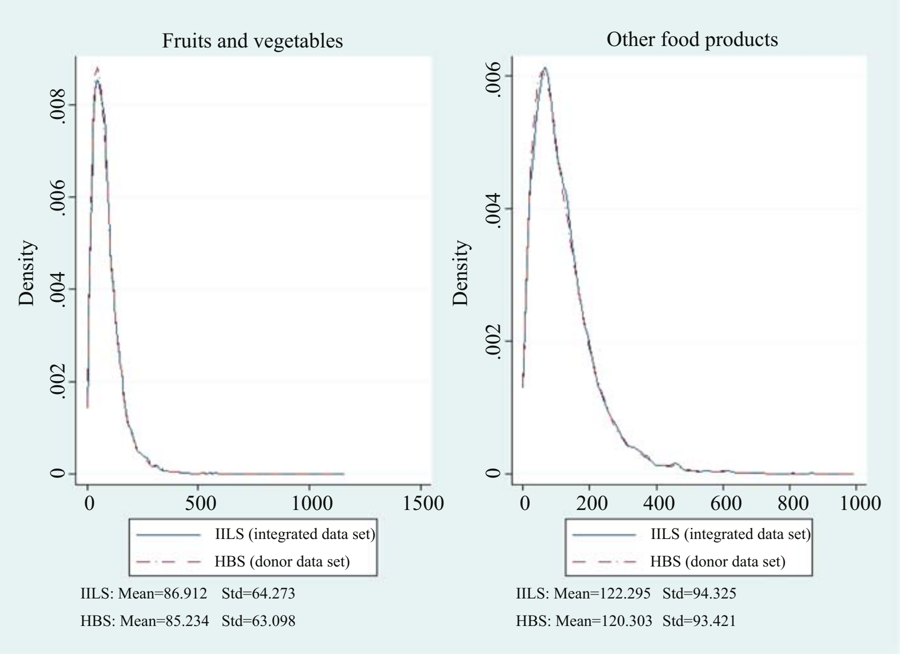


Fig. 4. Distribution of expenditure for "Fruits and vegetables" and "Other food products" in integrated and donor data sets.

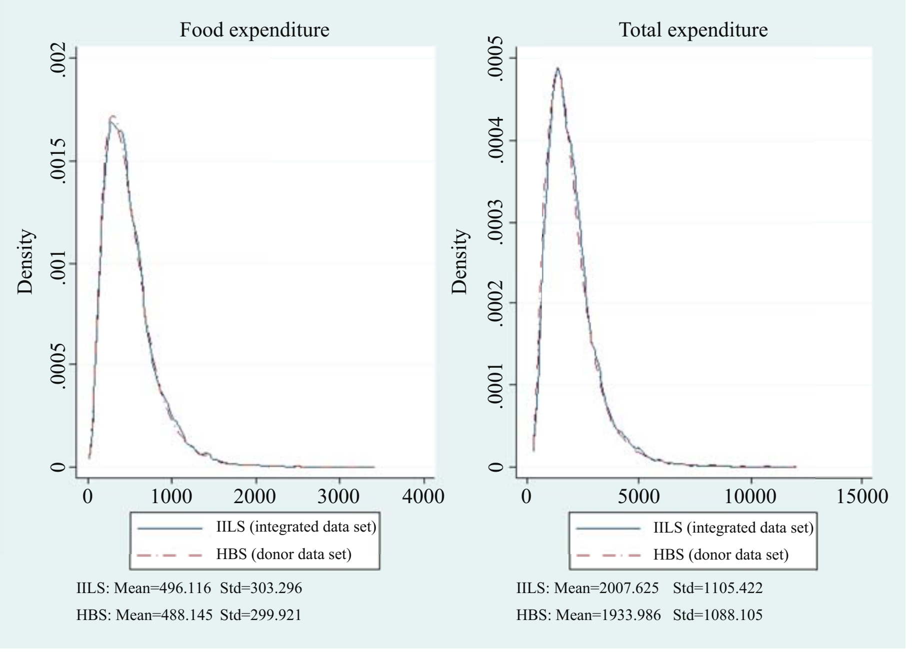


Fig. 5. Distribution of "Food expenditure" and "Total expenditure" in integrated and donor data sets.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 7. Cereals expenditure (in euros): Ratio of mean and median by covariates.

	Average			Median		
	HBS	IILS	Ratio	HBS	IILS	Ratio
Geographical area						
North-Western	71.23	71.05	99.74	60.84	60.79	99.92
North-Eastern	70.04	70.73	100.97	58.07	59.72	102.84
Center	69.30	72.67	104.87	59.95	62.35	104.00
Southern	67.80	68.38	100.84	58.97	58.83	99.76
Islands	66.40	69.24	104.27	56.96	59.36	104.21
Children 0–5 years old						
No	67.63	70.35	104.03	57.70	60.18	104.30
Yes	82.88	72.65	87.66	72.65	60.99	83.96
Children 6–14 years old						
No	65.10	70.47	108.24	55.32	60.11	108.66
Yes	89.44	71.40	79.84	78.93	60.79	77.02
Self-employed						
No	65.89	70.46	106.94	56.09	60.18	107.29
Yes	82.54	71.25	86.32	72.57	60.65	83.57
Single-parent						
No	69.41	70.63	101.75	59.36	60.14	101.31
Yes	66.44	70.51	106.13	57.97	62.59	107.97
Homeownership						
No	64.02	70.48	110.09	54.16	60.65	111.98
Yes	70.89	70.66	99.68	60.80	60.22	99.05
Average family education						
Primary	59.70	69.63	116.64	50.01	59.20	118.38
Middle	72.84	71.19	97.73	62.97	61.40	97.51
Middle-High	78.90	71.47	90.59	69.72	60.65	86.99
High	70.58	70.95	100.52	60.39	60.22	99.72
University	61.17	68.86	112.56	50.72	59.69	117.71
Household income						
1st quintile	51.05	58.84	115.26	42.89	48.60	113.30
2nd quintile	62.65	64.51	102.98	53.28	54.95	103.13
3rd quintile	68.90	69.78	101.28	60.49	61.85	102.25
4th quintile	77.12	76.29	98.93	68.29	68.18	99.83
5th quintile	87.48	81.60	93.28	76.71	69.64	90.78

or 6–14 years old and where a member of the family is self-employed. In this case, the divergence may be due to the number of children in each age group, rather than simply their presence.

We also investigated uncertainty generated by the lack of identifiability given the available data by calculating the Fréchet inequality for the contingency table associating income and expenditure classes. The Fréchet inequalities bound the probabilities of two joint events given the probabilities of the individual events conditioning on a set of common variables. In the present context where we use categorical variables, if we only know the conditional distributions $F(y|x)$ and $G(z|x)$ it is not possible to learn something about the association between y and z given x, but we can identify the bounds $max(0, F(y|x) + G(z|x) - 1) \leq H(y,z|x) \leq min(F(y|x), G(z|x))$ describing how uncertain the

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 8. Total household expenditure (in euros): Ratio of mean and median by covariates.

	Average			Median		
	HBS	IILS	Ratio	HBS	IILS	Ratio
Geographical area						
North-Western	2154.79	2021.39	93.81	1891.81	1783.93	94.30
North-Eastern	2147.08	2119.94	98.74	1914.22	1882.80	98.36
Center	1967.24	2197.14	111.69	1765.92	1989.06	112.64
Southern	1705.08	1766.91	103.63	1539.43	1562.48	101.50
Islands	1552.99	1748.30	112.58	1404.63	1552.45	110.52
Children 0–5 years old						
No	1884.86	1994.32	105.81	1655.32	1772.07	107.05
Yes	2364.10	2110.34	89.27	2133.66	1879.10	88.07
Children 6–14 years old						
No	1836.22	2015.54	109.77	1608.69	1797.63	111.74
Yes	2417.41	1966.00	81.33	2164.89	1729.95	79.91
Self-employed						
No	1806.41	2004.22	110.95	1591.03	1782.42	112.03
Yes	2449.27	2021.69	82.54	2174.32	1800.23	82.79
Single-parent						
No	1939.83	1993.21	102.75	1718.85	1769.21	102.93
Yes	1858.89	2160.95	116.25	1639.10	1954.15	119.22
Homeownership						
No	1817.00	2032.74	111.87	1638.38	1829.96	111.69
Yes	1972.39	1999.03	101.35	1738.17	1769.07	101.78
Average family education						
Primary	1340.58	1952.43	145.64	1155.36	1753.83	151.80
Middle	1929.09	1945.56	100.85	1719.70	1731.26	100.67
Middle-High	2307.72	2072.13	89.79	2083.60	1816.95	87.20
High	2260.24	2077.00	91.89	2012.14	1846.66	91.78
University	2293.16	2031.11	88.57	2062.34	1762.85	85.48
Household income						
1st quintile	1181.89	1481.45	125.35	1010.90	1235.49	122.22
2nd quintile	1593.65	1663.30	104.37	1436.04	1494.91	104.10
3rd quintile	1914.77	1931.64	100.88	1716.19	1730.43	100.83
4th quintile	2286.43	2247.94	98.32	2073.73	2015.75	97.20
5th quintile	2747.46	2581.01	93.94	2464.87	2332.84	94.64

association is between $yz|x$. When the intervals are statistically close, then the common variables of interest are suitable for matching.

