

Ibrahim Hanafy Mohammad, Hind

**Research Report**

## Welfare Effects of Social Cash Transfer Programs: Evidence from Egypt's Takaful and Karamah Programs

*Suggested Citation:* Ibrahim Hanafy Mohammad, Hind (2025) : Welfare Effects of Social Cash Transfer Programs: Evidence from Egypt's Takaful and Karamah Programs, ZBW – Leibniz Information Centre for Economics, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/319784>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# **Welfare Effects of Social Cash Transfer Programs: Evidence from Egypt's Takaful and Karamah Programs**

Hind Ibrahim Hanafy Mohammad

**Abstract**

Social cash transfers (SCTs) are considered a priority in least-developed countries, where the gap between the need for basic social protection and existing provisions is greatest. This study represents one of the first comprehensive impact evaluation treatments for Takaful and Karamah social cash transfer programs in Egypt. The results, based on propensity score matching (PSM) and odds-weighted regression, and data from the HIECS 2017-2018, confirm positive SCTs effects on per capita non-food consumption expenditures including healthcare and education for beneficiary households. The results also indicate threshold effects with SCTs mostly impacting healthcare expenditure among asset-poorer beneficiary households and education expenditure among asset-wealthier beneficiaries.

**Key words:** Social cash transfers, impact evaluation, consumption expenditure, propensity score matching, odds-weighted regression

## Table of Contents

Introduction .....	4
1.1. Propensity Score Methods in Observational Studies .....	5
1.1.1. The Evaluation Problem .....	5
1.1.2. Identification Assumptions .....	7
1.1.3. Observational Data vs. Randomized Control Experiments and the Role of Propensity Score .....	8
1.2. Practical Advantages of Propensity Score Methods .....	11
1.3. Propensity Score in Matching Methods .....	13
1.4. Model Specification .....	14
1.5. Propensity Score in Regression Adjustment: Propensity Score Weighting (Odds-Weighted Regression) .....	15
1.6. The Heterogeneous Impact .....	17
1.7. Data and Variables .....	17
1.8. Empirical Results and Analysis .....	19
1.8.1. Impact Estimation .....	22
1.8.2. Balancing Tests for Propensity Score Model .....	25
1.9. Discussion and Concluding Remarks .....	32
<b>References</b> .....	38
<b>Appendix (A)</b> .....	50
<b>Appendix (B)</b> .....	52

## Introduction

The recent economic literature on assessing the distributional impacts of public and social policies or interventions on target beneficiaries is increasingly relying on a class of estimators known as the propensity score estimators. These estimators belong to an array of inference methods for average treatment effects that depend on non-parametric approaches for estimation in a non-experimental environment (Caliendo and Kopeining, 2008; Heinrich et al, 2010; Imbens, 2004; Dehejia and Wahba, 2002; Smith and Todd, 2005). In many settings, estimating treatment effects based on propensity score estimators have several advantages over regression-based estimators. Propensity score methods provide the merit of studying the heterogeneity of intervention impact through allowing comparability conditional on knowledge of the pre-intervention state (Jalan and Ravallion, 2003). In addition, they minimize the dependence on functional forms and are more transparent in terms of the modeling process (Keele, 2010).

In principal, the non-experimental design techniques are called for in measuring the ex-post effects of social interventions on treated groups whenever there is a difficulty in employing the experimental design; i.e. when random selection is lacking (as when using observational data) (Abadie and Imbens, 2011; White and Sabarwal, 2014). These techniques were first introduced by the medical literature, and were developed to circumvent the problem of infeasible clinical trials (Cameron and Trivedi, 2005; Liu et al, 2013). From a practical point of view, the ease of implementation and interpretation of the non-experimental analyses techniques had largely gained them a considerable popularity across different disciplines and underscored these methods as important tools for estimating causal associations for impact evaluation studies (Imbens, 2004; Liu et al, 2013).

In this chapter, following Tembo et al (2014), the non-experimental approaches of Propensity Score Matching (PSM) and the odds-weighted regression approach—also referred to as Propensity Score Weighting (PSW)—<sup>1</sup> are applied for estimating the impact of the Egyptian social cash transfer schemes of Takaful and Karamah on per capita consumption expenditures of beneficiary households in relevance to their non-beneficiary counterparts using the Egyptian Household, Income, Expenditure and Consumption Survey (HIECS) 2017-2018 dataset.

---

<sup>1</sup> These two approaches are considered the most widely used approaches in the literature of treatment evaluation (Frölich, 2007).

The rest of this chapter is organized as follows. Part (1.1) sets up the framework for causal inference using propensity score methods and the identification assumptions. Part (1.2) addresses the practical advantages of propensity score methods over regression-based methods. Part (1.3) focuses on matching on the propensity score for the estimation of average treatment effect. Part (1.4) identifies the specifications of the applied propensity score model in the study at hand. Part (1.5) introduces the propensity score weighting (or odds-weighted regression) as an approach for enhancing the efficiency of propensity score estimators. Part (1.6) discusses the heterogeneous impact, which extends the impact measurement across subgroups defined according to asset-wealth. A brief description of the utilized dataset, Takaful and Karamah social schemes at the relevant time-point and variables selection are featured in part (1.7) and part (1.8) presents the empirical results and interpretation. Finally, part (1.9) provides a brief discussion of the results and some concluding remarks.

## 1.1. Propensity Score Methods in Observational Studies

### 1.1.1. The Evaluation Problem

In estimating the causal effects of interventions or treatments on subject populations, two measures are considered of interest. The first measure is the Average Treatment Effect (*ATE*), which is the average effect of treating the entire subject population. This measure is also referred to as the “marginal treatment effect” (Rosenbaum and Rubin, 1983). If ( $Y_i$ ) is to denote the outcome of interest that is realized by the  $i$ -th unit of an  $N$ -element population as a result of receiving a certain treatment, and ( $T_i$ ) denotes the treatment status for unit  $i$ , where ( $T$ ) is a dichotomous variable that takes the value of (1) for being treated and (0) otherwise;<sup>2</sup> the *ATE* is estimated as:

$$ATE = E[Y_{i1} - Y_{i0}] \quad ; \quad i = 1, \dots, N \quad (1)$$

---

<sup>2</sup> This is the case of a single treatment type. If the treatment applied has more than two categorical variations, then the variations are collectively referred to as “multivalued” or “multiple treatments” (Cameron and Trivedi, 2005; Imbens, 2004).

where  $(Y_{i1})$  is the outcome value for unit  $i$  estimated under the treatment state,  $(Y_{i0})$  is the outcome value estimated for unit  $i$  under the non-treatment state and  $N$  is the number of population. In this case, the impact of the treatment is estimated as the mean difference in outcome between the treated and the non-treated states for unit  $i$ .

The other relevant measure is the Average Treatment Effect on the Treated ( $ATT$ ), which is the average effect of the treatment on the treated group within the population,<sup>3</sup> and is referred to as the “conditional treatment effect” (Rosenbaum and Rubin, 1983):

$$ATT = E[Y_1 - Y_0 \mid T=1, x] \quad (2)$$

where  $(x)$  is a vector of covariates representing observable characteristics that influence the assignment to the treatment.<sup>4</sup> It should be noted that the observable covariates included in vector  $(x)$  are exogenous; in the sense of not being influenced by the potential outcome of interest ( $Y$ ) (Imbens, 2004). Hence, the treatment effect measure in equation (2) is estimated over the treated units in the subject population ( $N_T$ ) and is conditioned on a set of exogenous observable characteristics (such as age, gender, level of education, etc.).<sup>5</sup>

The major problem with identifying these measures is that only  $T=1$  or  $T=0$  can be observed for a single unit at a certain point in time ( $t$ ), where ( $t \neq 0$ ). This means that at a certain time ( $t$ ) a unit can be observed either in the treated state or the non-treated state, but not simultaneously in both (Cameron and Trivedi, 2005). One plausible solution to this problem would be to impute the missing data of the non-treated state for the treated units through the construction of a counterfactual<sup>6</sup> for  $(Y_{i0})$  using data collected from a group of units that possess similar observable characteristics at the baseline (i.e. at  $t=0$ ) as the treated group but are not exposed to the treatment.<sup>7</sup> These non-treated units are referred to as the control group (or the

---

<sup>3</sup> These measures ( $ATE$  and  $ATT$ ) can be estimated on the sample level through averaging by sample size. However, sample-level estimators cannot be used as estimators for the population-level (see Imbens (2004) for argument).

<sup>4</sup> These can be predictive variables, i.e. variables that are believed to influence the assignment to treatment.

<sup>5</sup> The discussion throughout this part is related to treatment selection on observables for a single (binary) treatment type. Selection on unobservables has other identification problems and follows different estimation methods (see Cameron and Trivedi (2005), Wooldridge (2002) or Heckman, Ichimura and Todd (1997) for a further discussion on this topic).

<sup>6</sup> A “counterfactual” means a hypothetical unobserved value (Cameron and Trivedi, 2005).

<sup>7</sup> In the econometric literature, the frameworks that are built on the idea of counterfactuals are traced to Roy’s potential outcome model following the work of Roy (1951) (Cameron and Trivedi, 2005).

comparison group (Dehejia and Wahba, 2002; Cameron and Trivedi, 2005)). Let  $(Y_{j0})$  denote the outcome of the  $j$ th element in the control subsample ( $N_C$ ). In this case, the treatment effect can be estimated as the mean difference between the potential outcomes of the treated unit ( $Y_{i1}$ ) and the control unit ( $Y_{j0}$ ).

