# A New Strategy for Reducing Selection Bias in Non-Experimental Evaluations, and The Case of How Public Assistance Receipt Affects Charitable Giving

*by*

Laura R. Peck, Associate Professor
School of Public Affairs, Arizona State University
laura.peck@asu.edu

Ida D'Attoma, Research Fellow
Department of Statistical Sciences, University of Bologna, Italy
ida.dattoma2@unibo.it

Furio Camillo, Associate Professor of Business Statistics and Data Mining
Department of Statistical Sciences, University of Bologna, Italy
furio.camillo@unibo.it

Chao Guo, Associate Professor
Department of Public Administration and Policy
University of Georgia

April 14, 2011

**Acknowledgements**
This paper was prepared for the Fall Research Conferences of the Association for Public Policy Analysis and Management and the American Evaluation Association. Thanks go to the agencies that fund the Panel Study of Income Dynamics (PSID), Atlantic Philanthropies for funding the collection of data in the first four waves of the Center on Philanthropy Panel Study (COPPS), and the Bill and Melinda Gates Foundation for funding the 2007 and 2009 data collection of the Center Panel as well as the dissemination of the 2005 data. We appreciate the research assistance of Andrea Mayo. We also acknowledge participants in the ASU School of Public Affairs Research Colloquium, and participants in the ASU Center for Population Dynamics Colloquium for their useful input.

**Keywords**
public assistance, welfare, charity, selection bias, cluster analysis

**A New Strategy for Reducing Selection Bias in Non-Experimental Evaluations, and The Case of How Public Assistance Receipt Affects Charitable Giving**


## Abstract

Prior work has analyzed the extent to which welfare recipients engage in giving money and time to charitable causes (Brooks, 2002, 2004; Author4 & Author1, 2009; Author1 & Author4, 2011), finding that public assistance is negatively associated with donating money with the relationship to volunteering being unclear. Nevertheless, sticky issues of selection bias compel more deliberate thinking about the strength of assertions about cause and effect. In response, we now conduct a multivariate cluster-based subgroup analysis approach to more confidently infer causality about the ways in which welfare receipt affects charitable activity. This approach to dealing with the problem of selection bias capitalizes on the known treatment-associated variance in the X matrix, transforming data to estimate unbiased treatment effects. We contribute to both the substantive and methodological literatures with this work.

**A New Strategy for Reducing Selection Bias in Non-Experimental Evaluations, and The Case of How Public Assistance Receipt Affects Charitable Giving**

Much scholarship is dedicated to understanding the factors that compel one to donate time or money to charitable causes; but only a small portion of that scholarship considers those within society are least able to make those contributions. Public opinion data shows that those who receive public assistance or are poor share the same values as the rest of society; and so on the face of it, there is no reason to believe they might behave differently in terms of their *interest* in volunteering their time or donating their money. Nevertheless, income is a major constraint on one's ability to donate money or time to charity, and it is income that public assistance recipients and other poor segments of our population lack. Some research on charitable activity among public assistance recipients suggests that welfare receipt tends to reduce the levels of charitable giving (e.g., Brooks, 2002; 2004); but more recent work shows that this reduction might be offset by relatively higher levels of volunteering, all else equal (Author4 & Author1, 2009). One of the main challenges in this research is selection bias. That is, those who use welfare are different – in ways that affect their outcomes of interest – from those otherwise, seemingly comparable individuals who do not use welfare, and this selection mechanism is not well understood or modeled with the kind of data commonly available.

Building on prior work, this paper adds to the small literature on how public assistance might influence recipients' charitable activity. We explore the issues of public assistance recipients' donations of their time and (quite limited) money for charitable purposes. We use the 1994-2005 waves of the Panel Study of Income Dynamics (PSID), along with 2005 wave of the Center on Philanthropy's Panel Study (COPPS), a PSID module that includes data on charitable giving and volunteering. Importantly, we also use this question to test a new analytic approach to overcoming selection bias and hope this has useful methodological lessons for program evaluators and policy analysts.

This paper is organized into four sections. We begin with background discussion both of the effects of earned and public assistance income on people's charitable behaviors. We then discuss the data and methods, including the issues of selection bias problem that challenge this research and proposing a new analytic method that minimizes selection bias in impacts estimates. The third section presents results of our data analysis and discusses the findings. We conclude with a discussion of the contributions and limitations of the study, as well as research and policy implications.

## Background

Prior research examines a variety of factors associated with individuals' donations of money and time. Here we discuss findings from that research, in particular what we know about the relationships between public assistance use, as well as demographic traits, and charitable activity.

### *Charitable Activity and its Correlates*

Several key personal characteristics are associated with charitable activity. As we have discussed in contextualizing prior work (Author4 & Author1, 2009; Author1 & Author4, 2011), research documents that age may be "the variable most consistently related to giving" (Clotfelter, 1997, 17), with donations increasing with one's age (e.g., Arai, 2004; Brooks, 2002; Van Slyke & Brooks, 2005; Wilson and Musik, 1997). Age may be related to volunteering in a curvilinear fashion due to physical infirmity among the elderly. Campbell and Yonish (2003) note that “while there is a generally monotonic increase in volunteering as someone ages, volunteering falls off sharply for the elderly as they become less physically able to perform volunteer work” (n14, p.233). Education is also consistently and positively correlated with charitable giving (e.g., Feldstein & Clotfelter, 1976; Jencks, 1987; Morgan, Dye & Hybels, 1977; Van Slyke & Brooks, 2005; Steinberg & Wilhelm, 2003) and volunteering (Choi, 2003; McPherson & Rotolo, 1996). Marital status, in contrast, has a less clear association with charitable giving, but research shows that it matters. Some have found a positive relationship (e.g., Andreoni, Brown & Rischall, 2003; Choi, 2003; Jencks, 1987; Rooney et al., 2005; Van Slyke & Brooks, 2005; Wilson, 2000), explained by the added resources and social connections that marriage brings; but others have found no difference in the volunteering that married and unmarried people do, controlling for other factors (Fischer et al., 1991; Herzog & Morgan, 1993; Musick & Wilson, 2008). Another clear link exists between religious affiliation and practice and charitable activity (Wilson & Musick, 1997), with being Christian (e.g., Van Slyke & Brooks, 2005) and more frequent service attendance (e.g., Steinberg & Wilhelm, 2003) being associated with greater giving of both time and money. Health status also influences volunteer activity, with poor health reducing both one’s ability to volunteer (Caputo 1997) and the amount of time one volunteers (Gallagher, 1994; Wilson & Musick, 1997), particularly among older adults (Choi, 2003). Some research finds that race and ethnicity affect charitable activity, with African American and Hispanic families being less likely to donate or volunteer, or to donate less money or fewer hours, than Whites (Hodgkinson & Weitzman, 1996); but others find no effect (e.g., O’Neill & Roberts, 2000; Rooney et al., 2005; Steinberg & Wilhelm, 2005).

In addition to these personal traits, two other correlates of charitable activity demand attention: income, and public policy. Substantial research has considered the impact of public policy on charitable giving (Brooks, 2000; Steinberg, 1990; Steinberg, 1993). Relevant policies include personal tax deductions for charitable donations and the tax exempt status of various investments. The policy most relevant to this analysis for instance is tax deduction itemizing. Another arena of policy is public assistance, which this project considers, and its role in inhibiting or encouraging charitable activity. This arena is much less studied, with our recent work (Author4 & Author1, 2009; Author1 & Author4, 2011) and that of Brooks (2004, 2002) being the most direct tests of policy effects.