In the present estimation of the Fréchet bounds, we consider the set of common variables used in the propensity score estimation. We first estimated these bounds, setting the intervals equal to income quintiles as used in our model specification. In order to analyze the influence of the width of the classes on the measure of uncertainty, we computed the same analysis also setting the intervals equal to income tertiles, eight fixed classes, as defined in Donatiello et al. (2014) that also match HBS and EUSILC, and income deciles. Consumption information was aggregated using the same classes defined for the income distribution. As reported in Table 9, the width of uncertainty is remarkably reduced from 20.3% to 5.9%, moving from tertiles to deciles. Donatiello et al. (2014) report an average width of the uncertainty bound

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 9. Average width of uncertainty bounds conditioning on common variables by different classes.

Classes	Average width of uncertainty bounds
Income tertile	0.203
Income quintile	0.125
Eight classes defined by Donatiello et al. (2014)*	0.069
Income decile	0.059

*Donatiello et al. (2014) defined the following classes: "Under EUR 1000", "EUR 1000–1500", "EUR 1500–2000", "EUR 2000–2600", "EUR 2600–3100", "EUR 3100–3600", "EUR 3600–5200" and "EUR 5200 or more".

equal to 7.8%, setting eight classes equal for income and consumption, which is comparable with our estimated range of 6.9% using the same intervals, though the comparison should be taken with caution because the number of conditioning variables is larger. If we take as a reference class definition the partition in deciles, we may consider an average width of 5.9% as a sound indication of a valid inference, though there still seems to be a good margin for improvement if, for example, auxiliary information was available. Inspection of Table 10 shows that, conditioning on the common variables, all cell probabilities for the eight selected classes are between the lower and upper bounds.

In the next Subsubsection, we study the economic robustness of the matching by investigating the Engel relationship linking the food share, an approximate indicator of well-being (Perali 2003, 2008), and the logarithm of total expenditure. This is a fundamental empirical relation that is stable independently of the society analyzed and the time period considered.

4.1.1. Economic Robustness of the Matched Data: The Engel Relationship and Material Well-Being

An immediate check of the economic robustness of the matched data is the comparison of income in the recipient EUSILC database and consumption from the HBS donor data set. Table 11 shows the number of households per income-expenditure and row frequencies of quintiles of household income and total expenditure grouped by the same classes of income quintiles. The marginal column of Table 11 shows that in the lowest quintiles, total expenditure exceeds income for almost 72% of the families, suggesting under-reporting of income (Meyer and Sullivan 2011). On the other hand, as is reasonable to expect, most families in the upper income quintiles have positive savings.

In Table 12, we focus on the relationship between total expenditure and specific expenditure items in the fused and donor data set. As shown in Table 12, all budget shares have a similar magnitude and pattern in both data sets. Food, clothing and housing shares decrease as total expenditure increases, as is typical for necessity goods. On the other hand, the budget share of transport and communication and recreation and education increase as total consumption increases.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 10. Uncertainty bounds for total household income and consumption.

Income classes	Consumption classes	Low.cx	CIA	Up.cx
1	1	0.00021	0.03565	0.10796
2	1	0.00010	0.03799	0.11970
3	1	0.00010	0.02918	0.10007
4	1	0.00000	0.02174	0.08065
5	1	0.00010	0.01085	0.05241
6	1	0.00005	0.00724	0.04055
7	1	0.00000	0.01094	0.04958
8	1	0.00000	0.00460	0.02827
1	2	0.00015	0.03701	0.12627
2	2	0.00008	0.04148	0.14724
3	2	0.00004	0.03939	0.14689
4	2	0.00000	0.03474	0.13315
5	2	0.00003	0.02051	0.08488
6	2	0.00005	0.01559	0.06708
7	2	0.00006	0.02669	0.10232
8	2	0.00001	0.01299	0.05510
1	3	0.00027	0.02766	0.10379
2	3	0.00016	0.03195	0.11913
3	3	0.00008	0.03486	0.14274
4	3	0.00008	0.03415	0.14402
5	3	0.00007	0.02252	0.09490
6	3	0.00000	0.01784	0.07493
7	3	0.00006	0.03193	0.12102
8	3	0.00012	0.01651	0.06482
1	4	0.00012	0.01876	0.07947
2	4	0.00016	0.02276	0.09596
3	4	0.00015	0.02689	0.11786
4	4	0.00005	0.02843	0.12807
5	4	0.00002	0.02015	0.09459
6	4	0.00000	0.01636	0.07610
7	4	0.00018	0.03097	0.12593
8	4	0.00016	0.01671	0.06658
1	5	0.00003	0.00760	0.04745
2	5	0.00010	0.00991	0.05767
3	5	0.00010	0.01228	0.06665
4	5	0.00000	0.01353	0.07412
5	5	0.00013	0.01039	0.07086
6	5	0.00005	0.00864	0.06341
7	5	0.00003	0.01595	0.07668
8	5	0.00018	0.00911	0.05257
1	6	0.00007	0.00408	0.03124
2	6	0.00006	0.00538	0.03740
3	6	0.00018	0.00699	0.04104
4	6	0.00005	0.00748	0.04426
5	6	0.00003	0.00584	0.04441
6	6	0.00005	0.00520	0.04258
7	6	0.00000	0.00980	0.04697
8	6	0.00002	0.00597	0.03840

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 10. Continued.

Income classes	Consumption classes	Low.cx	CIA	Up.cx
1	7	0.00001	0.00439	0.03289
2	7	0.00003	0.00599	0.03973
3	7	0.00010	0.00763	0.04482
4	7	0.00000	0.00858	0.04950
5	7	0.00005	0.00655	0.04814
6	7	0.00007	0.00614	0.04676
7	7	0.00017	0.01207	0.05559
8	7	0.00027	0.00789	0.04547
1	8	0.00000	0.00111	0.01156
2	8	0.00000	0.00154	0.01298
3	8	0.00000	0.00191	0.01388
4	8	0.00000	0.00223	0.01471
5	8	0.00002	0.00184	0.01446
6	8	0.00000	0.00175	0.01459
7	8	0.00006	0.00346	0.01559
8	8	0.00019	0.00251	0.01455

Notes: Classes are coded as: 1 = "Under EUR 1000"; 2 = "EUR 1000–1500"; 3 = "EUR 1500–2000"; 4 = "EUR 2000–2600"; 5 = "EUR 2600–3100"; 6 = "EUR 3100–3600"; 7 = "EUR 3600–5200"; 8 = "EUR 5200 or more".
Low.cx: The estimated lower bounds for the relative frequencies when conditioning on the common variables.
CIA: The estimated relative frequencies under the Conditional Independence Assumption (CIA).
Up.cx: The estimated upper bounds for the relative frequencies when conditioning on the common variables.

We further concentrate on the relationship between the food category and total expenditure because it is a robust relation whose main features should be maintained in the integrated database. Food expenditure and total expenditure have a similar distribution pattern in the original and fused data set both in the bottom and upper tail (Figure 5). This aggregate picture may hide significant differences, especially in the bottom and top five

Table 11. Conditional frequencies and percentages by income quintiles (in euros).

Quintiles of household income (Y)	Total expenditure (X)						Average savings (Y-X)
	<= 1199	1199⊣1788	1788⊣2545	2545⊣3662	>3662	Total	
<= 1199	1060	1026	912	501	237	3736	−1098.86
	28.37	27.46	24.41	13.41	6.34	100.00	
1199⊣1788	937	1068	895	558	272	3730	−450.12
	25.12	28.63	23.99	14.96	7.29	100.00	
1788⊣2545	754	1095	1098	616	299	3862	94.42
	19.52	28.35	28.43	15.95	7.74	100.00	
2545⊣3662	807	1058	1055	608	273	3801	1046.40
	21.23	27.83	27.76	16.00	7.18	100.00	
>3662	702	1092	1165	703	356	4018	3330.52
	17.47	27.18	28.99	17.50	8.86	100.00	
Total	4260	5339	5125	2986	1437	19147	
	22.25	27.88	26.77	15.60	7.51	100.00	

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 12. Average budget share by quintile group of total expenditure.