### 1.1.2. Identification Assumptions

In order to correctly identify the treatment effect measures and obtain unbiased estimators, two key assumptions underlie the consistent estimation of the potential outcome variable ( $Y$ ) (Cameron and Trivedi, 2005; Imbens, 2004). The first assumption is known as the “overlap” assumption and is concerned with the joint distribution of the treatment state ( $T$ ) and the observable covariates ( $x$ ) (Imbens, 2004):

*Assumption 1 (the “overlap” assumption):*  $0 < \text{Prob} [T = 1 | x] < 1$

Assumption (1) implies that each sampled unit has an equal chance for treatment assignment and that the probability of receiving the treatment for every unit among the treated and control groups is strictly positive and non-zero. This assumption implies random selection across individuals and ensures that for every treated unit there is an existing comparable control unit (Heckman, Ichimura and Todd, 1997). This assumption also ensures the elimination of the selection bias problem that could undermine the treatment effect measure (Cameron and Trivedi, 2005). The selection bias arises from the existence of systematic differences between the treated and non-treated groups that render the mean outcome of the non-treated group an invalid approximate for the outcome of the treated units (Caliendo and Kopeining, 2008).

The second assumption, as stated by Rosenbaum and Rubin (1983), is the “ignorability of treatment assignment”:

*Assumption 2 (ignorability of treatment assignment):*  $Y_1, Y_0 \perp\!\!\!\perp T | x$



This assumption is also known in the literature as the “conditional independence assumption” or the “unconfoundedness” assumption (Cameron and Trivedi, 2005; Imbens, 2004; Wooldridge, 2002). According to this assumption, the assignment to treatment is entirely based on observable characteristics, and there is no observable covariate that induces both outcome ( $Y$ ) and treatment assignment ( $T$ ) not explicitly accounted for in the vector ( $x$ ) (Caliendo and Kopeining, 2008; Tembo et al, 2014). This implies that the potential treatment outcomes ( $Y$ ), after controlling for ( $x$ ), are statistically independent of treatment assignment ( $T$ ) (i.e. ( $T$ ) is exogenous in ( $Y$ )’s estimation equation, and there is no omitted variable bias) (Cameron and Trivedi, 2005).

### 1.1.3. Observational Data vs. Randomized Control Experiments and the Role of Propensity Score

In practice, the randomized control experiments are considered the optimal method for estimating impacts of treatments on subject populations (Austin, 2011; Banerjee and Duflo, 2009; Heckman and Smith, 1995; Hill et al, 2004; Smith and Todd, 2005). This is because a properly conducted randomized experiment allocates treatment ( $T$ ) to subject units randomly over baseline characteristics ( $x$ ). This ensures that treated and control groups have similar distributions for baseline characteristics, and that the sole difference between the two groups would be the exposure to the treatment. In this case, the systematic differences in the resultant outcome variable between the treated and control groups can directly be attributed to the treatment effect (Austin, 2011). Moreover, since randomization is over ( $Y$ ) and ( $x$ ) (Cameron and Trivedi, 2005), the experimental design satisfies the strong “ignorability of treatment assignment” assumption for which:

*Assumption 3 (strong ignorability of treatment assignment):*  $Y_1, Y_0 \perp\!\!\!\perp T$

The stronger assumption (3)<sup>8</sup> implies that  $E[Y_1 | T=1] = E[Y_1]$  and  $E[Y_0 | T=0] = E[Y_0]$ . It follows that  $ATT$  can be an unbiased estimate for  $ATE$  (Austin, 2011; Wooldridge, 2002).<sup>9</sup>

---

<sup>8</sup> Assumption (3) is a translation for the satisfaction of both assumptions (1) and (2) (Rosenbaum and Rubin, 1983).

<sup>9</sup> However, randomized control experiments may suffer from other imbalances that could bias the experiment’s test results such as the chance bias or the substitution bias (Cameron and Trivedi, 2005; Roberts and Torgerson, 1999).

However, sometimes observational data are used to draw inferences about causal connections between treatments (or policies) and outcomes of interest.<sup>10</sup> When using observational data to evaluate the impacts of treatments, the simple direct comparison of outcome variables between treated and control units would lead to misleading conclusions, because baseline characteristics between treated and non-treated units are systematically different (Rosenbaum and Rubin, 1983). These systematic differences between the units in observational datasets can be traced to the non-random process through which the treatment was assigned to subject units. An example of such data is data on programs for which participation is through a self-selection process. The estimated treatment measure in this case would be prone to the selection bias problem. The Selection bias (or evaluation bias) comprises the difference between the outcomes of the control units [i.e.  $E(Y_0 | T=0)$ ] and the desired counterfactual outcomes for the treated units [i.e.  $E(Y_0 | T=1)$ ] (Smith and Todd, 2005). Therefore, the differences between the units must first be accounted for in order to isolate the causal treatment effect on the treated and obtain unbiased estimates (Rubin, 1997).

Consequently, a method was needed to simulate the design of a randomized control experiment and create a balance in the distribution of baseline covariates between treated and control units, i.e. satisfy the overlap condition, and ensure the independence between the potential outcome ( $Y$ ) and treatment assignment ( $T$ ) given the observed characteristics, i.e. satisfy the unconfoundedness assumption (Austin, 2011; Imbens, 2004).

The propensity score was one of the methods proposed to adjust for the imbalances between treated and control units.<sup>11</sup> Rosenbaum and Rubin (1983) developed the theoretical foundation for propensity score methods for observational data. The propensity score is a “scalar function” (one-dimensional) that represents the probability (propensity) of receiving the treatment conditional on a set of observed baseline characteristics ( $x$ ) (Rosenbaum and Rubin, 1983):

$$P(x) = \Pr [T = 1 | x] = E[T | x] \quad ; \quad 0 < P(x) < 1 \quad (3)$$

---

<sup>10</sup> This is because sometimes setting an experiment may not be a feasible option due to cost, social, geographic or policy limitations.

<sup>11</sup> This method especially used when selection is on observables and the target is to estimate *ATT* (Cameron and Trivedi, 2005) and the available pool of control units is quite large (Imbens, 2004). Other methods include nonparametric regression and matching on covariates (Imbens, 2004).

The observable characteristics in vector ( $x$ ) in equation (3) are assumed to be comparable for the treated and control groups, i.e. have closely similar distributions in order to satisfy the overlap condition.

The main assumption that underlies the propensity score methods is that, conditional on the propensity score  $P(x)$ , the potential treatment outcome ( $Y$ ) is assumed independent of treatment assignment ( $T$ ):

*Assumption 4 (independence of treatment assignment):*  $Y_1, Y_0 \perp\!\!\!\perp T \mid P(x)$

The reasoning behind assumption (4) that was provided by Rosenbaum and Rubin (1983) is as follows: as  $P(x)$  is a subset function in ( $x$ ), then if ( $T$ ) is assumed independent of ( $x$ ), it follows that ( $T$ ) is also independent of  $P(x)$  when conditioning on  $P(x)$ .

In addition, the average treatment measure ( $ATT$ ) is conditional on treatment assignment ( $T$ ). Hence, using a score that represents the probability of receiving the treatment conditional on a set of observable covariates ( $x$ ) can minimize the imbalance in baseline observables between treated and control groups, and reduces the estimator's bias that is due to differences in observable covariates (Cameron and Trivedi, 2005; Dehejia and Wahba, 2002; Heckman, Ichimura and Todd, 1997; Imbens, 2004). In this case, the differences in the covariates can be adjusted through adjusting for the differences in the propensity scores between the treated and control units (Imbens, 2004; Rosenbaum and Rubin, 1983). This feature represents one of the strength points of propensity score methods over the regression-based methods; Instead of adjusting for the differences in all covariates as in the regression-based methods,<sup>12</sup> the propensity score methods adjust only for the differences in the covariates that represent the probability of being treated. Moreover, using a subset of the covariates can partly solve the problem of “high dimensionality” of the covariates<sup>13</sup> that may occur when the pairing of the control units to treated

---

<sup>12</sup> Another down side of regression-based methods is that they are prone to the problem of the misspecification of the regression function (Imbens, 2004). The next section provides a lengthier comparison between propensity score methods and regression-based methods.

<sup>13</sup> The “high dimensionality” problem occurs when the covariates have large dimensional spaces (for example, if the covariates in vector ( $x$ ) have two dimensions or more). This may lead to the existence of units with outlying values of covariates. In this case, satisfying the overlap assumption can be very difficult and may lead to concluding that the average treatment effect using the dataset at hand is inestimable.

units is based on the whole set of possible covariates (Becker and Ichino, 2002; Dehejia and Wahba, 2002; Heckman, Ichimura and Todd, 1997; Imbens, 2004).

It is to be noted from the conditional independence assumption that propensity score methods address the correlation between the covariates in ( $x$ ) and the outcome variable ( $Y$ ) indirectly (Imbens, 2004). The independence between the two sets is achieved through the choice of covariates. Only variables that are simultaneously correlated with the outcome variable and the treatment assignment should be included in the analysis (Caliendo and Kopeining, 2008; Imbens, 2004). However, the level of correlation preferable between the covariates included in the analysis and outcome or treatment is debatable. Weakly correlated covariates are advised to be discarded from the analysis, especially for finite samples, as they may increase the mean squared standard error and the bias of the treatment effect measure; i.e. reduce the precision of the model (Augurzky and Schmidt, 2001; Imbens, 2004). On the other hand, Brookhart et al. (2006) argue, based on experimental simulations, that, for finite samples, the inclusion of covariates that are strongly related to outcome, whether related to treatment assignment or not, increases the precision of the treatment effect measure without increasing bias. The covariates chosen for the analysis should also satisfy the condition of being unaffected by the participation or the anticipation of participation in the treatment program. Hence, the covariates should be measured before participation (at the baseline) and should be considered fixed over time (Caliendo and Kopeining, 2008).