As we describe elsewhere (Author4 & Author1, 2009; Author1 & Author4, 2011), Brooks (2002, 2004) finds that welfare receipt negatively impacts charity, suggesting that welfare income functions differently from earned income in stimulating charitable activity. Notwithstanding their contribution to our understanding of the charitable activities of welfare recipients, existing studies neglect one key dimension: volunteering. This omission is important because, as Brooks (2002) notes, contributions by the poor "might be especially likely to take a non-pecuniary form" (111). In light of the negative effect of welfare receipt on charitable giving documented in the literature, the question arises: would public assistance use have a similar, negative influence on recipients' likelihood to volunteer? Such a negative effect would be consistent with the notion of a "culture of welfare," which suggests that long-term welfare dependency may suppress many pro-social attitudes (Brooks, 2002). It also echoes the proposition that "welfare benefits discourage political involvement by cultivating personal traits of dependence" (Soss, 1999, 363).

Alternatively, receiving public assistance might have a positive effect on recipients' likelihood to volunteer. That is, government assistance might encourage beneficiaries to become more civically engaged. Mettler (2002), for example, finds that the G.I. Bill's educational provisions lead to increased levels of voluntary participation by veterans (in the form of memberships in civic organizations and political activity). A positive association between public assistance receipt and volunteering may also be interpreted from the perspective of human capital and career development (e.g., Menchik & Weisbrod, 1987). Combined with job search (or other traditional welfare-to-work activities), volunteering can help public assistance recipients gain economic independence by building their confidence and the motivation necessary to succeed in regular employment and by providing opportunities for learning specific work-related skills (Herr, Wagner & Halpern, 1996). In addition, public assistance recipients can also benefit from the signaling effect of volunteering: their volunteer experience can signal to future employers the presence of certain skills or abilities (Schiff, 1990; Ziemek, 2006).

The effect of public assistance receipt on charitable activity can also be understood by considering whether giving and volunteering are complements or substitutes. On one hand, the Independent Sector's research on giving and volunteering reveals a positive correlation between these two types of individual philanthropy (Hodgkinson & Weitzman, 1996). Van Slyke and Brooks (2005) also find that individuals who volunteer for charitable organizations, regardless of their demographic attributes, are more likely to donate money as well. They reason that volunteers donate money out of a sense of community. Following this line of reasoning, then, government support for those who are in need might discourage their volunteer activities if assistance negatively affects their giving behavior. Although no existing evidence supports this possibility, research finds that giving and volunteering both increase when tax policy encouraging giving is implemented, and that they both decrease when tax policy discouraging giving is in place (e.g., Andreoni et al., 1996; Brown, 1999).

On the other hand, a standard economic model of constrained utility maximization would predict that giving and volunteering function as substitutes for each other (Slyke & Brooks, 2005). This is consistent with the work of Jencks (1987) and Duncan (1999), who find that when an individual's likelihood to give money declines, his or her contribution of time increases, on average. Following this line of reasoning, we might hypothesize that public assistance receipt would increase volunteerism. Although the relationship between public assistance receipt and giving is negative, the relationship between public assistance receipt and volunteering could well be positive: relying on public assistance reflects a severe income constraint that may rule out charitable giving, but, if giving and volunteering are substitutes, then volunteering would be more likely. Stated differently, it is possible that a public assistance recipient's diminished financial contributions might be off-set by an increase in the amount of time donated. Prior work suggests that this may be the case (Author4 & Author1, 2009).

Beyond the discussion of giving and volunteering as complements or substitutes, one might consider the effect of public assistance use on charitable activity in terms of the distinction between luxury and necessity goods. Luxury goods are goods whose income elasticity of demand exceeds one, whereas necessity goods are those whose income elasticity of demand is below one. As income rises, the proportion of expenditure on luxury goods increases while that on necessity goods decreases (Deaton & Muellbauer, 1980). It is plausible that gifts of money are luxury goods among people with low incomes. Gifts of time, on the other hand, are necessity goods. For low-income people such as welfare recipients, therefore, the "giving mix" is mostly in the form of time, but the mixture shifts towards money as income increases.

In light of this discussion, we propose to examine two competing hypotheses regarding the effect of public assistance receipt on charitable giving and volunteering. One hypothesis is that public assistance receipt has a negative effect on charitable giving, but has a positive effect on volunteering.  The competing hypothesis is that public assistance receipt has a negative effect on both charitable giving and volunteering.

*Research Question*.  The question that this motivates is: How does public assistance receipt affect charitable giving?  As noted, prior work examined two dimensions of charitable giving:  donations of time and money.  We revisit that same question here by extending a new methodological approach to increase our confidence in the *causal* story.

**Methodology**

This section details our data source, measures and the analytic approach we follow to answer the question.

*Data and Measures*

Our analysis uses the 2005 Center on Philanthropy Panel Study supplement to the PSID. In particular, we merge variables from the 1994-2005 panels of the PSID to the 2005 COPPS supplement.  The COPPS collects data on the amount of money and number of hours donated to several charitable purposes:  religious, combined funds, basic needs, health, education, youth and family services, the arts, neighborhoods, the environment, and international aid.  In addition to these key indicators, the dataset contains information about households characteristics and use of various forms of public assistance.  Our sample in particular is the 7,822 households that comprise the 2005 COPPS.

**Measures**

As elaborated in Author4 and Author1 (2009), from the raw COPPS data, we compute two dependent variables: the *Amount Donated* and the *Hours Volunteered*. The amount donated is simply the sum of households' contributions made in 2004 to a variety of charitable sources and is measured in dollars.[1] Likewise, the number of hours volunteered is the sum of the head of households' time donated to a variety of charitable organizations in 2002.[2] Our descriptive statistics include binary versions of these variables as well as the percent of the population that donates money or time at all.

Our specific measures of income combine specific sources of 2004 household income as reported by survey respondents. *Earned Income* is the sum of the household head and spouse's labor income (e.g., wage, salaries, etc.). *Public Assistance Income* is the sum of the household head and spouse income from cash public assistance (Temporary Assistance to Needy Families [TANF]), general assistance (GA), and food stamps. Both income variables are converted to reflect annual amounts. Although the scant prior research on public assistance use and charitable activity has focused on cash assistance in particular, we feel justified in using a broader conception of public assistance. There is substantial overlap between cash assistance and food stamps use. While some individuals receive only food stamps and not cash assistance, they may still be considered as somewhat reliant on the state for support. Including food stamp recipients also increases the number of members in the treatment subsample, thereby increasing our ability to detect effects, if they exist.

[1] Here we use the COPPS definition of charitable giving as donations of $25 or more made to charitable organizations. Since there is not a "total amount donated" variable in the COPPS data, we construct this variable by adding religious giving to all secular giving. For the few households who answered the unfolding brackets questions (e.g., "Did you make any donations specifically for religious purposes", etc.) rather than the amount given question, this variable is coded to the lower bound based on the brackets.

[2] Since there is not a "total hours volunteered" variable in the COPPS data, we construct this variable by adding hours volunteered to religious organizations to hours volunteered to secular organizations.

Next, we compute three additional public assistance variables. *Prior Public Assistance Receipt* identifies the number of years between 1993 and 2002 (with a total of seven possible) that a household had income from cash public assistance (TANF and GA) or food stamps. Similarly, we report the number of years of *Prior Public Housing Residence* between 1993 and 2002; and we report *Current Public Housing Residence* as whether the household reports living in public housing in 2004. We include these public housing variables as a way to control for other measures of disadvantage that might be associated with both public assistance use and the ability to donate one's money or time to charitable causes.

In order to capture the possible effect of the price of giving (i.e., marginal income tax rates), we include a binary variable, *Tax Itemizer*, which equals 1 if the household head itemized deductions in his or her tax return (and 0 otherwise). While the price of giving one dollar to charity is one dollar for taxpayers who do not itemize deductions in their tax returns, for itemizers the price of giving is less than a dollar, as they receive a "rebate" equal in value to their deductible contributions times the applicable marginal tax rate.