Expenditure category		Quintiles of total expenditure 1	2	3	4	5
Food	IILS	0.293	0.278	0.260	0.252	0.225
	HBS	0.301	0.280	0.272	0.257	0.226
Clothing	IILS	0.108	0.098	0.092	0.080	0.059
	HBS	0.107	0.098	0.092	0.079	0.058
Housing	IILS	0.327	0.293	0.274	0.255	0.212
	HBS	0.326	0.290	0.266	0.253	0.211
Transport and communication	IILS	0.103	0.151	0.166	0.178	0.185
	HBS	0.098	0.151	0.168	0.177	0.184
Recreation and education	IILS	0.058	0.083	0.107	0.134	0.198
	HBS	0.058	0.084	0.105	0.133	0.199
Health	IILS	0.110	0.096	0.101	0.102	0.121
	HBS	0.111	0.098	0.097	0.101	0.123

percent of the distribution. This is apparent when we compare the quantiles of the synthetic data set against the quantiles of the donor data set, as shown in the Q-Q Plot of Figure 6 referring to the whole sample. If the two groups belong to a population with the same distribution, the point should fall along the 45-degree reference line. Figure 6 shows a different pattern between the two samples for both food and total expenditure, only in the upper tail of the distribution. However, if we zoom in to the bottom and top five percent of

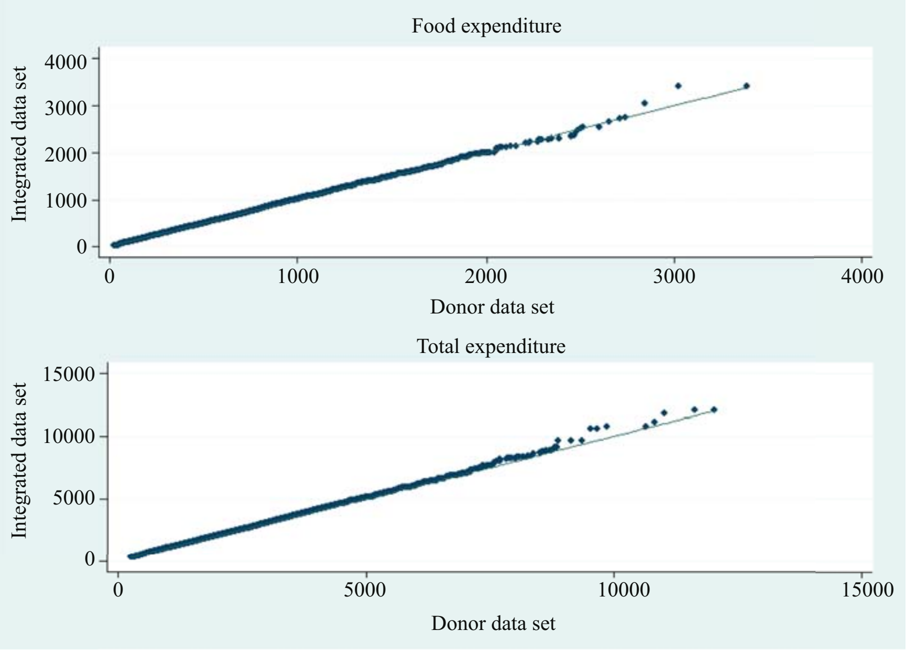


Fig. 6. Q-Q plot of food expenditure and total expenditure.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

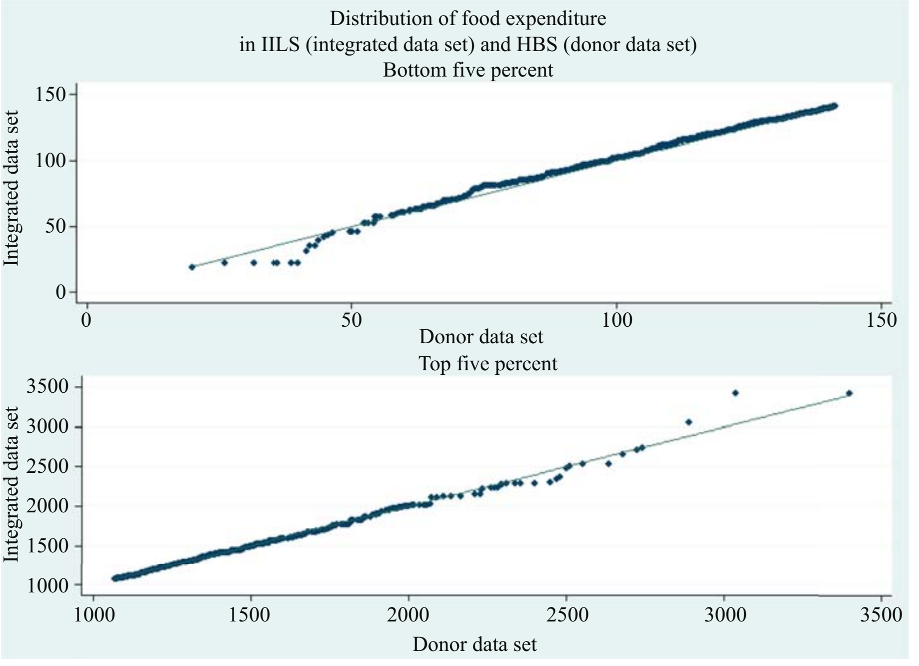


Fig. 7. Q-Q plot of food expenditure: focus on the tails.

the distribution, as shown in Figures 7 and 8, a similar departure in the lower tails can be seen, representing less than five percent of the sample.

In order to describe the shape of the food and total expenditure distributions at the tail, as shown in Tables 13 and 14, we test the statistical difference of the computed ratios of the 90th and 10th percentile describing the extent to which food or total consumption is larger at the top compared to the bottom of both the donor (HBS) and matched (IILS) population. As shown in Tables 13 and 14, we also summarize the dispersion of food and total expenditure with the Gini inequality index and test their difference. Table 14 also illustrates the Foster-Greer-Thorbecke (FGT) poverty measures and the associated statistics testing for the difference of the poverty measures in the donor and fused samples. The Foster-Greer-Thorbecke (Foster et al. 1984) indices are computed by substituting different values of the parameter α in the equation

$$\mathrm{FGT}_{\alpha} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{H}} \left(\frac{\mathrm{z} - \mathrm{y_i}}{\mathrm{z}} \right)^{\alpha},$$

where Z is the poverty threshold equal to 60% of the median of total expenditure respectively in the IILS integrated data (EUR 1082.944) and in the original HBS sample (EUR 1040.557), N is the sample size, H is the number of poor (those with total expenditure at or below z) and y_i is total expenditure of each individual i. With $\alpha = 0$, FGT_0 is the headcount ratio, the proportion of the population below the poverty line. With $\alpha = 1$ FGT_1 represents the poverty gap index, which summarizes the extent to which individuals fall below the poverty line. With $\alpha = 2$ FGT_2 measures the squared poverty gap ("poverty

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

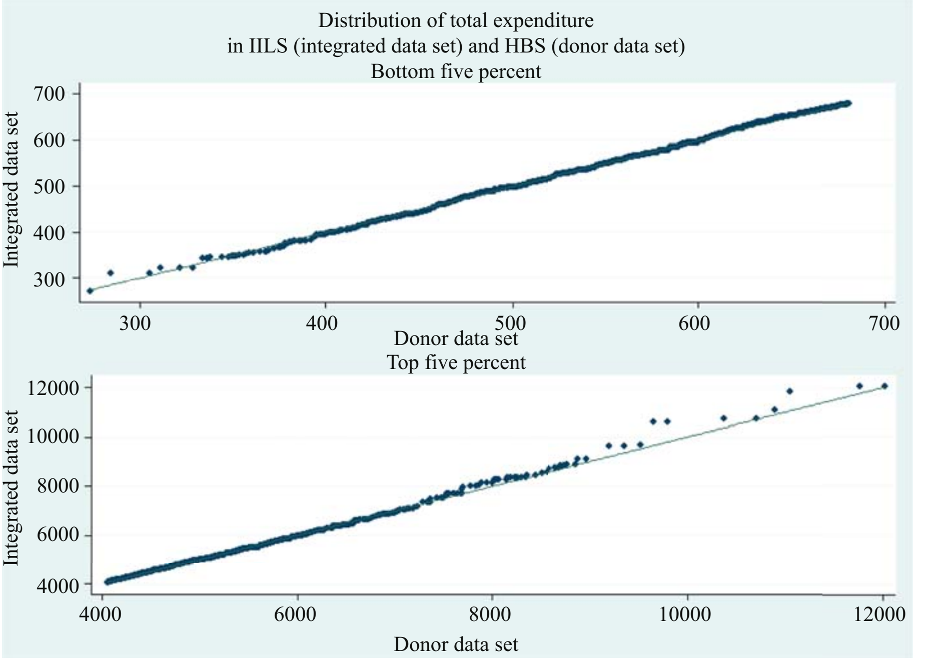


Fig. 8. Q-Q plot of total expenditure: focus on the tails.

severity") index, which places stronger emphasis on the poverty of the poorest individuals. With the exception of the percentile ratio for total expenditure, for all other comparisons we do not reject the null hypothesis that the estimates in the donor and matched data sets are the same at the .01 significance level. On the basis of this evidence, we conclude that the outcome of the matching is both statistically and economically robust.