## **1.2. Practical Advantages of Propensity Score Methods<sup>14</sup>**

One of the practical advantages of using propensity score methods is, as mentioned earlier, their ability to separate the study design from the study analysis, allowing the researcher to create an experimental environment that closely resembles a randomized control experiment (Austin, 2011).<sup>15</sup> In addition, the simplicity of calculation and the ability of drawing direct inference from

---

<sup>14</sup> Propensity score methods have, as other statistical analysis techniques, their own limitations and shortcomings (Abadie and Imbens, 2006; Heckman, Ichimura and Todd, 1998; Imbens, 2004; Rubin, 1997). However, this part is primarily concerned with the practical advantages that yielded these methods widely acknowledged and applied for evaluations of average treatment effects.

<sup>15</sup> Dehejia and Wahba (1998, 2002) provide an example and discussion on how propensity score matching methods permit the creation of an environment similar to experimental trials and provide highly comparable results.

the propensity score estimators had made these methods appealing for policy-relevant treatment assessments (Cameron and Trivedi, 2005; Zanutto, 2006).

Although the debate on favoring one method to the other is yet ongoing, the attractiveness of propensity score methods over regression-based methods lays, in principle, in their provision of a more flexible and straightforward approach for analyzing the causal effects of treatments in observational studies especially in the case of selection on observables (Dehejia and Wahba, 1997; Imbens, 2004; Rubin, 1997). Estimation of treatment effect using regression-based methods is considered highly sensitive to the correct specification of the functional form that defines the relationship between the outcome variable and the covariates, because regression methods rely on the specified functional form to extrapolate (predict) the impact estimate (Dehejia and Wahba, 1998; Imbens, 2004, 2014; Rubin, 1997; Zanutto, 2006). This is particularly relevant when the distributions of the covariates between the two treatment groups are very different (Imbens, 2004; Zanutto, 2006). As a result, the misspecification of the regression function may lead to biased treatment effect estimators (Hill, Reiter and Zanutto, 2004; Rubin, 1997). On the other hand, propensity score methods rely on much simpler assumptions and estimate the treatment effect through the direct comparison of average outcomes between the treated and control groups (Rosenbaum and Rubin, 1983). As long as the two treatment groups are balanced based on their pre-treatment covariates, propensity score methods require no further adjustments to produce unbiased and credible estimators (Imbens, 2004; Rubin, 1997).

Even though, regression methods may evade the previously stated problem by adjusting for the imbalance in pre-treatment covariates between the two treatment groups through adjusting the specifications of models by including more covariates or adding higher order or interactive terms (Dehejia and Wahba, 1997; Härdle, 1995). However, regression methods need to control for all covariates (or otherwise may suffer from the omitted-variables bias), and this may be very problematic in case of a large number of covariates or if the covariates are of a multi-dimensional nature (Dehejia and Wahba, 1997; Zanutto, 2006). Propensity score methods can overcome this problem by controlling only for the covariates that simultaneously affect the outcome and treatment assignment, which are those included in estimating the conditional

probability of being treated; i.e. the propensity score (Dehejia and Wahba, 1998; Heckman, Ichimura and Todd, 1997; Imbens, 2004; Rosenbaum and Rubin, 1983).

Another major drawback to the practical application of regression-based methods is their inability to recognize the heterogeneous nature of the treatment effect in the population even after observable variables are controlled for (DiPrete and Gangal, 2004). The standard regression methods implicitly assume that causal effects are constant either over the population or within subgroups of the population. Propensity score methods can overcome this limitation through their capacity for estimating the average of the causal effect of interest over the population distributions (DiPrete and Gangal, 2004; Hujer et al, 2003).

Propensity score methods can also easily detect the lack of overlap between the treated and control groups, and permit the comparability even if few control units are available for comparison with the treated units (Dehejia and Wahba, 2002).<sup>16</sup> The regression-based methods function poorly and their results are misleading when the overlap in the distribution of the covariates between the two treatment groups is limited (Imbens, 2004; Rubin, 1997). In this regard, the propensity score methods can better handle the problem of limited overlap between the treated and control groups in comparison to the standard regression-based methods.

### **1.3. Propensity Score in Matching Methods**

Matching on propensity score represents one of the methods that use the propensity score function for constructing an experimental setting out of a non-random non-experimental one. It involves comparing the potential outcomes of treated and control units through matching for each treated unit, based on the observable characteristics included in vector ( $x$ ) and are represented by the propensity score value  $P(x)$ , a unit within the control group (Tembo et al, 2014; Abadie and Imbens, 2016; Dehejia and Wahba, 2002). The control unit that is paired to the

---

<sup>16</sup> This is especially true for matching methods with replacement. Matching with replacement allows each treated unit to be matched to the control unit that has the closest propensity score even if this control unit was previously matched. In addition to providing a better match for the distribution of the propensity scores (i.e. the distribution of the baseline characteristics) of the treated units in case of small control sample size, matching with replacement reduces the bias in the estimates of treatment impact that occurs as the distance between the matched pairs increase. On the other hand, as the sample size increases the bias of the matching estimators for the average treatment effect disappears (Imbens, 2004) because of the availability of more close matches within the sampled pool and, in this case, matching without replacement could be a viable option. (See Dehejia and Wahba (2002) for an example and Cameron and Trivedi (2005), Dehejia and Wahba (2002) and Rosenbaum and Rubin (1983) for further discussion on different matching methods).

treated unit is the one with the  $P(x)$  value that is the closest to the treated unit's  $P(x)$  value. In this case,  $ATT$  is estimated as the weighted average of the difference in potential outcome between the treated units and their matched control units:

$$ATT = \frac{1}{N_T} \sum_{i \in N_T} [Y_{i1} - \sum_{j \in N_C} w_{ij} Y_{j0}] \quad ; \quad 0 < w_{ij} \leq 1 \quad (4)$$

where  $(j)$  represents the set of matched control units that corresponds to the treated unit  $(i)$  and  $(w_{ij})$  is the weight assigned to the control units relative to the treated units. It should be noted here that the formula in equation (4) represents the general formula for estimating  $ATT$  under the different methods of matching (Cameron and Trivedi, 2005).

#### 1.4. Model Specification

The propensity score  $P(x)$  is usually calculated through a discrete choice model, for example a probit or logit model. A Probit model is employed here to estimate the conditional probability of being selected for the program (the propensity scores). The Probit model has the following specification:

$$P(x) = \text{Prob} [T = 1 \mid x] = \Phi (\alpha + \beta x + \varepsilon) \quad (5)$$

where  $\Phi$  is the standard normal cumulative distribution function (CDF),  $\alpha$  and  $\beta$  are the model's estimated parameters,  $x$  is the vector of observed characteristics that represent the selection criteria for participation in the program and  $\varepsilon$  is the error term (Cameron and Trivedi, 2005; Tembo et al, 2014).

The matching algorithm follows Dehejia and Wahba (2002) radius caliper matching method and is conducted with replacement.<sup>17</sup> The idea of this method is that all of the comparison units within a specified propensity score range (caliper) in the neighborhood of the treated unit are used as matches and the composition of the counterfactual outcome consists of

---

<sup>17</sup> Other matching methods include the nearest-neighbor method (and its variations), stratification matching and radius matching. These matching estimators differ in their choices for matches and the weight given to the matched comparison units relative to their paired treated units (see Cameron and Trivedi (2005), Imbens (2004) or Smith and Todd (2005) for a detailed discussion on different matching methods).

the mean outcome of all the comparison units within the caliper (Dehejia and Wahba, 2002; Smith and Todd, 2005). Under this method, the treatment unit is weighted by 1 and the matched control units have equal weights.<sup>18</sup> The choice of the matching algorithm depends on various considerations and trade-offs (Baser, 2006; Caliendo and Kopeining, 2008; Heinrich et al, 2010; Imbens, 2004; Smith and Todd, 2005). One of these considerations is the reduction of the mean bias between the treated and control groups. For this model, radius caliper method is found to reduce the mean bias between the two groups to 5.4% after matching. The tolerance level for the caliper (or the caliper bandwidth) is estimated, as suggested by Rosenbaum and Rubin (1985), as one quarter of the total value of the standard deviation of the estimated propensity scores (Rosenbaum and Rubin, 1985).<sup>19</sup>

The standard errors are estimated, following most empirical literature, using the bootstrap method with 250 replications.

### **1.5. Propensity Score in Regression Adjustment: Propensity Score Weighting (Odds-Weighted Regression)**

If the propensity score model is correctly specified, propensity score methods can yield unbiased average treatment effect estimators that are more consistent compared to regression-based methods,<sup>20</sup> however, these estimators are not considered fully efficient (i.e. have the smallest variance) (Abadie and Imbens, 2002, 2006, 2011; Heckman, Ichimura and Todd, 1998; Imbens, 2004). Mixing propensity scores with other methods had been proposed for the purpose of attaining fully efficient estimators and arriving at a more robust inference (Hahn, 1998; Hirano and Imbens, 2002; Hirano, Imbens and Ridder, 2003; Imbens, 2004).