The general household characteristics are straightforward and capture the traits that prior research identifies as associated with charitable activity: age, sex, number of children, marital status, race, ethnicity, education, religiosity, urban residence, and health status. Our unit of analysis is the head of household, and some characteristics of the household more broadly are associated with those individuals. Table 1 presents the variable definitions and summary statistics for the overall sample, as well as for the overall subgroups of those who we consider to be welfare recipients and non-recipients.

**Table 1. Variable Names, Labels and Descriptive Characteristics (weighted)**

| Variable label | Overall | Welfare Rs | Non-Rs |
|---|---|---|---|
| Charity Variables | | | |
| Donated at all | 66.9 % | 29.2 | 69.8 ** |
| Volunteered at all | 28.5 % | 13.8 | 29.6 ** |
| Amount donated ($) | $1,418 | $159 | $1,515 ** |
| Hours volunteered | 48.3 | 29.9 | 49.7 ** |
| Income Variables | | | |
| Earned income | $47,822 | $10,288 | $50,712 ** |
| Public assistance income | $160 | $2,221 | 0 ** |
| Tax itemizer | 42.4 % | 6.5 | 45.1 ** |
| Public Assistance Use Variables | | | |
| Prior public assistance use (%) | 11.2 % | 58.2 | 7.6 ** |
| Prior public assistance use (yrs) | 0.28 | 1.9 | 0.2 ** |
| Current public housing residence (%) | 3.4 % | 15.0 | 2.5 ** |
| Prior public housing residence (yrs) | 0.13 | 0.57 | 0.10 ** |
| Household Characteristics | | | |
| Age of head (yrs) | 49.4 | 42.1 | 49.9 ** |
| Sex of head (female) | 29.2 % | 60.0 | 26.9 ** |
| Number of children | 0.6 | 1.3 | 0.5 ** |
| Marital status (married) | 50.2 % | 21.4 | 52.4 ** |
| African American | 14.1 % | 42.2 | 12.0 ** |
| Hispanic | 7.1 % | 13.3 | 6.6 ** |
| Education (<=high school) | 49.3 % | 76.7 | 47.2 ** |
| Education (yrs completed) | 17.4 | 14.8 | 17.6 ** |
| Catholic | 24.0 % | 17.1 | 24.5 ** |
| Jewish | 3.4 % | 0.0 | 3.6 ** |
| Protestant | 57.9 % | 66.0 | 57.3 ** |
| Rural location (1-10 scale) | 3.5 | 3.5 | 3.8 * |
| Good health (1-5 scale) | 3.5 | 3.5 | 2.9 ** |
| Number of observations | 7,822 | 891 | 6,931 |

*Notes*:
** statistically significant difference; p<0.05
* statistically significant difference; p<0.10

As Table 1 shows, major differences exist in the characteristics of those who are public assistance recipients and those who are not. For example, in terms of our dependent variables of interest, welfare recipients donate less money and less time to charitable causes. As expected, they also have much less income from work, are younger, are less white, and are more likely to have a female household head. These differences are statistically significant, are generally large in magnitude, and very likely contribute to important differences in their outcomes. This table reports only the descriptive characteristics, but controlling for between-group differences is essential to understanding whether, given these differences, welfare recipients are more or less generous with their time or limited money, when it comes to acting charitably. For this reason, we spend some time in the next section considering how to deal with the bias that this would undoubtedly influence any estimate of the impact of public assistance receipt on charitable activity.

*Analytic Methods*

Both dependent measures are continuous, are truncated at zero, and include relatively large numbers of observed zeroes. Prior research has shown that charitable giving data may have a non-normal and heteroskedastic error structure (Rooney, Steinberg & Schervish, 2001, 2004; Steinberg, Rooney & Chin, 2002). Under these circumstances, Ordinary Least Squares (OLS) regression coefficients are biased and inconsistent. In line with much of the literature on charitable giving, we estimate our models using Tobit, a censored regression technique (Brooks, 2002; Van Slyke & Brooks, 2005).[3]

With this functional form, our prior work has used a difference-in-difference strategy and propensity score matching (Author1 & Author4, 2011) to reduce the amount of bias in our impact estimates that derived from selection processes, primarily due to the selection process associated with welfare use. While we stand behind that approach and our prior work, at the very least for the transparency of the approach, we believe it is important to re-test our results using different methods, both as a robustness check and to advance research methods used in examining policy effects.

In turn, this paper uses an atheoretical approach that we propose minimizes bias from selection. The analysis clusters cases as similar; and within each cluster the comparison of

[3] It should be noted that, while Tobit accommodates the large number of zeroes in the dependent variable, it is not robust to either non-normality or heteroskedasiticity. However, a recent study by Wilhelm (2006) provides support for using Tobit with this particular data, arguing that despite rejection of the underlying assumptions of Tobit at higher levels of statistical significance, the Tobit estimates are numerically close to more robust methods.

welfare and non-welfare users generates an estimate of impact with minimal bias from selection. Two features of our analysis are especially unique, as previously described in Author2 (2009) and Author2 and Author3 (2010) and applied in Author1, Author3 and Author2 (2009). The first distinct feature is our use of cluster analysis, specifically incorporating data transformed through Multiple Correspondence Analysis (MCA). The MCA creates a data structure that eliminates variance in the X matrix that is associated with treatment assignment, and the cluster analysis enhances the possibility of finding local spaces in which variables are balanced.

The second feature – the measure of Global Imbalance – is a single measure of the balance in data and is based on the concept of inertia as a measure of dependence between categorical covariates and the treatment assignment indicator. Whereas common practice is to assess variable-by-variable the extent to which comparison and treatment groups are matched, the GI measure allows an overall assessment of how well-matched cases are. For example, while Table 1 considers one-by-one the differences between the two groups, the GI measure assesses the comparability between groups, taking into account variation in all baseline covariates simultaneously. This is particularly useful in our case because we expect to examine many subsets of matched treatment and comparison cases, and a variable-by-variable assessment of their balance is likely not only to be tedious but also to reveal differences that exist purely by chance, some of which may be real and others of which may be only random, the distinction between which is un-knowable. Appendix A details the technical aspects of the GI measure and its test of statistical significance.

In brief, we use a three-step approach for estimating unbiased treatment effects in non-experimental data, as follows:

1. Measure and test balance: compute GI measure on the whole sample and test its statistical significance.

2. If imbalance exists (which is probable in all non-experimental data), perform a subgroup analysis, which involves the following:

   - Use Multiple Correspondence Analysis (MCA) to obtain a continuous and a low-dimensional representation of the X-space.

   - Enter the MCA coordinates into a cluster analysis to identify homogeneous groups.

3. Measure and test balance within each cluster, and compute local treatment effects within balanced clusters, pruning observations in unbalanced clusters.

Intuitively, when data are not balanced, a transition from the global predictor space to local predictor spaces is done by means of cluster analysis in order to estimate unbiased treatment effects. This approach can be considered a subgroup analysis where the primary goal is to get a cluster partition that generates balanced groups, thereby minimizing selection bias and generating unbiased impact estimates.

**Results & Discussion**

This section reports on the results from each of the three steps described above, and then presents findings from this process, comparing them to results from our prior work and then discussing implications.

*Step 1*. In implementing the three-step analysis, we begin by computing the GI measure for the entire sample. As reported in Table 2, the resulting value of 0.0406 can be interpreted as demonstrating the presence of imbalance in data. This is not surprising given the obvious differences revealed in Table 1. The GI measure falls in the critical region, thereby demanding data adjustment in order to estimate treatment effects not biased by selection. Essentially, Table 2 reports a single measure of sample balance; whereas Table 1 identified the variable-by-variable differences in sample characteristics.