To further verify the economic robustness of the matched distribution in a welfare measurement context, we estimated the Engel relationship linking the food share, a reliable proxy for well-being (Perali 2003, 2008), and the logarithm of total expenditure, as shown in Figure 9, which plots the inverse relationship between the food share and the logarithm of total expenditure. As the level of total expenditure increases, the food share, and the associated level of household well-being, decreases in a similar fashion in both the recipient and donor distribution.

To also investigate the shape of the conditional distribution of food expenditure with the logarithm of total expenditure in the lower and upper tails where there is higher statistical

Table 13. Dispersion indexes for food expenditure.

		p90/p10	Gini coefficient
IILS (integrated data set)		4.7902	0.3189
HBS (donor data set)		4.7309	0.3206
DIFFERENCE		−0.0592	0.0017
	std. err.	*0.0339*	*0.0022*
	p-value	*0.0802*	*0.4474*

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 14. Inequality and poverty indexes for total expenditure.

		p90/p10	Gini coefficient	FGT$_\alpha$ poverty index* $\alpha = 0$	$\alpha = 1$	$\alpha = 2$
IILS (integrated data set)		3.6310	0.2766	0.1568	0.0334	0.0101
HBS (donor data set)		3.7593	0.2816	0.1658	0.0354	0.0106
DIFFERENCE		0.1283	0.0050	0.0090	0.0020	0.0005
	std. err.	*0.0218*	*0.0021*	*0.0035*	*0.0012*	*0.0006*
	p-value	*0.0000*	*0.0175*	*0.0103*	*0.1004*	*0.3832*

*$\alpha = 0$: headcount ratio, $\alpha = 1$: poverty gap index, $\alpha = 2$: squared poverty gap index.

noise, we estimated the Engel relation by using also a quantile regression for each distribution quantile not influenced by extreme values. We estimated five quantile regressions for the quantiles 0.10, 0.25, 0.50, 0.75, and 0.90. Figure 10 shows the estimated quantile coefficients with the associated confidence intervals (solid line) and the least squares coefficients (dashed line) that, by construction, do not vary by quantile. OLS estimates underestimate, especially in the lower tails, both the matched IILS data set and the donor HBS dataset. The underestimation is larger in the integrated data set. Figure 11 shows the estimated quantile and OLS coefficients in the same graph. The distance between the estimated OLS coefficients in the integrated and donor data set and by quantile is not economically significant, although it is slightly larger in the lowest quintiles. The difference between quantile regression coefficients at the level of the second quintile is .005. This means that even if the estimated parameter is statistically significant,

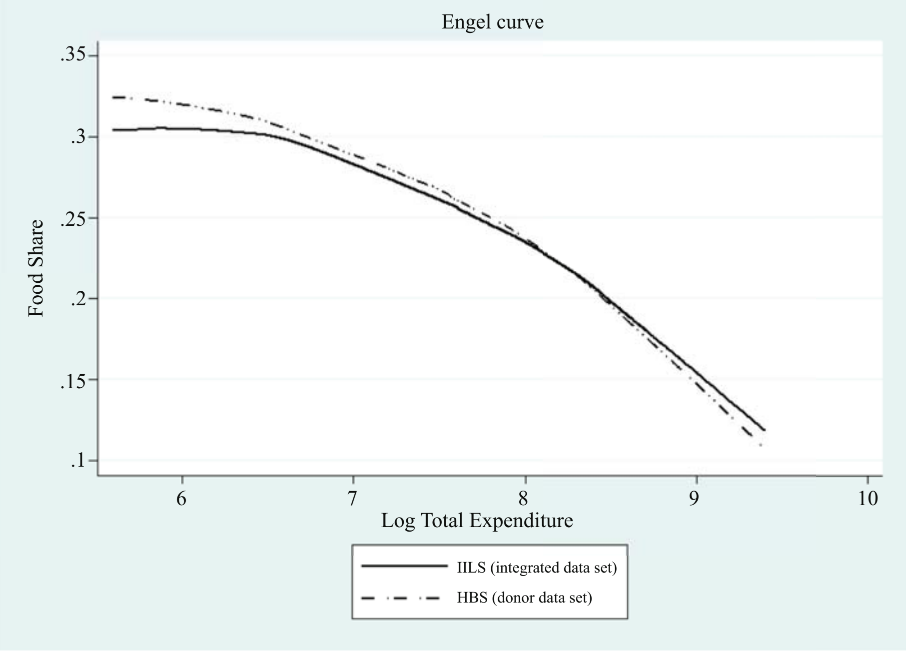
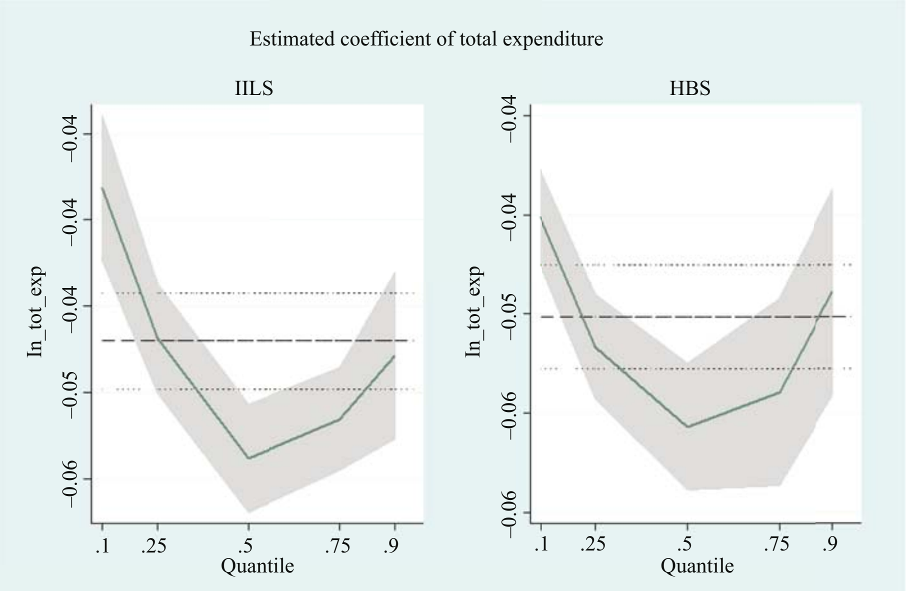


Fig. 9. Engel curve in integrated and donor data sets.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC



Note. Dashed line: least squares estimate
Solid line: Quantile regression estimate

Fig. 10. Estimated coefficient of total expenditure with OLS regression and quantile regression.

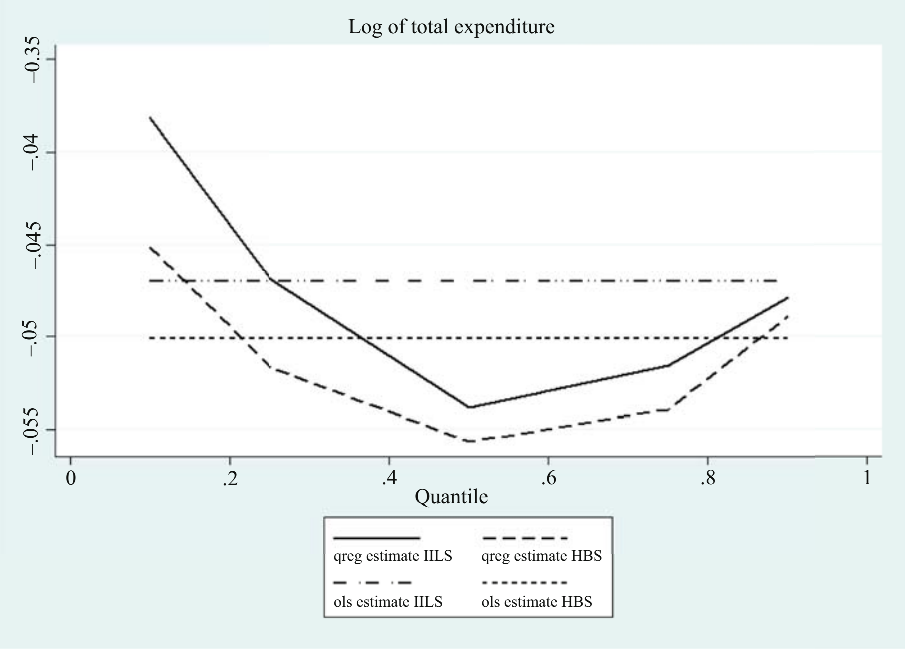


Fig. 11. Estimated coefficient of total expenditure with OLS regression and quantile regression.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

the variable's impact is economically negligible (Goldberger 1991). This evidence shows that in the fused data set the economic information is robustly maintained along all the relevant portions of the income distribution.