In this context, Hirano, Imbens and Ridder (2003) propose a semi-parametric method that uses the non-parametrically estimated propensity scores as weights in a regression model. They suggest that it may be possible to achieve full efficiency of the treatment effect estimator and

---

<sup>18</sup> The method applied does not weight the matched control units within the radius according to their distance from the treated unit as proposed by Lechner, Miquel and Wunsch (2011).

<sup>19</sup> Although there has not been an empirical consensus over the proper caliper bandwidth, Rosenbaum and Rubin's (1985) proposed approach is a widely accepted and recommended choice of caliper (Lunt, 2013; Pan and Bai, 2015).

<sup>20</sup> Consistency for propensity estimators is achieved if the assignment to treatment is exogenous in the sense of the inexistence of covariates that affect treatment assignment and outcome simultaneously and are not accounted for in the model (DiPrete and Gangl, 2004).



reduce the estimator's variance through weighting by the inverse of the estimated propensity scores.<sup>21</sup> It is argued that combining propensity scores with regression can remove the remaining bias in the matching estimator that is due to the remaining imbalance between the treated and control groups<sup>22</sup> leading to a more robust inference (Hirano, Imbens and Ridder, 2003; Imbens, 2004). This is because weighting the observations using the propensity score permits the exploitation of the control observations to adjust indirectly for the imbalances in the sampling of the covariates and create a balance between the treated and control units in the weighted sample (Imbens, 2004). In addition, the precision of the obtained estimator can be improved through the addition of covariates to the regression function (Imbens, 2004).

An advantage of the estimator proposed by Hirano, Imbens and Ridder (2003) in comparison to other “mixed” estimators is that it only requires the non-parametric estimation of the propensity scores and does not require the estimation of an initial consistent estimator of the full population parameters (Hirano, Imbens and Ridder, 2003).

Under the propensity score weighting approach proposed by Hirano, Imbens and Ridder (2003), the estimated impact is a “weighted average of treatment effect,” where weights are the inverse of the conditional probability of being treated, i.e. the inverse of the propensity scores. It follows that the treated unit is weighted by 1 and the matched unit from the control group is weighted by the odds ratio  $[P(x)/1-P(x)]$  (Hirano, Imbens and Ridder, 2003; Tembo et al, 2014) in the regression model:

$$\ln(Y) = \beta_0 + \beta_1 T + \varepsilon \quad (6)$$

where  $\ln(Y)$  is the natural logarithm of per capita consumption expenditure of food, non-food or both,  $\beta_1$  is the estimated impact coefficient and  $\varepsilon$  is a zero-mean error term.

---

<sup>21</sup> In Hirano, Imbens and Ridder's (2003) approach the weights are normalized to unity in order to achieve efficiency of the estimator (Abadie and Imbens, 2002; Imbens, 2004).

<sup>22</sup> The remaining bias arise from the fact that the covariates of the treated unit and its matched control unit, which are represented by the propensity score, are close through the matching process but are not exactly identical (Imbens, 2004).

## 1.6. The Heterogeneous Impact

In order to further examine the extent of the program's impact among the treatment groups, two categories are formed within the beneficiary households based on an assets-wealth index that is constructed using the principal-component factor analysis (PCF). These two categories are: (1) the assets-poor households with an assets-wealth index of zero or less and (2) the relatively assets-nonpoor households with an assets-wealth index above zero. The disaggregation of the impact by these two sub-groups would help identify whether or not Takaful and Karamah social schemes had benefited the poorer of the ultra-poor the most (Tembo et al, 2014). To estimate the heterogeneous impact, equation (6) is extended as follows:

$$\ln(Y) = \beta_0 + \beta_1 T + \beta_2 T * D + \varepsilon \quad (7)$$

where  $D$  is a dummy variable that takes the value of (1) if the household is categorized as an assets-poor household and (0) otherwise. In this case  $\beta_2$  is a measure of the additional impact that the assets-poor households would accrue from the program relative to their assets-nonpoor counterparts (Tembo et al, 2014). It should be noted that equation (7) is also estimated following the odds-weighted regression approach as is equation (6).

## 1.7. Data and Variables

The empirical part of this study applies the methodology on data at the household level. The data is extracted from the cross-sectional Household Income, Expenditure and Consumption Survey (HIECS) 2017-2018. The HIECS dataset is collected and disseminated by the Central Agency for Public Mobilization and Statistics (CAPMAS) and the survey is conducted periodically every other year and covers the entire 26 Egyptian governorates. The used dataset represents 50% of the 2017-2018 HIECS dataset and comprises of 12,485 households (50% of the total 26,000 surveyed households) of which 882 households host at least a beneficiary of Takaful and Karamah social schemes (7.064% of the sampled households) and represent 4,878 individual beneficiary family member. At the time point the data were collected, the monthly transfer amount of the "Takaful" program was LE 325 per household member for a maximum of three

children.<sup>23</sup> A household that hosts children attending school was getting an additional LE 60 per primary student, LE 80 per preparatory student and LE 100 per secondary student for a maximum of three students. The “Karamah” program supplemented a monthly payment of LE 350 per individual suffering from disability or of age above 65 years. The official minimum wage then was LE 1200 per month.

The focus on estimating the impact of participation on the subpopulation that is likely to participate in the program (i.e. estimating the ATT) for this study had allowed the use of the HIECS data. This is because, contrary to the estimation of ATE; which requires that the combinations of characteristics in the control group are also observed in the treated group, for estimating ATT it is sufficient to ensure the existence of potential matches within the control group (Bryson, Dorsett, and Purdon, 2002; Caliendo and Kopeining, 2008). Moreover, the assignment to treatment in Takaful and Karamah social schemes is considered to follow the “self-selection” feature, which justifies the use of the non-experimental methods.

The rationale behind using the non-experimental method of propensity score matching is its capacity in permitting the comparison between the households that are participants of Takaful and Karamah social schemes with the households belonging to the same poverty brackets and are non-participants.

For the Probit model, used to estimate the conditional probability of being selected for the program (the propensity scores), the vector of exogenous covariates ( $x$ ) includes variables pertaining to the selection criteria for Takaful and Karamah social schemes (hosting school-age children or being an elder above 60 years of age without formal pension coverage). The rest of the variables are control variables representing predictive characteristics assumed to influence the participation in the program. The chosen observed characteristics are believed to influence both the outcome of interest and participation in the program and, hence, capture part of the potential endogeneity problem between the outcome and participation. They are also considered stable over time and not likely to be affected by the participation status.

The construction of the treated and control groups follows, conceptually to a certain extent, the work of Dehejia and Wahba (1997, 1998 and 2002), which they built on Lalonde’s (1986) work, in drawing non-experimental observations from survey datasets and use these

---

<sup>23</sup> The Egyptian Ministry of Social Solidarity (MoSS), 2017.

observations to generate the two comparison groups.<sup>24</sup> Consequently, the treated group in the study is defined as the poor households that are non-beneficiaries of any formal pension scheme (government pension, social insurance pension or union pension) and within which at least a member is a beneficiary of Takaful and Karamah social schemes (comprises of 142 households). On the other hand, the control group is defined as the poor households that are neither beneficiaries of any formal pension scheme nor Takaful and Karamah social schemes (comprises of 448 households). The poverty line used to define the poor in this study is the lower poverty line (the regional-level lower poverty lines in 2017-2018, according to the CAPMAS's HIECS, were accounted for in the construction of the comparison groups). The subsamples of the two comparison groups were constructed as mentioned in order to isolate the effect of Takaful and Karamah social schemes on extremely poor beneficiary households and, in this form, the scheme can resemble a cash transfer program that is directed to the poorest of the poor (or ultra-poor) households.

The asset-wealth index used in estimating the heterogeneous impact is constructed following Filmer and Pritchett (2001) principal component analysis (PCA) approach. The variables included in the construction of the assets-wealth index are provided in a following appendix.

## **1.8. Empirical Results<sup>25</sup> and Analysis**

Table (1) depicts some of the characteristics of the treatment sample (Takaful and Karamah social cash transfer schemes' target households) as well as the sample's breakdown into the two comparison groups.

---

<sup>24</sup> The advantage of the approach followed in the study at hand in constructing the treated and comparison groups compared to the work of Dehejia and Wahba (1997, 1998 and 2002) is that both the treated and comparison groups are from the same data source, which is the HIECS 2017-2018. According to the findings of Smith and Todd (2005), this criterion is crucial for a reliable estimation strategy and for arriving at high-performance estimators (Smith and Todd, 2005).

<sup>25</sup> The calculations and the estimations presented in this part are conducted using the STATA software package. For the propensity score matching results, "psmatch2" and "pstest" (Leuven and Sianesi, 2003) STATA routines are utilized.