**Table 2. Balance in the Overall Sample**

| Treatment | Comparison | GI | Interval | balance |
|---|---|---|---|---|
| 891 | 6,931 | 0.0406 | (0;0.0006) | no |

*Step 2*. The second step involves cluster analysis to identify homogeneous groups on the basis of MCA coordinates. The MCA was carried out using the following variables: age of head, sex of head, number of children, age of youngest child, African-American, Hispanic, marital status, education, single female, years of school completed, rural location, health of head, work status (working, retired, disabled), head hours worked, spouse hours worked, earned income, prior public assistance use, current and prior public housing residence.

The result of the MCA is a set of new variables (factorial coordinates) that are continuous and orthogonal one other. On the basis of these new MCA coordinates, we perform a cluster analysis to generate homogeneous groups. Although cluster analysis is an atheoretical approach, the variables we enter in to this preceding MCA regardless have theoretically justification as useful predictors of public assistance receipt. While others have used cluster analysis to understand how various treated subgroups are impacted (e.g., Author1, 2005; Yoshikawa, Rosman, & Hsueh, 2001), what makes this approach different is both (1) the MCA that transforms the data to minimize bias in cluster-based treatment-comparison group impact analyses, and (2) the introduction of the GI measure to conclude if a particular cluster has balanced treatment-comparison cases. In previous research (Author1, Author3, & Author2, 2009), we used a Ward algorithm on the MCA coordinates to generate clusters. Here, the sample size is greater than 5,000 units and therefore requires substantial computational resources to perform the distance matrix computation. To overcome this computational challenge, we use a mixed clustering method (Lebart et al., 1984), a combination of clustering around moving centers and Ward's hierarchical aggregation of stable groups, which is the best adapted to large datasets. We use the SAS software for MCA and cluster analysis.

The result of the mixed clustering method is depicted on a dendrogram, a tree diagram used to document the clustering process. Based on the dendrogram's structure, we mostly closely examine the 8-, 13-, 19-, 24-, 31- and 40-cluster solutions. We retain the 40-cluster solution because it provides balance within a suitable number of clusters, compared with other examined cluster solutions. With the 40-cluster solution, we measure and test balance within each group, using the GI measure, which is a clear advantage over the more traditional variable-by-variable assessment of group similarity.

Table 3 shows the results of this cluster analysis in terms of balance, including the number of treatment and comparison cases that each cluster includes. In this illustration, 13 of the clusters result in having unbalanced characteristics by our GI measure. In total, they represent about 35.7 percent of the observations of the original sample, which are then excluded from our third analytic step. Compared to a more commonly-used propensity score matching approach (e.g., Author1 & Author4, 2011), this cluster-based approach retains much more of the sample.

**Table 3. Balance by Clusters, 40-Cluster Solution**

| Cluster | Comparison | Treatment | GI | Interval | Balance |
|---|---|---|---|---|---|
| 1 | 801 | 5 | 0.003 | (0;0.004) | Yes |
| 2 | 250 | 14 | 0.024 | (0;0.015) | No |
| 3 | 446 | 8 | 0.006 | (0;0.007) | yes |
| 4 | 250 | 9 | 0.017 | (0;0.017) | yes |
| 5 | 111 | 3 | 0.025 | (0;0.030) | yes |
| 6 | 40 | 2 | 0.072 | (0;0.073) | yes |
| 7 | 91 | 10 | 0.022 | (0;0.034) | yes |
| 8 | 31 | 3 | 0.065 | (0;0.102) | yes |
| 9 | 171 | 3 | 0.024 | (0;0.021) | no |
| 10 | 97 | 6 | 0.028 | (0;0.033) | yes |
| 11 | 68 | 26 | 0.040 | (0;0.042) | yes |
| 12 | 137 | 10 | 0.035 | (0;0.025) | no |
| 13 | 133 | 4 | 0.058 | (0;0.027) | no |
| 14 | 309 | 29 | 0.010 | (0;0.010) | yes |
| 15 | 119 | 43 | 0.036 | (0;0.026) | no |
| 16 | 169 | 8 | 0.014 | (0;0.019) | yes |
| 17 | 138 | 10 | 0.036 | (0;0.022) | no |
| 18 | 263 | 14 | 0.015 | (0;0.015) | yes |
| 19 | 218 | 19 | 0.010 | (0;0.014) | yes |
| 20 | 427 | 9 | 0.009 | (0;0.009) | yes |
| 21 | 59 | 15 | 0.022 | (0;0.048) | yes |
| 22 | 288 | 20 | 0.017 | (0;0.012) | no |
| 23 | 63 | 10 | 0.018 | (0;0.050) | yes |
| 24 | 70 | 11 | 0.030 | (0;0.042) | yes |
| 25 | 195 | 25 | 0.027 | (0;0.015) | no |
| 26 | 85 | 99 | 0.020 | (0;0.020) | yes |
| 27 | 130 | 36 | 0.035 | (0;0.019) | no |
| 28 | 139 | 12 | 0.023 | (0;0.025) | yes |
| 29 | 366 | 17 | 0.009 | (0;0.009) | yes |
| 30 | 258 | 78 | 0.020 | (0;0.010) | No |
| 31 | 45 | 24 | 0.038 | (0;0.059) | Yes |
| 32 | 63 | 33 | 0.039 | (0;0.039) | Yes |
| 33 | 11 | 12 | 0.070 | (0;0.161) | Yes |
| 34 | 43 | 43 | 0.033 | (0;0.046) | Yes |
| 35 | 75 | 118 | 0.024 | (0;0.021) | Yes |
| 36 | 123 | 12 | 0.034 | (0;0.028) | No |
| 37 | 13 | 38 | 0.044 | (0;0.068) | Yes |
| 38 | 59 | 38 | 0.021 | (0;0.039) | Yes |
| 39 | 243 | 14 | 0.016 | (0;0.013) | no |
| 40 | 334 | 2 | 0.023 | (0;0.011) | no |

*Notes*: GI refers to the measure of global imbalance, which assesses holistically whether within-cluster treatment and comparison cases are balanced (similar) or imbalanced (dissimilar).

*Step 3*. During the final stage of the procedure, we analyze the effects of public assistance receipt on two outcomes of interest, the amount donated to charity and the number of hours volunteering within each of the remaining 27 balanced clusters. We estimate our giving and volunteering models by cluster, fitting a Tobit model (Verbeek, 2008) as follows:

$$y^*_i = \alpha + \beta w_i + \varepsilon_i \qquad i = 1,2,\ldots,N$$

$$\begin{aligned} y_i &= y_i^* \quad \text{if} \quad y_i^* > 0 \\ y_i &= 0 \quad \text{if} \quad y_i^* \leq 0 \end{aligned}$$

where

*y* is the outcome of interest (dollars donated or hours volunteered),

*w* is a vector of indicators of charitable activity,

$\boldsymbol{\varepsilon}$ is a random disturbance term, assumed to be $NID(0, \sigma^2)$ and independent from $w_i$, and

the subscript *i* indexes individuals.

Using the SAS qlim procedure, we fit a Tobit model for each cluster to estimate the effect of public assistance use on the outcome of interest. At this point, we do not include any additional variables in the model, at least in part because the number of observations of either treatment or comparison group members is sometimes smaller than the number of explanatory variables we would include. Instead, we perform a single regression within each of the 27 balanced clusters, under the presumption that they include well-matched treatment and comparison group members on the basis of the GI measure and test's result within clusters; and, while heterogeneous, their selection as cluster-matched cases means that, on aggregate, they should also retain the properties of supporting solid internal validity.

Because Tobit coefficients are not directly interpretable, we also compute and report marginal effects (Greene, 1999) to judge the magnitude and importance of estimated effects. The marginal effect represents the instantaneous effect that a change in a certain variable has on the dependent variable while keeping all the other covariates constant.