4.2. Data Fusion Between EUSILC and TUS

The EUSILC survey does not collect information about how Italians spend their time. These detailed data are traditionally gathered within specifically designed time use surveys. Both samples constructed by ISTAT are drawn from the same population with the same sampling design. They share a large set of common variables. Both fundamental conditions are satisfied, so that we could reliably perform the statistical matching technique.

We used the same covariates to match the activities on a weekday, Saturday, and Sunday. We obtained the same conclusion for the time spent on a weekday, on Saturday, and on Sunday. Consequently, as supplementary data we only report the results for the time spent on main activities during a weekday.

The common covariates used in the specification of the propensity score model included the region of residence (three dummies coded as North, Center, South), age (nine dummies for the age classes 3–5, 6–14, 15–19, 20–26, 27–36, 37–46, 47–56, 57–66, and older than 66), gender (1 if the individual is male), the presence in the family of a worker (1 if there was at least one working member), the presence of students (1 if there was at least one student in the family), the presence of children by age classes (0–5, 6–13 and 14–18 years of age), single-parent family (1 if there was a mother or father without partner, 0 otherwise) and the educational level attained (1 if the highest education level was high school or more, 0 otherwise). This set of common variables is the same as the common set used for the EUSILC and HBS, except the income variable that is not present in the TUS.

The distributions of these variables do not show any significant relationship between the two samples (Table A4). The largest absolute differences are recorded for geographical area but, as highlighted by Cramer's V test, these differences are not statistically significant. We also tested the equal distribution of the covariates before and after matching (Table A5). The largest standardized differences before matching are observed between the categories that refer to the geographical area. These differences disappear after matching. The p-values highlight the equality of means of covariates after matching. The covariates that reject the null hypothesis of equality of means before matching are the same covariates recording higher standardized difference.

The estimated propensity score shows a similar density distribution and its values show a common support in the recipient and donor databases. Observations have the same probability of belonging to one of the two samples (Figure A1) and we can be confident about obtaining unbiased results after implementing the matching algorithm to impute the missing values in the recipient database. To lend further support to this assertion, we investigated the matching quality for the variables in which we are most interested, such as rest, work, study and mobility (Tables A6–A9). These figures describe the differences in the original and matched database and in the ratio of the means by each covariate used in the propensity score specification. Almost all ratios are close to 100. This implies that the average in the two groups is similar. Marked deviations from 100 are explained by the

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

presence of some outliers in the donor data set that are not used to "impute" the missing values in the integrated database as shown by the heavy upper tails (Figures A2 and A3). This problem can be solved by computing the ratio of medians that gives statistical values not influenced by outliers. Note that it is not possible to use the ratio of medians because in most cases the median is equal to 0 and therefore the ratio cannot be calculated. In fact, the time spent on a particular activity does not depend only on one socio-demographic variable as represented in the tables, that is, work time should be compared jointly in relation to age and occupational status.

4.3. Data Fusion Between EUSILC and CISF Surveys

This matching involves the EUSILC survey, which does not present information about social capital, and the CISF survey, which collects detailed information on both bridging and bonding social capital and relational well-being (Menon et al. 2015). The set of common variables is the same as the common set used for the EUSILC and HBS, and EUSILC and TUS with the addition of the occupational status of women. Here, the income variable is not part of the set because it did not pass the balancing procedure.

To link these data sets, we implemented two different propensity score specifications because some variables about family relationships are pertinent only for some types of family. One propensity score specification concerned questions about family relations and the relationship with children. As a consequence, this specification related to a subsample of the EUSILC and CISF data set that does not include singles. We also excluded the families defined as "other types of family" because this typology is not defined in the same way in the two questionnaires and comparison is impossible with the available information. The other specification, on the other hand, analyzed the whole sample because the questions of interest are not related to family composition.

Statistical matching between these two questionnaires can be applied because the surveys refer to the same target population and share a set of common covariates with the same definition. Some variables are used in both specifications. We describe both because the sample size differs and this may affect the shape of the distribution.

4.3.1. Propensity Score Specification Excluding Singles and Other Family Types

In this propensity score specification, which excludes singles and other family types, we included the following variables: region of residence (five dummies coded as North-West, North-East, Center, South, Islands), age of the household head (three dummies coded as less than 35, 35–64, older than 64), dummies for the presence of children by age class (0–5, 6–13 and 14–18 years of age), main activity of the head of the household (four dummies coded as Employee, Unemployed, Retired, Inactive person), woman's occupational status (dummy equal to 1 if the household's wife/partner works), single-parent family (1 if there is a mother or father without partner, 0 otherwise) and education level attained by the household head (four dummies coded as Primary, Middle, High, University).

The distribution of these variables after their harmonization and aggregation is reported in the Table A10. Only the different levels of education have relatively higher values of absolute differences, although they are not statistically different in the two groups as measured by Cramer's V test.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

The specification used in the propensity score model achieved the balance in observed covariates (Table A11). Almost all values of standardized differences are reduced after matching, and the p-values show that the means of the recipient and the donor database are not statistically different. The propensity score distribution is similar in the same common support region, so we conclude that the observations have the same probability of being assigned to one of the two samples (Figure A4).

The quality of the matching outcome is high. The ratios of mean are close or very close to 100, revealing that the two databases have similar distributions of the extra information (Tables A10–A14, Figures A5 and A6).

4.3.2. Propensity Score Specification for the Whole Sample

This specification involved the whole sample because the extra information was not related to family type but concerns the attitude to participation in social life and social framework that pertains to singles and families as well. The specification also included variables regarding family composition, because the time spent on social events and voluntary activities also depends on family characteristics. We considered the region of residence (three dummies coded as North, Center, South), three dummies for the presence of children by age class (0–5, 6–13 and 14–18 years of age), two dummies describing man and woman's occupational status (1 if the man/woman was an employee), single-parent family (1 if there was only the mother or the father without partner) and level of education of the head of the household (four categories coded as Primary, Middle, High, University).

The frequency distribution of these variables shows a similar trend in the two samples (Table A15). The level of education of the head of the household displays the largest absolute differences between categories, but these differences are not statistically different, as pointed out by the result of Cramer's V test.

This specification proves that the observed variables are balanced between the recipient and the donor database. After matching, the standardized differences of all covariates are close to 0 and the p-values of the t-tests do not reject the null hypothesis of equality of means in the two samples (Table A16). The distribution of the propensity score value shows that the observations with the same characteristics have the same probability of extraction from both the synthetic and original data set (Figure A7). For simplicity's sake, we show the matching outcome for the variable "Take part in social activities or voluntary work", which is one of the variables of keenest interest in the present matching design because of its relevance to the measurement of well-being. The distribution is similar in the donor and integrated data set. Its ratios of mean are close to 100 (Figure A8 and Table A17).

5. An Example of an Empirical Application to the Measurement of Multidimensional Poverty

To better communicate at least some of the insights that can be obtained using the fused living standard data, we propose some salient results, also from a policy point of view, from an empirical exercise related to the multidimensional measurement of poverty. The monetary dimension of poverty is not sufficient to capture the multifaceted reality of poverty. A person with a relatively low standard of living may suffer from multiple

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

deprivations. A person in poverty may be jobless and houseless, a single parent, lacking good health, sufficient education or time to invest in the family. It could also be a person poor in the relation or social capital dimensions. Some of these dimensions are not strongly associated with income and can be highly informative about non-material dimensions of well-being. In our analysis, the monetary dimension can take the traditional form of disposable (after-tax) household income, may include the current income derived from the property's net worth (Brandolini et al. 2010), or may additionally include the evaluation of time invested in household production to form an extended notion of income.