**Table 1. Characteristics of the treated and control groups: sample means**

	Treatment Sample	Sub-Samples		
		Treated	Control	Difference (Control-Treated)
Female-headed households (%)	8.305 (0.011)	6.338 (0.021)	8.929 (0.013)	2.591 (0.027)
Age of household head (years)	43.283 (0.420)	42.014 (0.762)	43.685 (0.497)	1.671* (0.982)
Education level of household head	1.453 (0.058)	1.415 (0.120)	1.464 (0.067)	0.049 (0.137)
Widow-headed households (%)	3.559 (0.008)	2.817 (0.014)	3.795 (0.009)	0.978 (0.018)
Divorcee-headed households (%)	0.678 (0.003)	0.704 (0.007)	0.670 (0.004)	-0.035 (0.008)
Effective dependency ratio	3.490 (0.081)	3.908 (0.150)	3.357 (0.095)	-0.551*** (0.188)
Number of children 14 years or younger	2.949 (0.060)	3.366 (0.109)	2.817 (0.070)	-0.549*** (0.139)
Number of male members 15-30 years	0.663 (0.040)	0.634 (0.076)	0.672 (0.046)	0.038 (0.093)
Number of female members 15-30 years	0.690 (0.030)	0.655 (0.057)	0.701 (0.035)	0.046 (0.071)
Number of male members 31-45 years	0.549 (0.021)	0.585 (0.042)	0.538 (0.024)	-0.047 (0.048)
Number of female members 31-45 years	0.622 (0.021)	0.676 (0.042)	0.605 (0.025)	-0.071 (0.050)
Number of male members 46-60 years	0.283 (0.019)	0.282 (0.039)	0.283 (0.021)	0.002 (0.044)
Number of female members 46-60 years	0.166 (0.016)	0.092 (0.024)	0.190 (0.020)	0.098*** (0.037)
Number of male members (61+) years	0.059 (0.010)	0.021 (0.012)	0.071 (0.012)	0.050** (0.023)
Number of female members (61+) years	0.066 (0.010)	0.085 (0.023)	0.060 (0.011)	-0.024 (0.024)
Orphan children 18 years or younger (%)	2.203 (0.006)	1.408 (0.010)	2.455 (0.007)	1.047 (0.014)
Asset wealth index†	-0.869 (0.027)	-1.065 (0.045)	-0.806 (0.032)	0.258*** (0.063)
Agricultural assets ownership‡ (%)	19.492 (0.016)	26.056 (0.037)	17.411 (0.018)	-8.646** (0.038)
Per capita income (‘000 EGP)	5852.579 (67.537)	5499.306 (105.511)	5964.554 (81.766)	465.248*** (156.950)
Per capita consumption expenditure (‘000 EGP)	4822.349 (33.932)	4679.079 (66.456)	4867.760 (39.208)	188.681** (79.060)

Level of significance: \*=10 percent, \*\*=5 percent, \*\*\*=1 percent

Standard errors in parenthesis.

†A unit-less wealth index estimated using the principal component analysis (PCA) approach.

‡Agricultural assets include owned or rented cultivated lands, farm animals and agricultural machinery that generate annual income for the household.

In comparison to the control units, the treated households tend to be relatively of larger composition (have relatively higher effective dependency ratio and host higher number of children of 14 years and younger and higher number of elderly members especially females), asset-poorer households with household heads of low levels of education. They also have significantly lower levels of per capita income and consumption expenditure (averaging about LE 5499.306 and 4679.079, respectively). Households hosting elder members, especially females, are more likely to benefit from Takaful and Karamah social schemes. A sizable proportion of the treated households are in possession of agricultural assets. This could be an indication that most beneficiary households reside in rural regions (71.356% of the total treatment sample and 80.986% of the treated subsample belong to rural areas). However, this could also signal a misspecification in the programs' targeting criteria.

The parameter estimates and the marginal effects (the marginal probabilities of the participation decision) from the fitted Probit model for participant and non-participant households are presented in Table (2).

**Table 2. Probit model results for household participation**

Variable description	Parameter estimate	Marginal effect
Constant	-1.730 (0.297)	
1. Education level of most educated member	-0.114** (0.058)	-0.033
2. Children (14 years or younger)	0.160*** (0.043)	0.046
3. Male elders (61+ years old) dummy	-0.643* (0.328)	-0.144
4. Female elders (61+ years old) dummy	0.320 (0.240)	0.103
Joint significance of location dummies		46.140***
Goodness-of-fit Chi-square		147.330*
Log likelihood		-279.177
Cragg and Uhler's Pseudo $R^2$		0.196
Observed probability of participation		0.250
Predicted probability of participation		0.213
Number of observations †		568

Level of significance: \*=10 percent, \*\*=5 percent, \*\*\*=1 percent

Standard errors in parenthesis.

† The total number of treated and control units included in the estimation.

It is to be noted that the probability of participation in Takaful and Karamah social schemes is significantly correlated with hosting children (14 years of age or younger). Hosting a female elder also increases the household's chance of participation in the schemes by 10.30% relative to its counterparts; however, this positive correlation is statistically insignificant. The specific geographical location of the household is also a determining factor for the decision of participation in the program as indicated by the significant joint test of location dummies.

It should be highlighted that the probit model represents only the selection process from among the eligible group and does not constitute an assessment process of the targeting effectiveness of the programs. As stated earlier, the vector of exogenous covariates ( $x$ ) includes variables pertaining to the selection criteria for Takaful and Karamah social schemes and the sampled households were drawn from the program beneficiaries and those who are non-beneficiaries but are most likely to qualify for participation based on similar eligibility criteria.

### 1.8.1. Impact Estimation

Tables (4) through (8) present the impact that Takaful and Karamah social schemes have on the per capita consumption expenditure for the treated households relative to their counterparts in the control group. Columns (1) and (2) estimate the average impact realized by the entire group of beneficiary households compared to their non-beneficiary counterparts, while column (3) disaggregates the impact realized by beneficiary households by their wealth status.

**Table 4. Propensity score matching and odds-weighted regression estimates of the impact of Takaful and Karamah social schemes on per capita consumption expenditure**

Variable	Variable description	Propensity score matching	Odds-weighted regression	
			Overall	Heterogeneous
		(1)	(2)	(3)
Constant	Treatment indicator (1=participant)	-	8.4463***	8.4463***
$T$	Treatment indicator (1=participant)	-0.0250 (0.0210)	-0.0118 (0.0238)	0.0957** (0.0446)
$T \times D$	$T$ interacted with wealth dummy ( $D$ , 1=poor)	-	-	-0.1090** (0.0439)
Goodness of fit (F statistic)		-	0.2500	3.0800**
Number of observations		568	568	568

Level of significance: \*=10 percent, \*\*=5 percent, \*\*\*=1 percent

Dependent variable: natural log of per capita consumption expenditure. In parenthesis: robust standard errors, except for PSM for which bootstrapped standard errors with 250 replications are reported.

According to the results in table (4), the average impact of Takaful and Karamah social schemes on total per capita consumption expenditure was negative and ranged between -1.18% and -2.50% for beneficiary households, taking into consideration that the results for the propensity score matching and the overall PSW are not statistically significant. When disaggregating the treated group by wealth index, the reduction in total per capita consumption expenditure was particular for asset-poorer households compared to their wealthier counterparts, with a statistically significant reduction of 10.90% between the two groups.

This reduction in total per capita consumption expenditure for treated households, as is illustrated in the following impact-estimation results, could be attributed to the observed reduction in spending on the food component due to the support services provided to beneficiary households through the food ration cards. According to the dataset, the treated households receive a food support in the form of bread and food rations that amounts to 15.78% of their total food consumption. The reduction in total spending on food items could also be translated into the direction of household resources toward other investment options, as is observed in the increase in total spending on the non-food component. Nevertheless, this reduction could also be attributed to the high inflation rate, which had reached 20.90% in the reference survey year,<sup>26</sup> and the insignificance of the transfers' amounts in raising the consumption expenditures of poor households.

**Table 5. Propensity score matching and odds-weighted regression estimates of the impact of Takaful and Karamah social schemes on per capita food consumption expenditure**

Variable	Variable description	Propensity score matching	Odds-weighted regression	
			Overall	Heterogeneous
		(1)	(2)	(3)
Constant	Treatment indicator (1=participant)	-	7.6352***	7.6352***
<i>T</i>	Treatment indicator (1=participant)	-0.0757** (0.0340)	-0.0664** (0.0317)	0.1359 (0.1723)
<i>T</i> x <i>D</i>	<i>T</i> interacted with wealth dummy ( <i>D</i> , 1=poor)	-	-	-0.2052 (0.1728)
Goodness of fit (F statistic)		-	4.380**	2.8300*
Number of observations		568	568	568

Level of significance: \*=10 percent, \*\*=5 percent, \*\*\*=1 percent

Dependent variable: natural log of per capita food consumption expenditure. In parenthesis: robust standard errors, except for PSM for which bootstrapped standard errors with 250 replications are reported.

<sup>26</sup> The inflation rate was, in June/July 2017/2018, 20.90% (2010=100) (based on consumer price index (urban population, January 2010=100). Source: Central Agency for Public Mobilization and Statistics (CAPMAS) monthly bulletin, August 2018).



In tables (5) and (6), the per capita consumption expenditure is disaggregated into its food and non-food components. The results for the food component (in table (5)) exhibit a negative impact, ranging between -7.57% and -6.64%, and the impact is statistically significant for the propensity score model and the overall PSW model. The asset-poorer households realized the largest impact here, with a difference of 20.52% compared to asset-wealthier households. On the other hand, concerning the non-food component (in table (6)), the participation in the schemes seemed to raise the non-food consumption expenditure for the asset-wealthier households by 11.18%, with a difference of 4.09% compared to their poorer counterparts. Nevertheless, the impact-results for the non-food component are statistically insignificant across the three models.