Our main findings appear in Table 4, which summarizes the Tobit regression results along with the computed marginal effects of receiving public assistance on the amount of money donated and on the number of hours volunteered.  Considering the dependent variable of the amount of money donated, among these 27 balanced clusters, five show a positive effect of welfare receipt on donations and the remaining 22 show a negative effect.  None of the positive effects are statistically significant.  Fourteen of the negative effects are statistically significant; and among those, just five are in clusters where we have sufficient sample size to be confident of the results.  The magnitude of the effect between the treatment and comparison group members in Cluster 18, for example, would be interpreted as follows:  holding all other variables constant at the mean, being a public assistance recipient is associated with making $1,052 *less* in charitable contributions relative to those who do not receive assistance.

**Table 4. The Effect of Public Assistance Use on Charitable Activity, by Balanced Cluster**

| | Amount of Money Donated | | | Number of Hours Volunteered | | |
|---|---|---|---|---|---|---|
| Cluster | Parameter Estimate | Marginal Effect | P-value | Parameter Estimate | Marginal Effect | P-value |
| 1^ | -2,321.1 | -1,583.36 | <0.01 ** | -226.0 | -83.1 | 0.26 |
| 3^ | -1,254.9 | -816.94 | <0.01 ** | -171.7 | -49.1 | 0.36 |
| 4^ | -2,062.2 | -1,254.06 | <0.01 ** | -122.2 | -34.0 | 0.43 |
| 5^ | -3,894.6 | -2,240.43 | <0.01 ** | -769.6 | -209.3 | <0.01 ** |
| 6^ | -1,358.1 | -751.48 | <0.01 ** | 93.1 | 20.9 | 0.72 |
| 7 | -784.7 | -370.96 | 0.35 | -295.1 | -99.3 | 0.13 |
| 8^ | -11,939.0 | -4,000.14 | <0.01 ** | -1,651.3 | -327.1 | <0.01 ** |
| 10^ | -2,485.0 | -1,137.76 | <0.01 ** | 247.1 | 43.5 | 0.16 |
| 11 | -164,453.0 | -510.47 | 0.03 ** | -55.0 | -8.6 | 0.64 |
| 14 | -1,022.4 | -454.91 | 0.13 | -141.6 | -30.5 | 0.44 |
| 16^ | -5,101.0 | -2,459.59 | <0.01 ** | 437.9 | 100.6 | 0.03 ** |
| 18 | -1,961.0 | -1,052.91 | <0.01 ** | -176.9 | -40.8 | 0.36 |
| 19 | -638.4 | -256.03 | 0.34 | -122.6 | -19.2 | 0.52 |
| 20^ | -1,795.9 | -1,055.62 | <0.01 ** | -58.1 | -14.4 | 0.69 |
| 21 | -1,081.9 | -406.93 | 0.17 | -59.6 | -9.9 | 0.82 |
| 23 | 498.6 | 82.96 | 0.32 | 68.9 | 6.7 | 0.31 |
| 24 | 443.9 | 153.90 | 0.41 | -37.9 | -8.1 | 0.58 |
| 26 | 14.3 | 3.32 | 0.98 | -106.8 | -25.8 | 0.56 |
| 28 | -1,456.6 | -779.13 | 0.01 ** | -51.5 | -10.9 | 0.59 |
| 29 | -1,357.4 | -808.69 | 0.17 | -23.9 | -9.3 | 0.82 |
| 31 | -113.3 | -39.78 | 0.78 | 14.4 | 2.1 | 0.65 |
| 32 | -642.1 | -206.03 | 0.05 ** | -231.8 | -35.9 | 0.48 |
| 33 | 104.6 | 24.28 | 0.47 | -2,999.6 | -437.6 | <0.01 ** |
| 34 | -864.0 | -186.92 | 0.10 * | -6.0 | -0.8 | 0.98 |
| 35 | -138.1 | -34.38 | 0.68 | 68.8 | 14.6 | 0.59 |
| 37 | 172.2 | 37.36 | 0.32 | 3,422.6 | 67.9 | <0.01 ** |
| 38 | -188.5 | -60.85 | 0.68 | -39.3 | -3.6 | 0.72 |

*Notes*:
^ Although we estimate parameters, we urge caution in interpretation because of low statistical power due to few observations for a treatment level.
The parameter estimate is that associated with the treatment variable, indicating that the sample member received welfare. So, these are interpreted as the effect of receiving welfare on the money and time donations. No additional controls are included.
** Statistically significant, p<0.05
* Statistically significant, p<0.10

Next, considering volunteer hours, only two of the five statistically significant results are in clusters where we have enough observations to be confident in the results. That said, those two effects operate in opposite directions: the finding from Cluster 33 suggests that those public assistance recipients volunteer many fewer hours (437 per year) than non-recipients; whereas those recipients in Cluster 37 volunteer 68 more hours annually than their non-recipient counterparts, holding other variables constant.

Each of these individual analyses within cluster is between matched members of the treatment group (received public assistance) and the comparison group (did not receive public assistance). In part for the observation that a cluster-by-cluster analysis involves relatively few cases within some selected clusters, we are prevented from controlling for other variables that might explain charitable activity. Recall that cluster membership was determined by the MCA coordinates, which were based on a theoretical model of public assistance use, and so we assert that the treatment and comparison cases within each cluster minimize selection bias related to public assistance use; but there may be variability in the data related to charitable activity that we could control for in increasing the precision of our estimates of the impact of public assistance receipt on charitable activity. If this is the case, then future analysis might consider collectively all of the observations within the balanced clusters in an aggregate analysis that would permit more precise estimate of impacts through controlling for other variables.

While we undertake this analysis to maximize the internal validity of our findings (that is minimize bias in impact estimates), we also remain concerned with the external validity. That is, our process resulted in some unbalanced clusters, the cases we pruned from subsequent analysis. Further, several of the remaining clusters show no effect of having received public assistance on either money or time donated to charity. We might be interested to know the characteristics of those whose public assistance receipt has no effect on their charitable activity; and we are also likely interested in the characteristics of those clusters in which there is a statistically significant effect, at least in part to gauge the extent to which these results might be generalizable.

In response, we provide the descriptive characteristics of those clusters in which balance between the treatment group (welfare recipients) and comparison group (non-recipients) exists and there is some evidence of a statistically significant effect. Tables 5.1 and 5.2 present these results.