In general, an individual receives income Y from labour, pensions, and other transfers and may hold a certain level of net worth or wealth W. Net worth, obtained as total income minus total liabilities, is thus an indicator of long-run economic security, while access to liquid assets is an indicator of the ability to cope with unanticipated emergencies. Current income CY is then defined as the sum of income Y and property income rW, where r is the average rate of return on assets, $CY = Y + rW$. Current income is an important determinant of the "economic situation" of an individual that depends on the flow of services over which it has command (Brandolini et al. 2010).

Adding to current income the value of time invested in household production gives a measure of extended income. The problem of estimating the value of the production of household services stems from the fact that the household product is not marketable. It is therefore difficult to know the value of the marginal product generated within the family enterprise. Therefore, the value of time devoted to paid market or unpaid domestic activities differ. Household production is a nonmarket activity whose value can be measured by its opportunity or market cost. A reasonable practice is to evaluate the time devoted to children at the market value, that is the wage at which families would pay the person that would substitute parents' care (Sharpe et al. 2011, Caiumi and Perali 2015; Poissonnier and Roy 2017).

Such a comprehensive picture of a deprivation profile can be described only using Integrated Living Standards data sets. In the present case, consumption information comes from the household budget survey, income and wealth from the standard of living survey, household time allocation from the time use survey and information on relational well-being from the social capital survey. A multidimensional measure of poverty counts the different forms of deprivation that a person experiences at the same time in different indicators of poverty that, in the present application, are equally weighted. By convention,

Table 15. Incidence of poverty (headcount ratio – H).

	North	Centre	South	Italy
Italian sample				
Equivalent total expenditure	0.1076	0.0893	0.2151	0.1356
Equivalente disposable income	0.0654	0.0852	0.1991	0.1100
Equivalent current income	0.0679	0.0871	0.2007	0.1121
Equivalent extended income	0.0388	0.0505	0.0870	0.0559
Subsample of Italian families with children				
Equivalent total expenditure	0.1983	0.1450	0.2957	0.2200
Equivalente disposable income	0.0709	0.0960	0.2194	0.1276
Equivalent current income	0.0773	0.1067	0.2416	0.1406
Equivalent extended income	0.0462	0.0710	0.0836	0.0647

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Table 16. Comparison of poverty measures by number of deprivations and first dimension (total expenditure, disposable/current/extended income).

	North		Centre		South		Italy	
First dimension	H	MPI	H	MPI	H	MPI	H	MPI
6 dimensions (cutoffs 3)								
Equivalent total expenditure	0.087	0.046	0.115	0.062	0.149	0.083	0.115	0.062
Equivalent disposable income	0.084	0.046	0.133	0.074	0.154	0.089	0.119	0.067
Equivalent current income	0.086	0.048	0.138	0.077	0.161	0.093	0.124	0.070
Equivalent extended income	0.082	0.045	0.127	0.071	0.121	0.071	0.105	0.060
10 dimensions (cutoffs 6)								
Equivalent total expenditure	0.041	0.022	0.075	0.039	0.116	0.063	0.074	0.040
Equivalent disposable income	0.044	0.024	0.084	0.045	0.118	0.065	0.079	0.043
Equivalent current income	0.047	0.025	0.086	0.046	0.123	0.068	0.082	0.045
Equivalent extended income	0.041	0.023	0.087	0.046	0.104	0.058	0.073	0.040

Notes:
1. Cutoffs: the minimum number of dimensions in which a person is deprived to be considered multidimensionally poor.
2. The dimensional cutoffs are: half the median for total expenditure or income (disposable/current/extended), net worth and women's time use; secondary school as level of parents education, single parenthood, at least one unemployed in the family and a satisfaction level of eight (in a range 0–10) for social capital dimensions.
3. Incomes are equivalized using Italian Equivalent Scales introduced January 1st, 2015. This scale accounts for family composition, parents' working condition and number of parents present in the family.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

a household is identified as multidimensionally poor if it is deprived in some combination of indicators whose weighted sum exceeds 30% of all deprivations (Alkire and Foster 2011). The traditional unidimensional approach to measure poverty is to calculate the proportion of the population who are poor, or headcount ratio H, on the basis of disposable income or total household expenditure. We also compute the index H considering the current and extended notion of income. We further calculate the multidimensional poverty index ($MPI = HA$), or adjusted headcount ratio, as the product of the incidence of poverty (H) and the average intensity of deprivation (A) reflecting the proportion of dimensions in which households are, on average, deprived.

Table 15 reports the incidence of poverty H for both the Italian sample and the subsample of Italian families with children also distinguishing the North, Center and South macroregions based on equivalent disposable, current, extended incomes and total household expenditure. Table 16 presents both the H and MPI measures for six and ten deprivation dimensions. These deprivation dimensions are: 1) equivalent household total expenditure or income (disposable/current/extended), 2) net worth, 3) parents education, 4) number of parents, 5) presence in the family of unemployed members, 6) women's time use for child care and household chores, 7) trust in family members, 8) trust on friends or acquaintances, 9) satisfaction of the relationship with children, 10) satisfaction about time spent together. The results are limited to the subsample of Italian families with children, because only in this context these relational variables are observable. Interestingly, the relative contribution of the dimensions "trust on friends" and "satisfaction about time spent together" are the two most important contributions of all deprivation dimensions. The striking result is that the poverty gap between the North and the South reduces increasingly as we integrate deprivation dimensions in terms of both H and MPI. This is a completely new map of poverty of great utility to policy-makers that we have been able to draw thanks to the construction of the Integrated Italian Living Standard data set.

6. Conclusions

This study has described a procedure used to construct a data set integrating Italian consumption, income, time use, and social capital surveys, adopting propensity score matching. The choice of fusing four data sets was motivated by the recommendations of the Fitoussi Commission (Stiglitz et al. 2010) and the interest of the Italian National Institute of Statistics in estimating well-being from an equitable and sustainable point of view. In general, integrated information is crucial for improving the quality of the estimation of household and individuals' well-being and of the comparisons of their standards of living.

Statistical matching can be seen as an imputation procedure for missing values from a donor data to a recipient data set. We used the propensity score value as a synthetic indicator of the common variables used in the specification model. This study gives detailed information on the matching variables and the statistical tests of the independence of the covariates playing special attention to the main data fusion between the EUSILC and HBS surveys, which we evaluated also exploring uncertainty.

We also compared the distributions of the extra information in the original and synthetic database. For the imputed information, we computed the ratio of mean and median between the two databases for the covariates used in the propensity score specification. We

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

also tested the economic robustness of the related data set by the Engel relationship, often used as a benchmark measure for welfare measurement. The matched data set passed all statistical and economic tests. To illustrate the value of the integrated information about standards of living we describe an example related to the multidimensional measurement of poverty. The noticeable result is that the poverty gap between the North and the South of Italy reduces increasingly as we integrate deprivation dimensions. This approach revealed a novel map of poverty of significant policy interest that we have been able to draw thanks to the construction of the Integrated Italian Living Standard data set.

The objective of this study is undeniably challenging because it deals with independent data sources not designed with integration purposes. Indeed, from a methodological point of view, we share the common hope that the international institutional effort to produce greater harmonization across HBS, EUSILC and TUS of both socio-demographic and other key economic variables will soon generate significant changes in their questionnaire design. As an example, a useful anchoring between HBS and EUSILC for matching purposes may occur if both surveys are record linked with administrative registers on income and wealth. Further, the *ex-ante* collection of auxiliary variables for integration purposes may involve both food consumed at home or away from home and clothing and footwear (not only in EUSILC, but also in HBS as aggregate recall questions), cumulated and short-term savings, housing value and expenses, transport, health conditions and, not last, stylized time use questions. This evolution would provide important auxiliary information and more meaningful logical constraints that can be effective in making the bias due to the conditional independence assumption negligible by reducing uncertainty.

An underexplored empirical issue that seems worth investigating in a systematic fashion is the comparison of the matching quality between propensity score matching and nonparametric matching methods placing especial emphasis on the selection procedure of the best set of matching variables and on the opportunities to deal with complex sample designs through weights' calibration procedures during the execution of the process. Another relevant empirical issue that may be more thoroughly analyzed is the measurement of the impact on the estimated standard errors derived from the fusion of multiple complex sample surveys.

Despite the lack of valuable auxiliary information, the results are satisfactory. Therefore, we can conclude that the integrated database to measure living standards in Italy can be reliably used to implement multidimensional inequality and poverty analysis explicitly assessing the value of time and social capital and, in general, to measure individual, household and social welfare more thoroughly.

7. References

Albayrak, Ö. and T. Masterson. 2017. *Quality of Statistical Match of Household Budget Survey and SILC for Turkey*. Levy Economics Institute (Working Paper no. 885). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2924849 (accessed October 2018).