**Table 6. Propensity score matching and odds-weighted regression estimates of the impact of Takaful and Karamah social schemes on per capita non-food consumption expenditure**

Variable	Variable description	Propensity score matching	Odds-weighted regression	
			Overall	Heterogeneous
Constant	Treatment indicator (1=participant)	-	7.7814***	7.7814***
<i>T</i>	Treatment indicator (1=participant)	0.0284 (0.0319)	0.0714 (0.0648)	0.1118 (0.0958)
<i>T x D</i>	<i>T</i> interacted with wealth dummy ( <i>D</i> , 1=poor)	-	-	-0.0409 (0.0784)
Goodness of fit (F statistic)		-	1.210	0.8100
Number of observations		565	565	565

Level of significance: \*=10 percent, \*\*=5 percent, \*\*\*=1 percent

Dependent variable: natural log of per capita non-food consumption expenditure. In parenthesis: robust standard errors, except for PSM for which bootstrapped standard errors with 250 replications are reported.

In tables (7) and (8), the per capita consumption expenditures on healthcare services and education for treated households are inspected against their counterparts. It is to be noted that the participation in the schemes had contributed in raising the total spending on healthcare services and education for treated households. The increase in healthcare spending was particular for asset-poorer households (statistically significant result), while the increase in spending on education was particular for asset-wealthier households (statistically insignificant result).

**Table 7. Propensity score matching and odds-weighted regression estimates of the impact of Takaful and Karamah social schemes on per capita consumption expenditure on healthcare services**

Variable	Variable description	Propensity score matching	Odds-weighted regression	
			Overall	Heterogeneous
Constant	Treatment indicator (1=participant)	-	5.3148***	5.3148***
<i>T</i>	Treatment indicator (1=participant)	0.1633 (0.1108)	0.1980* (0.1118)	-0.6280 (0.4102)
<i>T</i> x <i>D</i>	<i>T</i> interacted with wealth dummy ( <i>D</i> , 1=poor)	-	-	0.8380** (0.4121)
Goodness of fit (F statistic)		-	3.1400*	3.3100**
Number of observations		560	560	560

Level of significance: \*=10 percent, \*\*=5 percent, \*\*\*=1 percent

Dependent variable: natural log of per capita non-food consumption expenditure. In parenthesis: robust standard errors, except for PSM for which bootstrapped standard errors with 250 replications are reported.

**Table 8. Propensity score matching and odds-weighted regression estimates of the impact of Takaful and Karamah social schemes on per capita consumption expenditure on education**

Variable	Variable description	Propensity score matching	Odds-weighted regression	
			Overall	Heterogeneous
Constant	Treatment indicator (1=participant)	-	4.8974***	4.8974***
<i>T</i>	Treatment indicator (1=participant)	0.1318 (0.1080)	0.1133 (0.1038)	0.9691 (0.6298)
<i>T</i> x <i>D</i>	<i>T</i> interacted with wealth dummy ( <i>D</i> , 1=poor)	-	-	-0.8689 (0.6311)
Goodness of fit (F statistic)		-	1.1900	1.5500
Number of observations		504	504	504

Level of significance: \*=10 percent, \*\*=5 percent, \*\*\*=1 percent

Dependent variable: natural log of per capita non-food consumption expenditure. In parenthesis: robust standard errors, except for PSM for which bootstrapped standard errors with 250 replications are reported.

### 1.8.2. Balancing Tests for Propensity Score Model<sup>27</sup>

In order to validate the adequacy of the propensity score model, the specifications of the model have to be examined against a number of criteria. First, it is important to assess the balance of the distribution of the covariates among treated and control groups; i.e. test that the overlap assumption holds. This can be achieved through inspecting the region of common support (overlap) between the treated and control groups (Austin, 2011; Caliendo and Kopeining, 2008; Imbens, 2004). Figure (1), in appendix (B), represents a histogram of the propensity score by the treatment status. The minima and maxima of the propensity score for the control and treated groups are [0.0067, 0.7132] and [0.0598, 0.6138], respectively. Therefore, the interval of

<sup>27</sup> This section follows, to a large extent, the guidelines provided by Caliendo and Kopeining (2008) and Imbens (2004) for the steps of the implementation of propensity score methods.

common support lies within the bounds [0.0067, 0.7132]; where the values of the propensity score for the two groups intersect. The total number of households in the region of support is 568 (426 control and 142 treated).

Another way to inspect the region of common support is through inspecting the density distribution of the propensity score in the two treatment groups (Caliendo and Kopeining, 2008; Smith and Todd, 2005). Figure (2), in appendix (B), portrays the density distribution of the propensity score for the two groups. The common support falls in an area that includes only strictly positive values of propensity scores within the distributions of the two groups.

A second check is assessing the quality of the employed matching method. This entails checking whether the matching algorithm succeeded in balancing the distribution of the covariates between the two treatment groups by reducing the systematic differences between the covariates of the two groups after conditioning on the propensity score (Caliendo and Kopeining, 2008). The applied algorithm in this study (radius caliper matching method) was found to reduce the mean standardized bias, after matching, between the two treatment groups to 5.4% (table 7).

Table (9) provides an overview of the covariates imbalance for the matched sample before and after the matching procedure. Column (1) reports the value of the Pseudo- $R^2$  from the Probit estimation of the propensity score; which indicates the degree of the ability of the chosen covariates in predicting the participation in the program (i.e. the fitness of the model). The value of the Pseudo- $R^2$  is minimal after the match, which means the nonexistence of systematic difference in the distribution of the covariates between the two groups after matching (Caliendo and Kopeining, 2008; Sianesi, 2004). Column (3) reports the p-values for the likelihood-ratio (LR) test for the joint significance of all of the chosen covariates before and after matching. The results establish that the test is not rejected before matching and is rejected after, which validates the deduction arrived at from the results in column (1). Table (1), in appendix (B), reports the results of a t-test on the selected observed covariates before and after the matching. A considerable reduction on the bias for each covariate is observed in the results, and the mean differences are not significant after matching (identified by the p-values).

**Table 9. Overall measures of covariate imbalance for participants and matched non-participants before and after matching**

Matched Sample	PS ( $R^2$ )	LR ( $\chi^2$ )	p> $\chi^2$	Mean bias	Median bias
	(1)	(2)	(3)	(4)	(5)
Before match	0.123	71.660	0.000	36.300	28.800
After match	<b>0.005</b>	1.900	<b>0.928</b>	<b>5.400</b>	5.400

Table notation: PS ( $R^2$ ): pseudo ( $R^2$ ) derived from the estimation of the propensity score on the sample of treated and their matched control units. LR ( $\chi^2$ ): the likelihood-ratio test of the joint significance of all covariates in the probit model of the propensity score before and after matching. Mean bias: standardized mean pre- and post-matching absolute bias. Median bias: standardized median pre- and post-matching absolute bias.

The propensity score methods are built on the idea of selection on observables. This idea dictates one of the main identifying assumptions that underlie these methods, which is the conditional independence or unconfounded assumption. The assumption postulates that all the relevant covariates influencing both outcome and participation in treatment are included in the model and no unobserved confounding factors are unmeasured or unaccounted for. The deviation from this assumption leads to a biased estimate for the treatment effect based on the propensity score model and, therefore, it is crucial to inspect the validity of this assumption in the specified model. However, the conditional independence assumption cannot be directly examined; what can be examined is the robustness of the study findings to the existence of potential unobserved confounding factors that may affect the outcome and participation simultaneously.

The non-parametric bounds sensitivity test for hidden (unobserved) bias, proposed by Rosenbaum (2002),<sup>28</sup> is one of the widely applied methods and is developed for propensity score matching estimators. The test assesses the robustness of the average treatment effect estimated with matching methods against the possible presence of unobserved confounders that may affect the participation decision and outcome simultaneously but are unaccounted for in the matching analysis, and quantify the magnitude of the hidden bias that might invalidate the relations observed through the model's results (Rosenbaum, 2002, 2005).

The test is identified as follows: If the probability of participation is specified by the following formula:

$$P_i = P(x_i, u_i) = \Pr [T_i = 1 \mid x_i, u_i] = F(\beta x_i + \gamma u_i) \quad (8)$$

<sup>28</sup> Also referred to as Rosenbaum's primal sensitivity approach (Rosenbaum, 2002, 2005).

where  $(P_i)$  is the participation probability for unit  $(i)$ ,  $(x_i)$  is the vector of observed characteristics for unit  $(i)$ ,  $(u_i)$  is an unobserved variable,  $(\gamma)$  is the effect of  $(u_i)$  on the participation decision and  $(F)$  is the logistic distribution. The odds that unit  $(i)$  participates in the program is given by:

$$\frac{P_i}{(1-P_i)} = \exp(\beta x_i + \gamma u_i) \quad ; \quad 0 \leq u_i \leq 1 \quad (9)$$

and the odds ratio of participation in the program, for two units  $(i)$  and  $(j)$ ; where  $(i \neq j)$ , is:

$$\frac{P_i/(1-P_i)}{P_j/(1-P_j)} = \frac{P_i(1-P_j)}{P_j(1-P_i)} = \frac{\exp(\beta x_i + \gamma u_i)}{\exp(\beta x_j + \gamma u_j)} \quad (10)$$