**Table 5.1. Descriptive Statistics, by Cluster, among Selected Balanced Clusters (with small sample size)**

| Variable | Sample mean | Cluster 1 | 3 | 4 | 5 | 6 | 8 | 10 | 16 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|
| Dollars effect | | – ** | – ** | – ** | – ** | – ** | – ** | – ** | – ** | – ** |
| Hours effect | | – | – | – | – | + | – ** | + | + ** | – |
| Observations | | 806 | 454 | 259 | 114 | 42 | 34 | 103 | 177 | 436 |
| Treatment | | 5 | 8 | 9 | 3 | 2 | 3 | 6 | 8 | 9 |
| Comparison | | 801 | 446 | 250 | 111 | 40 | 31 | 97 | 169 | 427 |
| Age of head | 44.7 | 44.9 | 43.2 | 41.3 | 43.0 | 36.5 | 35.6 | 39.3 | 45.1 | 43.0 |
| Female head (%) | 29.5 | 0.0 | 1.3 | 4.6 | 2.7 | 0.0 | 17.6 | 2.9 | 1.7 | 0.0 |
| # children | 0.90 | 1.04 | 0.92 | 0.97 | 0.83 | 0.95 | 1.11 | 1.16 | 0.72 | 1.06 |
| African-Am. (%) | 33.7 | 13.5 | 26.6 | 31.7 | 9.7 | 7.1 | 5.9 | 25.2 | 24.3 | 37.8 |
| Hispanic (%) | 6.1 | 0.0 | 0.0 | 0.0 | 5.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Married (%) | 51.1 | 91.6 | 77.3 | 69.5 | 73.7 | 69.1 | 67.7 | 68.0 | 76.3 | 90.1 |
| ≤high school (%) | 31.7 | 0.0 | 0.0 | 0.0 | 100.0 | 0.0 | 26.5 | 100.0 | 0.0 | 100.0 |
| Working (%) | 75.4 | 97.8 | 95.2 | 93.8 | 93.9 | 92.9 | 91.2 | 92.2 | 97.2 | 97.5 |
| Retired (%) | 10.8 | 0.0 | 0.0 | 0.0 | 0.9 | 2.4 | 2.9 | 1.0 | 0.0 | 0.0 |
| Head work hrs | 1,696 | 2,344 | 2,236 | 2,194 | 223 | 2,141 | 2,138 | 2,279 | 2,225 | 2,315 |
| Spouse work hrs | 732 | 1,511 | 1,154 | 1,145 | 1,040 | 1,276 | 1,098 | 1,155 | 1,193 | 1,643 |
| Earned income($) | 47,269 | 8,279 | 55,791 | 66,091 | 51,686 | 54,852 | 52,070 | 53,216 | 49,817 | 76,261 |
| Prior PA (yrs) | 0.44 | 0.01 | 0.05 | 0.08 | 0.20 | 0.19 | 0.14 | 0.08 | 0.06 | 0.06 |
| Pub. housing (%) | 5.13 | 0.12 | 1.10 | 1.93 | 1.75 | 2.38 | 0.00 | 1.94 | 0.56 | 0.69 |
| Prior housing (yrs) | 0.19 | 0.02 | 0.07 | 0.07 | 0.02 | 0.04 | 0.02 | 0.17 | 0.05 | 0.103 |

*Notes*:
These selected balanced clusters are those marked in Table 1 as having low statistical power due to few observations for a treatment level. We therefore urge caution in interpretation of the effects.
** Statistically significant, $p<0.05$
* Statistically significant, $p<0.10$

**Table 5.2. Descriptive Statistics, by Cluster, among Selected Balanced Clusters (with larger sample size)**

| Variable | Sample mean | Cluster 11 | 18 | 28 | 32 | 34 | 37 |
|---|---|---|---|---|---|---|---|
| Dollars effect | | – ** | – ** | – ** | – ** | – ** | + |
| Hours effect | | – | – | – | – | – | + ** |
| Observations | | 94 | 277 | 151 | 96 | 22 | 51 |
| Treatment | | 26 | 14 | 12 | 33 | 12 | 38 |
| Comparison | | 68 | 263 | 139 | 63 | 11 | 13 |
| Age of head | 44.7 | 38.7 | 42.8 | 44.4 | 41.3 | 58.3 | 51.3 |
| Female head (%) | 29.5 | 78.3 | 3.3 | 0.0 | 99.6 | 59.1 | 56.8 |
| # children | 0.90 | 1.06 | 0.89 | 0.80 | 1.58 | 1.22 | 0.56 |
| African-Am. (%) | 33.7 | 69.2 | 16.3 | 55.6 | 88.5 | 50.0 | 78.4 |
| Hispanic (%) | 6.1 | 0.0 | 0.0 | 0.0 | 7.3 | 9.1 | 0.0 |
| Married (%) | 51.1 | 2.1 | 72.9 | 0.0 | 0.0 | 36.4 | 17.7 |
| ≤high school (%) | 31.7 | 36.2 | 46.2 | 100.0 | 43.8 | 18.2 | 43.1 |
| Working (%) | 75.4 | 72.3 | 93.9 | 95.4 | 88.5 | 4.6 | 0.0 |
| Retired (%) | 10.8 | 0.0 | 0.0 | 0.0 | 0.0 | 31.8 | 0.0 |
| Head work hrs | 1,696 | 1,509 | 2,217 | 2,028 | 1,981 | 156 | 50 |
| Spouse work hrs | 732 | 37 | 1,083 | 0.0 | 0.0 | 195 | 167 |
| Earned income($) | 47,269 | 18,254 | 52,609 | 36,378 | 17,915 | 3,903 | 1,824 |
| Prior PA (yrs) | 0.44 | 0.39 | 0.10 | 0.32 | 3.46 | 3.54 | 6.01 |
| Pub. housing (%) | 5.13 | 18.09 | 0.36 | 2.65 | 20.83 | 18.18 | 13.73 |
| Prior housing (yrs) | 0.19 | 0.26 | 0.06 | 0.30 | 1.09 | 0.86 | 1.19 |

*Notes*:
These selected balanced clusters are those marked in Table 1 as having sufficient numbers of observations to instill greater confidence in interpretation of the effects.
** Statistically significant, $p<0.05$
* Statistically significant, $p<0.10$

In brief, it is clear that our clusters represent relatively heterogeneous subgroups of the overall sample. Many dimensions of variation exist, even along just those variables that we have measured here. The consistent effect across most of the clusters (where results are statistically significant) is that current public assistance receipt is associated with lower donations to charity. This is in line with our prior work, which found that public assistance use itself suppressed donations of money, but that as public assistance income rises so too do money donations (Author4 & Author1, 2009). Note, however, that the present analysis does not (yet) consider the role of the amount of public assistance income, only public assistance receipt, since the cluster-by-cluster analysis includes only an unbiased selection-to-welfare (treatment) indicator. Most of the reported effects on volunteer hours are not statistically significant; but, among the three that are, two of them are positive effects, which is sync with our prior research and supports the possibility that donations of money and time are substitutes for one another. Our research here, however, suggests that this is not a widespread finding but instead isolated to some specific sub-populations. The characteristics of the subgroup represented in cluster 37 suggest that it might be classified as “children of the Great Society.” Specifically, with an average age of 51, this somewhat older group was born around 1953, and they are relatively evenly split between men and women. With very low income from work, and a much greater proportion being African American, compared to the sample mean, they have greater connection to public assistance, with an average of six years (of the prior seven) as public assistance recipients, and greater proportions living in public housing. Despite these characteristics, this group has greater levels of volunteering, perhaps suggesting that their upbringing during the Great Society and Civil Rights movements of the sixties is associated with greater evidence of giving back.

**Conclusion**

Prior research has considered the effects of public policy on charitable activity, although relatively little of that considers the role of public assistance. Prior work has aimed to minimize problems associated with selection bias in estimating policy impacts, and the current work makes another attempt to do so. Here we apply a creative clustering approach devised and applied in Author2 (2009), Author3 and Author2 (2010) and Author1, Author3, and Author2 (2009).

Although our prior work made the case for this approach on the basis of it minimizing researcher influence over elements of the process, this particular application highlights that this objectivity may in fact be somewhat compromised. Specifically, the choice of algorithm to use in cluster analysis can have important implications for analytic results (e.g., Aldenderfer & Blashfield, 1984), and here we chose to use a mixed classification method. It is possible that a different choice at this point in the analysis might have resulted in a different combination of treatment and comparison cases being grouped together, perhaps in a different number of clusters. This observation suggests that future research that uses this approach might test the sensitivity of various algorithms at the cluster analysis stage. Nevertheless, as Lebart et al. (1984) mention, some of the greatest minds in Multivariate Statistical Analysis of the French School of *Analyse des données*, the efficacy of the method of classification around moving centers is largely attested to by empirical results and is the partitioning method that best accommodates large datasets like ours.