Alkire, S. and J. Foster. 2011. "Counting and Multidimensional Poverty Measurement." *Journal of Public Economics* 95(7–8): 476–487. Doi: http://dx.doi.org/10.1016/j.jpubeco.2010.11.006.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Andridge, R.R. and R.J.A. Little. 2010. "A Review of Hot Deck Imputation for Survey Non-response." *International Statistical Review* 78(1): 40–64. Doi: http://dx.doi.org/10.1111/j.1751-5823.2010.00103.x.

Attanasio, O., E. Hurst, and L. Pistaferri. 2015. "The Evolution of Income, Consumption, and Leisure Inequality in the US, 1980–2010." In *Improving the Measurement of Consumer Expenditures*, edited by C.D. Carroll, T.F. Crossley, and J. Sabelhaus, 100–140. University of Chicago Press.

Augurzy, B. and C.M. Schmidt. 2001. *The Propensity Score: A Means to an End*, IZA Discussion Paper No. 271. Available at: https://ssrn.com/abstract=270919 (accessed September 2015).

Austin, P.C., N. Jembere, and M. Chiu. 2018. "Propensity Score Matching and Complex Surveys." *Statistical Methods in Medical Research* 27(4): 1240–1257. Doi: http://dx.doi.org/10.1177/0962280216658920.

Black, D.A. and J.A. Smith. 2004. "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching." *Journal of Econometrics* 121: 99–124. Doi: http://dx.doi.org/10.1016/j.jeconom.2003.10.006.

Blundell, R. and I. Preston. 1995. "Income, Expenditure and the Living Standards of UK Households." *Fiscal Studies* 16(3): 40–54. Doi: https://doi.org/10.1111/j.1475-5890.1995.tb00226.x.

Brandolini, A., S. Magri, and T. Smeeding. 2010. "Asset-based Measurement of Poverty." *Journal of Policy Analysis and Management* 29(2): 267–284. Doi: https://doi.org/10.1002/pam.20491.

Brewer, M. and C. O'Dea. 2012. *Measuring Living Standards with Income and Consumption: Evidence from the UK*. Institute for Social and Economic Research, University of Essex and Institute for Fiscal Studies (Working Paper n. 2012-05). Available at: https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2012-05.pdf (accessed April 2015).

Caiumi, A. and F. Perali. 2015. "Who Bears the Full Cost of Children? Evidence From a Collective Demand System." *Empirical Economics* 49: 33–64. Doi: http://dx.doi.org/10.1007/s00181-014-0854-2.

Caliendo, M. and S. Kopeinig. 2008. "Some Practical Guidance for the Implementation of Propensity Score Matching." *Journal of Economic Surveys* 22(1): 31–72. Doi: https://doi.org/10.1111/j.1467-6419.2007.00527.x.

Chen, J. and J. Shao. 2000. "Nearest Neighbor Imputation for Survey Data." *Journal of Official Statistics* 16(2): 113–131. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/nearest-neighbor-imputation-for-survey-data.pdf (accessed April 2015).

Coli, A., F. Tartamella, G. Sacco, I. Faiella, M. Scanu, M. D'Orazio, M. Di Zio, I. Siciliani, S. Colombini, and A. Masi. 2005. *La costruzione di un archivio di microdati sulle famiglie italiane ottenuto integrando l'indagine ISTAT sui consumi delle famiglie italiane e l'indagine Banca d'Italia sui bilanci delle famiglie italiane*. Technical Report, Working Group ISTAT- Bank of Italy, Rome. Available at: https://www.istat.it/it/files/2018/07/2006_12-1.pdf (accessed April 2015).

Conti, P.L., D. Marella, and M. Scanu. 2012. "Uncertainty Analysis in Statistical Matching." *Journal of Official Statistics* 28: 69–88.

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Conti, P.L., D. Marella, and M. Scanu. 2016. "Statistical Matching Analysis for Complex Survey Data with Applications." *Journal of the American Statistical Association* 111: 1715–1725. Doi: http://dx.doi.org/10.1080/01621459.2015.1112803.

Conti, P.L., D. Marella, and A. Neri. 2017. "Statistical Matching and Uncertainty Analysis in Combining Household Income and Expenditure Data." *Statistical Methods & Applications* 26(3): 485–505. Doi: http://dx.doi.org/10.1007/s10260-016-0374-7.

Dehejia, R.H. and S. Wahba. 1999. "Casual Effects in Nonexperimental Studies: Revaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94: 1053–1062. Doi: http://dx.doi.org/10.1080/01621459.1999.10473858.

Dehejia, R.H. and S. Wahba. 2002. "Propensity Score Matching Methods for Non-Experimental Causal Studies." *Review of Economics and Statistics* 84(1): 151–161. Doi: http://dx.doi.org/10.1162/003465302317331982.

D'Orazio, M., M. Di Zio, and M. Scanu. 2006a. *Statistical Matching: Theory and Practice*. New York: Wiley.

D'Orazio, M., M. Di Zio, and M. Scanu. 2006b. "Statistical Matching for Categorical Data: Displaying Uncertainty and Using Logical Constraints." *Journal of Official Statistics* 28: 137–157. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/statistical-matching-for-categorical-data-displaying-uncertainty-and-using-logical-constraints.pdf (accessed September 2017).

D'Orazio, M., M. Di Zio, and M. Scan. 2009. *Uncertainty Intervals for Non-Identifiable Parameters in Statistical Matching*. 57th Session of the International Statistical Institute, Durban (South Africa), August 2009. Available at: http://isi.cbs.nl/iamamember/CD8-Durban2009/A5%20Docs/0074.pdf (accessed April 2015).

D'Orazio, M., M. Di Zio, and M. Scanu. 2017. "The Use of Uncertainty to Choose Matching Variables." In *Statistical Matching*, edited by Ferraro et al. Soft Methods for Data Science. New York: Springer.

Donatiello, G., M. D'Orazio, D. Frattarola, A. Rizzi, M. Scanu, and M. Spaziani. 2014. "Statistical Matching of Income and Consumption Expenditures." *International Journal of Economic Science* Vol. III(3): 50–65. Available at: https://www.iises.net/download/Soubory/soubory-puvodni/pp50-65-ijoes-V3N3.pdf (accessed April 2015).

Foster, J., J. Greer, and E. Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52(3): 761–766. Doi: http://dx.doi.org/10.2307/1913475.

Garrido, M.M., A.S. Kelly, J. Paris, K. Roza, D.E. Meier, R.S. Morrison, and M.D. Aldridge. 2014. "Methods for Constructing and Assessing Propensity Score." *Health Services Research* 49(5): 1701–1720. Doi: http://dx.doi.org/10.1111/1475-6773.12182.

Goldberger, A. 1991. *A Course in Econometrics*. Cambridge, Massachusetts: Harvard University Press.

ISTAT. 2013. *Rapporto Bes 2013: il Benessere Equo e Sostenibile in Italia*. Rome: ISTAT. Available at: https://www.istat.it/it/archivio/84348 (accessed August 2018).

ISTAT. 2014. *Rapporto Bes 2014: il Benessere Equo e Sostenibile in Italia*. Rome: ISTAT. Available at: https://www.istat.it/it/archivio/126613 (accessed August 2018).

Kiesl, H. and S. Rässler. 2009. "How Useful Are Uncertainty Bounds? Some Recent Theory with an Application to Rubin's Causal Model." 57th Session of the International Statistical Institute, Durban (South Africa), 16–22 August 2009. Available at: http://isi.cbs.nl/iamamember/CD8-Durban2009/A5%20Docs/0169.pdf (accessed April 2015).

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Krug, E.G., L.L. Dahlberg, J.A. Mercy, A.B. Zwi, and R. Lozano. 2002. *World Report on Violence and Health*. Geneva: World Health Organization. Available at: https://www.who.int/violence_injury_prevention/violence/world_report/en/ (accessed April 2015).

Kum, H. and T.N. Masterson. 2010. "Statistical Matching Using Propensity Score: Theory and Application to the Analysis of the Distribution of Income and Wealth." *Journal of Economic and Social Measurement* 35: 177–196.

Lechner, M. 2008. "A Note on the Common Support Problem in Applied Evaluation Studies." *Annales d'Economie et de Statistique* 91/92: 217–235. Doi: http://dx.doi.org/10.2307/27917246.

Lee, W.S. 2013. "Propensity Score Matching and Variations on the Balancing Test." *Empirical Economics* 44(1): 47–80. Doi: http://dx.doi.org/10.1007/s00181-011-0481-0.