If the two units  $(i)$  and  $(j)$  have matching observed characteristics (i.e.  $x_i = x_j$ ), the  $x$ -vector is canceled out in equation (10), and the odds ratio for units  $(i)$  and  $(j)$  becomes:

$$\frac{P_i/(1-P_i)}{P_j/(1-P_j)} = \exp(\gamma(u_i - u_j)) \quad (11)$$

Equation (11) implies that the relative odds of participation for two matching units  $(i)$  and  $(j)$ ; where  $(i \neq j)$ , could differ by a factor equal to  $[\gamma(u_i - u_j)]$  if  $(u_i \neq u_j)$ . If  $(u_i = u_j)$  or if the difference between  $(u_i)$  and  $(u_j)$  does not influence the participation decision (i.e.  $\gamma = 0$ ), then the odds ratio for units  $(i)$  and  $(j)$  is equal to (1) and no hidden bias is observed. However, if the difference between  $(u_i)$  and  $(u_j)$  does affect the participation decision, then the odds of participation differ by  $[\gamma(u_i - u_j)]$  and the participation is not solely determined by the vector  $(x)$ . Rosenbaum (2002) further simplifies equation (11) for a binary unobserved factor  $(u \in \{0,1\})$  to the following bounds:

$$\frac{1}{e^\gamma} \geq \frac{P_i/(1-P_i)}{P_j/(1-P_j)} \geq e^\gamma \quad ; \quad \forall i \text{ and } j \text{ for which } (i \neq j) \text{ and } (x_i = x_j) \quad (12)$$

According to equation (12), the two matched units  $(i)$  and  $(j)$  feature even relative odds for participation if  $(e^\gamma = 1)$  and, in this case, the results of the model are not sensitive to the hidden

bias problem. If  $(e^\gamma > 1)$ , then the results of the model are sensitive to hidden bias and the odds for participation for two matched units (for which  $x_i = x_j$ ) differ by a factor equal to the value of  $(e^\gamma)$ . It is considered that the closer the value of  $(e^\gamma)$  to (1) the more sensitive the results are to the hidden bias problem, and the farther the value of  $(e^\gamma)$  from (1) (for example  $e^\gamma \geq 2$ ) the less sensitive the results of the model are (Aakvik, 2001; Becker and Caliendo, 2007; Keele, 2010; Liu et al, 2013; Rosenbaum, 2002, 2005). In this sense,  $(e^\gamma)$  is a measure of the degree of the departure from the probability of equal participation (Rosenbaum, 2002).

The null hypothesis of the test specifies that the outcome (in this case it is the consumption expenditure) is unaffected by the participation status and that participation has no impact on the outcome. If outcome differs by the participation status, then the participation has some positive (or negative) effect, and the assumption of even odds of participation in the program for any paired units from the control and the treated groups holds. If the null hypothesis is not rejected, then the observed impact estimated by the model is possibly not due to participation but could be due to an unobserved confounding factor and the difference in this unobserved confounder does affect the odds of participation for two units that appear comparable based on their baseline observed characteristics. The significance threshold for the test is the 5%-significance level (or 0.05); the point at which the upper bound (p-value) is greater than (0.05) is the point at which the outcome turns sensitive to the existence of an unobserved confounder and the propensity score estimators are not robust against the problem of hidden bias. This means that the propensity score model fails to account for some relevant covariate that has a strong impact on the outcome and the participation decision. This inference may lead to undermine the implications of the matching analysis.

The application of the test is based on trials for a range of different specified odds ratios of participation (or values for parameter  $(\gamma)$ ), starting from (1) and with steady increments of, for example, (0.01) or (0.05). Table (10) reports the results of the conducted sensitivity test for the 142 pairs of matched units. The values reported in the table are the upper and lower bounds on the significance levels (p-values)<sup>29</sup> for the impact of participation at different values of  $(e^\gamma)$ .<sup>30</sup>

---

<sup>29</sup> The values reported are the upper and lower significance levels (p-values) for the upper and lower bounds of Hodges-Lehmann point estimate for the signed rank test-statistic for continuous outcomes (on which the sensitivity analysis is based) (Rosenbaum 2005, Keele 2010). The extended test results are featured in appendix (B).

<sup>30</sup> The sensitivity test is conducted using the “rbounds” (DiPrete and Gangl, 2004) STATA routine for continuous outcome variables.

**Table 10. Rosenbaum (2002) bounds sensitivity test: upper and lower bounds' significance levels (p-values)**

(γ) value*	Total consumption		Food consumption		Non-food consumption		Health expenditure		Education expenditure	
	p-upper	p-lower	p-upper	p-lower	p-upper	p-lower	p-upper	p-lower	p-upper	p-lower
<b>1.00</b>	<b>0.4955</b>	<b>0.4955</b>	0.0019	0.0019	0.0076	0.0076	0.0157	0.0157	<b>0.0558</b>	<b>0.0558</b>
1.05	0.3961	0.5952	0.0008	0.0041	0.0147	0.0037	0.0285	0.0081	0.0889	0.0332
1.10	0.3071	0.6850	0.0004	0.0081	0.0263	0.0017	0.0479	0.0041	0.1322	0.0192
<b>1.15</b>	0.2314	0.7617	0.0001	0.0147	0.0435	0.0008	<b>0.0749</b>	<b>0.0020</b>	0.1853	0.0109
<b>1.20</b>	0.1698	0.8245	0.0001	0.0249	<b>0.0675</b>	<b>0.0004</b>	0.1103	0.0010	0.2469	0.0060
1.25	0.1216	0.8738	0.0000	0.0396	0.0989	0.0002	0.1542	0.0005	0.3148	0.0033
1.30	0.0851	0.9113	0.0000	0.0597	0.1380	0.0001	0.2058	0.0002	0.3865	0.0017
1.35	0.0584	0.9389	0.0000	0.0856	0.1844	0.0000	0.2640	0.0001	0.4594	0.0009
1.40	0.0393	0.9587	0.0000	0.1177	0.2372	0.0000	0.3270	0.0000	0.5309	0.0005
1.45	0.0260	0.9726	0.0000	0.1560	0.2950	0.0000	0.3928	0.0000	0.5991	0.0002
1.50	0.0169	0.9821	0.0000	0.1998	0.3563	0.0000	0.4594	0.0000	0.6622	0.0001
1.55	0.0109	0.9885	0.0000	0.2486	0.4194	0.0000	0.5250	0.0000	0.7193	0.0001
1.60	0.0069	0.9927	0.0000	0.3011	0.4825	0.0000	0.5878	0.0000	0.7697	0.0000
1.65	0.0043	0.9954	0.0000	0.3563	0.5441	0.0000	0.6467	0.0000	0.8134	0.0000
1.70	0.0027	0.9971	0.0000	0.4129	0.6028	0.0000	0.7007	0.0000	0.8505	0.0000
1.75	0.0016	0.9982	0.0000	0.4696	0.6578	0.0000	0.7493	0.0000	0.8815	0.0000
1.80	0.0010	0.9989	0.0000	0.5254	0.7082	0.0000	0.7922	0.0000	0.9071	0.0000
1.85	0.0006	0.9994	0.0000	0.5791	0.7537	0.0000	0.8294	0.0000	0.9278	0.0000
1.90	0.0004	0.9996	0.0000	0.6300	0.7940	0.0000	0.8614	0.0000	0.9444	0.0000
1.95	0.0002	0.9998	0.0000	0.6776	0.8293	0.0000	0.8883	0.0000	0.9576	0.0000
2.00	0.0001	0.9999	0.0000	0.7213	0.8598	0.0000	0.9108	0.0000	0.9678	0.0000

\*(γ) is the log odds of differential assignment to treatment due to the existence of unobserved confounding factors.

Except for the outcome of food component, the sensitivity test results reveal the high sensitivity of the matching estimators to unobserved factors that affect the participation into the program. These factors may be, for example, individual-specific traits, such as motive or desire, or program-related, community-related or geographical-related barriers, that were not controlled for by the specified propensity score matching model. According to Table (10), for the outcomes of per capita total consumption expenditure and education expenditure, the null is directly rejected at  $(\gamma) = (1)$ .

For the outcome of per capita non-food consumption expenditure rate, the largest value of parameter  $(\gamma)$  for which the upper probability bound was lower than (0.05), was (1.15). At  $(\gamma) = (1.20)$ , the relationship between participation and outcome is no longer significant ( $p\text{-upper} = 0.0675 > 0.05$ ). This means that the outcome is insensitive to a bias that would increase the odds of participation by (1.05-1.15) folds, but is sensitive to a bias that would increase the odds by

(1.20) folds. The value of  $(\gamma)$  of (1.20) can be interpreted as follows: for two units with the same values of covariates (or propensity scores), the odds of assignment is (1.20) folds higher for one unit compared to the other unit or their odds of participation differ by a rate of (20%). For the outcome of per capita health expenditure, the null hypothesis of the test is rejected at  $(\gamma)$  value of (1.15). Conversely, the results for food-consumption expenditure are considered relatively more robust against the unobserved variable or the hidden bias.

Table (10) indicates that most of the estimated participation impacts are sensitive to the hidden bias problem. This is because a very small difference in an unobserved covariate would change the inference made from the matching model. The general remark is that the findings of the propensity score matching model on the impact of participation in the program on the rate of per capita consumption expenditures are not robust to plausible hidden bias due to unobserved confounders. The direction of the bias can be inferred from the values of the bounds of the Hodges-Lehmann (H-L) point estimates for the signed rank test-statistic (provided in tables (2-6) in appendix (B)). The upper bound (H-L+) adjusts downward for the case of a positive hidden bias. A Positive hidden bias occurs when those units most likely to participate tend to have a higher (or lower) consumption even in the absence of participation and given that they have similar values in the  $x$ -vector as the units in the control group. This leads to an upward bias in the estimated treatment effects (Becker and Caliendo, 2007). This means that the estimated impacts for the propensity score matching models in tables (4) through (8) over estimate the true impact of participation in the program.