Related is the point that the selection of an optimal cluster solution also involves some subjectivity. Although several reasonable solutions can be identified, the rationale for choosing one over another is not objective. In fact, our selection of the 40-cluster solution was based on the rationale that it seemed like that solution retained enough of its cases to be useful; but as analysis continued, we observed that several of the resulting clusters had too few treatment cases to instill confidence. This is another arena where this method might be subjected to further sensitivity testing

Despite these caveats, we believe this approach has promise. It includes an improvement over conventional propensity score matching, which requires subjective judgment on model specification. In contrast, our strategy groups each observation in the treatment group with those in the control group whose observed characteristics are similar; within balanced clusters, units are different only with respect to whether they received treatment. In turn, estimated impacts are not biased in any systematic way. Such a conclusion is enhanced by the observation that the GI measure is a global measure of comparability between groups, objective because it is based on the concept of variance observed in the data, among baseline covariates (considered simultaneously) and the assignment-to-treatment indicator variable. It is also enhanced by the use of the multivariate imbalance test that allows determining the imbalance's significant, thereby overcoming standard variable-by-variable test of balance that can not consider interaction among variables. This strategy is based on the assumption of "no omitted variable bias," meaning that relevant variables involved in the selection process are known (Author3 & Author2, 2010). Future work might also explore the sensitivity of the multivariate test of imbalance to specific failure of the unconfoundedness assumption.

This paper has taken a question – how does public assistance receipt affect charitable giving? – and applied a new, cluster-based approach to minimize bias that arises specifically in the selection-to-welfare process. The preliminary substantive findings reported accord with prior research. Specifically, we observe some evidence that public assistance suppresses money donations, in line with Brooks (2002, 2004) and Author4 & Author1 (2009). Findings regarding volunteer hours are less clear, with evidence that there may be no effect (as in Author1 & Author4, 2011) or a positive effect (as in Author4 & Author1, 2009). What is clear is that there is substantial heterogeneity within the population, and this heterogeneity demands an analytic approach that recognizes it, as we had done here, rather than on ignoring it and assessing only average treatment effects. This work offers a useful application of a new method to capitalize on treatment group heterogeneity and minimize the effect of selection bias in computing treatment effects in non-experimental evaluations.

**Works Cited** (several author citations removed for blind review)

Andreoni, J., Brown, E., & Rischall, I. (2003). Charitable giving by married couples: Who decides and why does it matter? *Journal of Human Resources*, 38(1), 111-133.

Andreoni, J., Gale, W., & Scholz, J. (1996). *Charitable contributions of time and money*. Working paper. Madison, WI: University of Wisconsin.

Arai, S.M. (2004). Volunteering in the Canadian context: Identity, civic participation and the politics of participation in serious leisure. In Stebbins, R.A. & Graham, M. (eds.), *Volunteering as leisure/Leisure as volunteering* (pp. 151-176). Cambridge, MA: CABI Publishing.

Aldenderfer, M.S., & R.K. Blashfield. (1984). Cluster Analysis. Beverly Hills, CA: Sage Publications.

Brooks, Arthur. (2002). Welfare receipt and private charity. *Public Budgeting and Finance*, 22(3), 100-114.

Brooks, Arthur. (2004). The effects of public policy on private charity. *Administration and Society*, 36(2), 166-185.

Brown, E. (1999). Patterns and purposes of philanthropic giving. In C.T. Clotfelter & T. Ehrlich (Eds.), *Philanthropy and the nonprofit sector in a changing America*. Bloomington: Indiana University Press.

Brown, E., & Lankford, H. (1992). Gifts of money and gifts of time: Estimating the effects of tax prices and available time. *Journal of Public Economics*, 47, 321-341.

Bryant, W.K., Jeon-Slaugher, H., Kang, H., & Tax, A. (2003). Participation in philanthropic activities: Donating money and time. *Journal of Consumer Policy*, 26, 42-73.

Campbell, D.E., & Yonish, S.J. (2003). Religion and volunteering in America. In C. Smidt (ed.), *Religion as Social Capital: Producing the Common Good* (pp. 87-106). Waco, Texas: Baylor University Press.

Caputo, R. K. 1997. Women as volunteers and activists. *Nonprofit and Voluntary Sector Quarterly*, 26(2), 156-174.

Choi, L. (2003). Factors affecting volunteerism among older adults. *Journal of Applied Gerontology*, 22(2), 179-196.

Clotfelter, C.T. (1997). The economics of giving. In J.W. Barry & B.V. Manno (Eds.), *Giving better, giving smarter: Working papers of the National Commission on Philanthropy and Civil Renewal.* Washington, DC: National Commission on Philanthropy and Civic Renewal.

Duncan, B. (1999). Modeling charitable contributions of time and money. *Journal of Public Economics*, 72(2), 213-242.

Escofier, B. (1988). Analyse des Correspondances Multiples Conditionelle. In E. Diday (ed.), Data Analysis and Informatics, 333-342. North Holland, Amsterdam: Elsevier Science Publishers.

Fischer, L.R., Mueller, D.P., & Cooper, P.W. (1991). Older volunteers: A discussion of the Minnesota Senior Study. *The Gerontologist*, 31(2), 183-194.

Gallagher, S. K. (1994). Doing their share: Comparing patterns of help given by older and younger adults. *Journal of Marriage and the Family*, 56, 567-578.

Gittel, R. & E. Tebaldi. (2006). Charitable giving: Factors influencing giving in the U.S. states. *Nonprofit and Voluntary Sector Quarterly*, 35(4), 721-736.

Greenacre, M. J. (1984). Theory and Applications of Correspondence Analysis. London, UK: Academic Press.

Greene, W. (1999). Marginal Effects in the Censored Regression Model. *Economics Letter*, 64, 43-49.

Herzog, A.R.,& Morgan, J.S. (1993). Formal volunteer work among older Americans. In S.A. Bass, F.G. Caro, & Y.-P. Chen (Eds.), *Achieving a productive aging society* (pp.119-142). Westport, CT: Auburn House.

Hodgkinson, V.A., & Weitzman, M.S. (1996). *Giving and volunteering in the United States: Findings from a national survey*. Washington, DC: Independent Sector.

Jackson, E., Bachmeier, M., Wood, J.R., & Craft, C. (1995). Volunteering and charitable giving: Do religious and associational ties promote helping behavior? *Nonprofit and Voluntary Sector Quarterly*, 24(1), 59-78.

Jones, K.S. (2006). Giving and volunteering as distinctive forms of civic engagement: The role of community integration and personal resources in formal helping. *Nonprofit and Voluntary Sector Quarterly*, 35(2), 249-266.

Lebart,L., Morineau, A., & Warwick, K.M. (1984). *Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices.* New York, NY: Wiley & Sons.

Lebart, L., Morineau, A., & Piron, M. (1997). Statistique exploratoire multidimensionelle. Paris: Dunod.

McClelland, R., & Brooks, A.C. (2004). What is the real relationship between income and charitable giving. *Public Finance Review*, 32(5), 483-497.

McPherson, J.M., & Rotolo, T. (1996). Testing a dynamic model of social composition: Diversity and change in voluntary groups. *American Sociological Review*, 61, 179-202.

Menchik, P.L., & Weisbrod, B.A. (1987). Volunteer labor supply. *Journal of Public Economics*, 32(2), 159-183.

Mesch, D.J., Rooney, P., Steinberg, K., & Denton, B. (2006). The effects of race, gender, and marital status on giving and volunteering in Indiana. *Nonprofit and Voluntary Sector Quarterly*, 35(4), 565-587.

Mettler, S. (2002). Bringing the state back in to civic engagement: policy feedback effects of the GI Bill for World War II veterans. *American Political Science Review*, 96(2), 351-365.

Musick, M., & Wilson, J. (2008). *Volunteers: A social profile*. Bloomington and Indianapolis: Indiana University Press.