Leulescu, A. and M. Agafitei. 2013. *Statistical Matching: a Model Based Approach for Data Integration*. Luxembourg: Eurostat, European Commission. Available at: https://ec.europa.eu/eurostat/documents/3888793/5855821/KS-RA-13-020-EN.PDF/477dd541-92ee-4259-95d4-1c42fcf2ef34?version=1.0 (accessed April 2015).

Masterson, T. 2010. *Quality of Match for Statistical Matches Used in the 1999 and 2005 LIMEW Estimates for Canada*. Levy Economics Institute (Working Paper n. 615). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1670765 (accessed April 2015).

Masterson, T. 2014. *Quality of Statistical Match and Employment Simulations Used in the Estimation of the Levy Institute Measure of Time and Income Poverty (LIMTIP) for South Korea, 2009*. Levy Economics Institute (Working Paper n. 793). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2416850 (accessed April 2015).

Menon, M., R. Pendakur, and F. Perali. 2015. "All in the Family: How Do Social Capital and Material Wellbeing Affect Relational Wellbeing?" *Social Indicators Research* 124(3): 889–910. Doi: https://doi.org/10.1007/s11205-014-0816-2.

Meyer, B. and J. Sullivan. 2011. "Further Results on Measuring the Well-being of the Poor Using Income and Consumption." *Canadian Journal of Economics* 44(1): 52–87. Doi: http://dx.doi.org/10.1111/j.1540-5982.2010.01623.x.

Mittag, N. 2013. *Imputations: Benefits, Risks and a Method for Missing Data*. Harris School of Public Policy, University of Chicago. Available at: http://home.cerge-ei.cz/mittag/papers/Imputations.pdf (accessed April 2015).

Perali, F. 2003. *The Behavioral and Welfare Analysis of Consumption*. Dordrecht: Kluwer Academic Publishers.

Perali, F. 2008. "The Second Engel Law: Is it a Paradox?" *European Economic Review* 52(8): 1353–1377. Doi: http://dx.doi.org/10.1016/j.euroecorev.2008.01.005.

Poissonnier, A. and D. Roy. 2017. "Household Satellite Accounts for France. Methodological Issues on the Assessment of Domestic Production." *Review of Income and Wealth* 63(2): 353–368. Doi: http://dx.doi.org/10.1111/roiw.12216.

Rässler, S. 2002. *Statistical matching: A Frequentist Theory, Practical Applications, and Alternative Bayesian Approaches*. New York: Lecture Notes in Statistics, 168, Springer.

Rässler, S. 2004. "Data Fusion: Identification Problems, Validity, and Multiple Imputation." *Austrian Journal of Statistics* 33: 153–171. Available at: https://www.researchgate.net/publication/228528513_Data_fusion_Identification_problems_validity_and_multiple_imputation (accessed April 2015).

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Renssen, R.H. 1998. "Use of Statistical Matching Techniques in Calibration Estimation." *Survey Methodology* 24: 171–183. Available at: https://www150.statcan.gc.ca/n1/en/catalogue/12-001-X19980024354 (accessed August 2018).

Ridgeway, G., S.A. Kovalchik, B.A. Griffin, and M.U. Kabeto. 2015. "Propensity Score Analysis with Survey Weighted Data." *Journal of Causal Inference* 3(2): 237–249. Doi: http://dx.doi.org/10.1515/jci-2014-0039.

Rios-Avila, F. 2014. *Quality of Match for Statistical Matches Using the American Time Use Survey 2010, the Survey of Consumer Finances 2010, and the Annual Social and Economic Supplement 2011*. Levy Economics Institute (Working Paper no. 798). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2432153 (accessed April 2015).

Rios-Avila, F. 2015. *Quality of Match for Statistical Matches Using the Consumer Expenditure Survey 2011 and Annual Social Economic Supplement 2011*. Levy Economics Institute (Working Paper n. 830). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2554089 (accessed October 2015).

Rios-Avila, F. 2016. *Quality of Match for Statistical Matches Used in the Development of the Levy Institute Measure of Time and Consumption Poverty (LIMTCP) for Ghana and Tanzania*. Levy Economics Institute (Working Paper n. 873). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2838475 (accessed September 2017).

Rodgers, W.L. 1984. "An Evaluation of Statistical Matching." *Journal of Business & Economic Statistics* 2: 91–102. Doi: http://dx.doi.org/10.2307/1391358.

Rosenbaum, P.R. and D.B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Casual Effects." *Biometrika* 70: 41–55. Doi: http://dx.doi.org/10.1093/biomet/70.1.41.

Rosenbaum, P.R. and D.B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician* 39: 33–38. Doi: http://dx.doi.org/10.2307/2683903.

Rubin, D.B. 1986. "Statistical Matching Using File Concatenation with Adjusted Weights and Multiple Imputations." *Journal of Business, Economics and Statistics* 4: 87–94. Doi: http://dx.doi.org/10.2307/1391390.

Rubin, D.B. and N. Schenker. 1986. "Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse." *Journal of the American Statistical Association* 81: 366–374. Doi: http://dx.doi.org/10.1080/01621459.1986.10478280.

Sharpe, A., A. Murray, B. Evans, and E. Hazell. 2011. *The Levy Institute Measure of Economic Well-Being: Estimates for Canada, 1999 and 2005*. Levy Economics Institute (Working Paper n. 680). Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1895637&rec=1&srcabs=1670765&alg=7&pos=5 (accessed August 2018).

Singh, A.C., H. Mantel, M. Kinack, and G. Rowe. 1990. *On Methods of Statistical Matching with and Without Auxiliary Information*. Methodology Branch, Statistics Canada (Technical Report SSMD-90-016E). Available at: http://publications.gc.ca/collections/collection_2017/statcan/11-613/CS11-613-90-16-eng.pdf (accessed April 2015).

Singh, A.C., H. Mantel, M. Kinack, and G. Rowe. 1993. "Statistical Matching: Use of Auxiliary Information as an Alternative to the Conditional Independence Assumption."

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC

Survey Methodology 19: 59–79. Available at: https://www150.statcan.gc.ca/n1/pub/12-001-x/1993001/article/14475-eng.pdf (accessed April 2015).

Sisto, A. 2006. *Propensity Score Matching: un'Applicazione per la Creazione di un Database Integrato ISTAT-Banca d'Italia*. Dipartimento di Politiche Pubbliche e Scelte Collettive – POLIS, Università del Piemonte Orientale (Working Paper n. 63). Available at: http://polis.unipmn.it/pubbl/RePEc/uca/ucapdv/sisto63.pdf (accessed April 2015).

Slesnick, D. 1993. "Gaining Ground: Poverty in the Postwar United States." *Journal of Political Economy* 101: 1–38. Doi: http://dx.doi.org/10.1086/261864.

Smith, J. and P. Todd. 2005. "Rejoinder." *Journal of Econometrics* 125: 365–375. Doi: http://dx.doi.org/10.1016/j.jeconom.2004.04.013.

Stiglitz, E., A. Sen, and J.P. Fitoussi. 2010. *Report by the Commission on the Measurement of Economic Performance and Social Progress*, Paris. Available at: http://www.stiglitz-sen-fitoussi.fr/en/index.htm (Accessed August 2018).

Tedeschi, S. and E. Pisano. 2013. *Data Fusion Between Bank of Italy-SHIW and ISTAT-HBS*. Munich Personal RePEc Archive (Working Paper n. 51253). Available at: https://mpra.ub.uni-muenchen.de/51253/ (accessed April 2015).

Webber, D. and R. Tonkin. 2013. *Statistical Matching of EU-SILC and the Household Budget Survey to Compare Poverty Estimates using Income, Expenditures and Material Deprivation*. EUROSTAT, Methodologies & Working papers, European Union. Available at: https://ec.europa.eu/eurostat/documents/3888793/5857145/KS-RA-13-007-EN.PDF/37d4ffcc-e9fc-42bc-8d4f-fc89c65ff6b1 (accessed April 2015).

Wolff, E.N. and A. Zacharias. 2003. *The Levy Institute Measure of Economic Well-Being*. The Levy Economics Institute (Working Paper n. 372). Available at: http://www.levyinstitute.org/pubs/wp/372.pdf (accessed April 2015).

Wolff, E.N., A. Zacharias, T. Masterson, S. Eren, A. Sharpe, and E. Hazell. 2012. *A Comparison of Inequality and Living Standards in Canada and the United States Using an Expanded Measure of Economic Well-Being"*. Levy Economics Institute (Working Paper n. 703). Available at: http://www.levyinstitute.org/pubs/wp_703.pdf (accessed April 2015).

Received October 2016
Revised June 2018
Accepted September 2018

Bereitgestellt von Universita degli Studi di Verona (Biblioteca Frinzi) | Heruntergeladen 11.09.19 07:05 UTC