O'Neill, M., & Roberts, W. (2000). *Giving and volunteering in California*. San Francisco: University of San Francisco, Institute for Nonprofit Organization Management.

Rooney, P.M., Mesch, D.J., Chin, W., & Steinberg, K.S. (2005). The effects of race, gender, and survey methodologies on giving in the U.S. *Economic Letters*, 86, 173-180.

Schiff, J.A. (1990). *Charitable giving and government policy: An economic analysis*. New York: Greenwood Press.

Smith, W, & Chang, C. (2002). Shipping the good apples out: A note on contributions of time and money. *Economics Bulletin*, 10(1), 1-14.

Soss, J. (1999). Lessons of welfare: Policy design, political learning, and political action. *American Political Science Review*, 93(2), 363-380.

Steinberg, R. (1990). Taxes and giving: New findings. *Voluntas*, 1(2), 61-79.

Steinberg, R. (1993). Does government spending crowd out donations? Interpreting the evidence. In A. Ben-Ner and B. Gui (eds.), *The Nonprofit Sector in the Mixed Economy* (99-125). Ann Arbor: The University of Michigan Press.

Steinberg, K., Rooney, P., & Chin, W. (2002). Measurement of volunteering: A methodological study using Indiana as a test case. *Nonprofit and Voluntary Sector Quarterly,* 31(4), 484-501.

Steinberg R., & Wilhelm, M. (2003). Patterns of giving in COPPS 2001. Presented at the annual conference of Association for Research on Nonprofit Organizations and Voluntary Action, Denver, CO.

Steinberg, R., & Wilhelm, M. (2005). Religious and secular giving, by race and ethnicity. *New Directions for Philanthropic Fundraising*, 48, 57-66.

Van Slyke, D.M., & Brooks, Arthur C. (2005). Why do people give? New evidence and strategies for nonprofit managers. *American Review of Public Administration,* 35(3), 199-222.

Verbeek, M. (2008). *A guide to modern econometrics*. New York, NY: Wiley & Sons.

Wilhelm, Mark O. (2006). Practical considerations for choosing between Tobit and SLCS or CLAD estimators for censored regression models with an application to charitable giving. Mimeo, IUPUI.

Wilson, J. (2000). Volunteering. *Annual Review of Sociology*, 26, 215-240.

Wilson, J., & Musick, M. (1997). Who cares? Toward an integrated theory of volunteer work. *American Sociology Review*, 62(5), 694-713.

Wolff, E. (1999). The economy and philanthropy. In C.T. Clotfelter & T. Ehrlich (Eds.), *Philanthropy and the nonprofit sector in a changing America*. Bloomington, IN: Indiana University Press.

Yoshikawa, H., E. A. Rosman, and J. Hsueh. (2001). Variation in teenage mothers' experiences of child care and other components of welfare reform: Selection processes and developmental consequences. *Child Development*, 72(1), 299-317.

Ziemek, S. (2006). Economic analysis of volunteers' motivations - A cross-country study. *Journal of Socio-Economics*, 35, 532–555.

**APPENDIX A** (See Author1, Author 3 & Author2, 2009)

This technical appendix explains the computation of the measure of Global Imbalance (GI) and the imbalance test we propose using to ascertain whether selection bias (1) poses a problem for an evaluation's impact analysis and (2) has been eliminated from cluster subgroups.

### The GI Measure

Author2 (2009) and Author3 and Author2 (2010) report that the between-group inertia of a cloud of units denotes the GI measure expressed as:

$$GI = I_b = \frac{1}{Q}\sum_{t=1}^{T}\sum_{j=1}^{J_Q}\frac{b_{tj}^2}{k_{.t}k_{.j}} - 1$$

where

Q denotes the number of baseline covariates introduced in the analysis

T denotes the number of treatment levels;

$J_Q$ denotes the set of all categories of the Q variables considered;

$b_{tj}$ is the number of units with category $j \in J_Q$ in the treatment group $t \in T$;

$k_{.t}$ is the group size $t \in T$; and

$k_{.j}$ is the number of units with category $j \in J_Q$.

The GI measure is the result of using Conditional MCA (Escofier, 1988) that allows quantifying the between-group inertia. Such a measure originates from the consideration that when the dependence between **X** and **T** is out of control of researchers displaying the relationship among them on a factorial space represents a first step for discovering the hidden relationship. In fact, if dependence between **X** and **T** exists, any descriptive factorial analysis may exhibit this link.

A conventional method dealing with the factorial decomposition of the variance related to the juxtaposition of the **X** matrix and the **T** variable is Multiple Correspondence Analysis (MCA) framework.[4] Given that the variability (inertia) of a data matrix can be decomposed into eigenvalues and eigenvectors, and referring to MCA for the study of the relationship between variables and of the structure induced by variables on the population, the presence of a conditioning variable (**T**) will strongly influence the structure of the matrix decomposition process. Hence, a conditional analysis could be useful in order to isolate the part of the variability of the **X**-space due to the assignment mechanism. Conditioning applied to problems arising from the dependence between categorical covariates and an external categorical variable was first studied by Escofier (1988) with the resulting Conditional Multiple Correspondence Analysis (MCA_cond).

Referring to Huygens' overall inertia decomposition of total inertia ($\mathbf{I}_T$) as within-groups ($\mathbf{I}_W$) and between-groups ($\mathbf{I}_B$), MCA_cond consists in a factorial decomposition of the within-group inertia. In turn, MCA_cond could be also considered as an *intra* analysis since the inertia induced by the conditioning variable (**T**) is not taken into account. Specifically, an *inter-group* analysis considers the relative position of groups, whereas an *intra-group* analysis detects and describes differences between units within each group by not considering the effect due to the partition's structure. In the evaluation context, this structure is induced by the non-random selection mechanism. An intra-analysis allows measuring the influence of conditioning, which means, as reported in Author3 & Author2 (2010), obtaining a measure of comparability between treatment groups.

---

[4] For a comprehensive description of this method, computational details, and its applications, refer to Lebart et al. (1984), and for problems in the presence of a conditioning variable, refer to Author 3 and Author 2 (2010) and Author 2 and Author 3 (2011).

This method especially works in the presence of categorical covariates. Eventually, continuous variables could be transformed into categorical by dividing them in classes. The need to work with categorical covariates stems from the consideration that, as reported in Cox and Wermuth (1998), in the social sciences, background knowledge tends to be qualitative.

The key result of using MCA_cond is represented by the quantified "Between-group Inertia" ($\mathbf{I}_b$). The *no omitted variable bias* assumption underlying the approach assumes a crucial role and, thus, must be emphasized. The assignment mechanism is assumed to be known, which means that the **X** matrix includes all baseline variables associated with both the treatment assignment and the observed outcome.

### The Imbalance Test

To determine the significance of the detected imbalance, we perform an Imbalance test. We specify the null hypothesis of no dependence between **X** and **T** as:

$$\mathbf{H}_0 : \mathbf{I}_W = \mathbf{I}_T$$

To establish an interval of plausible values for $\mathbf{I}_B$ under the null hypothesis, we use results obtained by Estadella and Aluja (2005), who have studied the asymptotic distribution function of $\mathbf{I}_B$. Once derived the distribution of the between group inertia as $\mathbf{I}_B \approx \frac{\boldsymbol{\chi}^2_{(T-1)(J-1)}}{\mathbf{nQ}}$, the interval of plausible values for GI defined as:

$$GI \in (0, \frac{\chi^2_{(T-1)(J-1),\alpha}}{nQ})$$

Specifically, if the GI calculated on the specific dataset is outside the interval, then the null hypothesis of no dependence between **X** and **T** is rejected and data are deemed unbalanced.

Simulation results show that where the test detects balance unbiased estimates of the ATE are obtained (Author2, 2009; Author3 & Author2, 2010).