The sensitivity results for the outcome of total consumption expenditure also highlight that the significance level on the sensitivity bounds might be high first and then fall. At  $(\gamma)$  value of (1.4), we get a significant upper-p value of (0.0393). This significant upper-p value indicates greater uncertainty introduced by departure from random treatment assignment but does not alter the qualitative conclusion that the null hypothesis of no treatment effect is not plausible. It does imply, however, that the observed association does not imply causation; that is, a sufficiently large departure from the model can explain away as non-causal of any observed association (Rosenbaum, 2010).

However, it should be noted that the results of the sensitivity test does not nullify the estimated impact of participation. A sensitivity test demonstrates the extent to which the



presence of hidden bias might alter the model results, but it does not confirm the actual presence of the bias (Aakvik, 2001; Becker and Caliendo, 2007). The conclusion arrived at from the test is as follows: the overall impact of participation is sensitive to a hidden bias which is positive in its direction; if units with high (or low) values of ( $u$ ) are over represented, then the estimated impact over estimates the true participation impact. Moreover, as stated earlier, the sensitivity test does not directly validate the conditional independence assumption and, therefore, it remains inconclusive whether or not the assumption holds for the given model specification and the chosen observed covariates or the used dataset. What could be stated is that the results of the matching model are sensitive to possible deviation from the underlying identifying conditional independence assumption and this should be taken into consideration when accounting for the results.

## **1.9. Discussion and Concluding Remarks**

The empirical part of this study sets out the results of the impact of the Egyptian Takaful and Karamah social schemes on a single (short-term) dimension of welfare, which is the per capita consumption expenditure. The valuation of the impact is upheld through comparing the outcomes of per capita total consumption expenditure, per capita food consumption expenditure, per capita non-food consumption expenditure and per capita consumption expenditures on healthcare and education services for the participant and the nonparticipant households using a non-experimental data of the Egyptian HIECS 2017-2018. The per capita total consumption expenditure is conceived a proxy measurement for income and an indicator for the household's welfare, whereas per capita non-food consumption expenditures are indicators for asset-accumulation<sup>31</sup> (Ahmed et al, 2009). The propensity score matching method as well as propensity score weighting (or odds-weighted regression) were used to estimate the impact of participation for different subgroups of the sampled households.

The impact of participation in Takaful and Karamah social schemes on per capita consumption was found negative for the categories of total consumption expenditures and its food component, while it was positive for the non-food component as well as healthcare and

---

<sup>31</sup> The household's accumulated assets could be in the form of consumption or productive assets (Ahmed et al, 2009).

education services for the two applied models. For the propensity score matching model, the specification of the model succeeded in eliminating the systematic differences (balancing) between the selected observed covariates of the treated and the control groups when adjusting for the observed differences between the participants and the non-participants. However, except for the outcome of food component, the results of the propensity score model for the impact of participation on consumption were statistically insignificant. As for the results of the odds-weighted regression model, they were in accord with the results of the propensity score model except for the heterogeneous impact on per capita total consumption expenditure and for the impact on per capita health consumption expenditure.

Compared to the propensity score matching model, the odds-weighted regression results indicated much stronger treatment impact with magnitudes that surpassed considerably those obtained from the propensity score model. It should be underscored in this regard that, because of being based on the assumption of randomization and not being built on actual random trials, the matching methods typically tend to increase the statistical power of the evaluation design (Ahmed et al, 2009; Stürmer and Brenner, 2001).<sup>32</sup> It follows that the true size of the impact could be even smaller than the impact identified by the matching model.

In addition, the standard errors of the odds-weighted regression model were greater than those of the propensity score matching model. This reflects the less precision of the odds-weighted regression model and the more uncertainty that is stemmed from the stringent nature of the semi-parametric model, which is inherent from the much stronger assumptions that have to be complied for relative to the non-parametric propensity score matching model. A misspecification of the model through, for example, the existence of unobservable factors that might cause potential endogeneity of participation can lead to the overestimation of the true treatment impact on the outcome under evaluation (DiPrete and Gangl, 2004).<sup>33</sup>

---

<sup>32</sup> The statistical power of the analysis refers to the pre-probability that the analysis will yield a statistically significant outcome; given a particular study design, sample size and data analysis tool, if the true impact of the treatment is of a particular size (Djimeu and Houndolo, 2016; Juras et al, 2016).

<sup>33</sup> In an investigation of the finite-sample properties of propensity score estimators through a Monte-Carlo simulation analysis, Frölich (2004) found that the weighted estimators for average treatment effect performed the worst compared to pairwise matching and a number of other estimators. The weighted estimators were found the least efficient estimators (have the largest mean squared error) and were highly sensitive to the specifications of the regression functions and the densities of the propensity scores of the treated and control groups (Frölich, 2004).

At this point, a number of considerations should be taken into account. First, it should be noted that, in order for the approach that combines weighting using propensity score with regression to yield consistent estimators, either the regression model or the propensity score model should be correctly specified (Abadie and Imbens, 2002; Imbens, 2004). Second, in observational data, the regression and matching methods rule out the confounding that arises from omitted variables in case of selection on observables (i.e. under the exogeneity of confounding covariates) (Cameron and Trivedi, 2005; Imbens, 2004; Smith and Todd, 2005; Todd, 2010). If there are unobservable factors that affect both participation decision and outcome (i.e. the treatment participation is endogenous) then the identification of the average treatment impact through these methods is infeasible and other methods for selection on unobservables should be considered (Cameron and Trivedi, 2005).

The specifications of the propensity score model were evaluated through, first, the inspection of the satisfaction of the common support (overlap) assumption. The model was found to fairly balance between the covariates of the treated and control groups. This eliminates the misspecification of the propensity score model.<sup>34</sup>

The results of the propensity score matching model were then tested for the sensitivity (robustness) to hidden bias due to correlation between unobserved confounding factors and the status of participation. The test is conducted by bounding the matching estimators at different values of the hidden bias under the null hypothesis of no impact for participation. The findings of the test were that the impacts of participation, based on the matching estimators, were sensitive (not robust) to the hidden bias problem (except for the results on per capita food consumption, which demonstrated a slightly moderate robustness against the bias).

An alternative identification strategy would involve using the instrumental variable (IV) approach to complement the results of the matching model in order to permit for selection on unobservables and deal with biases originated by the potential endogeneity of the participation decision in the program (Cameron and Trivedi, 2005). The IV approach relaxes the conditional

---

<sup>34</sup> The specified model can be considered a parsimonious model that contains the minimum set of conditioning covariates that can be extracted from the core information available in the HIECS survey. Including higher order or interaction terms might lower the evaluation bias and improve the efficiency of the model (Diaz and Handa, 2005; Smith and Todd, 2005). It is worth to mention that, according to the findings of Heckman, Ichimura, Smith and Todd (1998) and Smith and Todd (2005), the matching estimators are sensitive to the set of matching variables. Moreover, the composition of the control group is an important determinant of the success of the matching technique (Diaz and Handa, 2005).

independence assumption and can provide complementary information about the potential average causal relationship between the program participation and the outcome variables (Cameron and Trivedi, 2005; DiPrete and Gangl, 2004). Consequently, models of “selection on unobservable” are considered a much fitting approach in this case and the alternative identification strategy would involve the estimation of a Local Average Treatment Effect (LATE) model, which adopts the IV approach, following the advocacy of Angrist and Imbens (1995), Heckman, Ichimura, and Todd (1997) and Smith and Todd (2005). The identification of treatment impact through LATE relies on much weaker assumptions compared to other treatment impact estimators (Frölich, 2007; Heckman, Ichimura, and Todd, 1997). The incentive for participation is randomized rather than the participation status (through finding a valid instrument that shifts the participation probability and does not affect the participation status directly) allowing the identification of the average treatment impact for those whose participation can be changed by changing the participation incentives (Angrist and Imbens, 1995). A further discussion of methods of selection on unobservables is, however, beyond the scope of this study.

From an economic perspective, on the other hand, the results of the insignificant impact of Takaful and Karamah social schemes’ transfers on per capita consumption expenditures that the propensity score matching model yielded is in line with the fact that increasing incomes significantly or being an income-generating mechanism are not the primary objectives of the social programs. Rather, the aim of the social programs is to provide a supplementary assistance that facilitates the process of consumption smoothing during the periods of monetary congestion that are encountered through the life cycle of the eligible households. For most beneficiary households, Takaful and Karamah social transfers are not the sole source of income. As a result, the small amounts of the transfers and their income-multiplier effects are not enough for the extremely poor beneficiary households to increase their consumption expenditures to significant levels (Ahmed et al, 2009).

Another note to highlight is that, since most of the beneficiary households are from rural areas, it follows that agricultural assets and landownership is the most important constituent in a household’s asset-base—as was eminent from the characteristics of Takaful and Karamah social schemes’ beneficiaries in table (1). This fact could have induced the result obtained for the heterogenous impact on the non-food consumption expenditures. When the results are

disaggregated by asset wealth, the impact of participation on per capita non-food expenditure was significantly larger among non-poor households compared to their poorer counterparts. The greater impact on non-food expenditure among non-poor households suggests that these households use a substantial portion of their cash transfers to further re-enforce their asset base. It also signifies the presence of possible threshold effects, where a minimum level of asset base is required for a beneficiary household to be able to use a substantial portion of the transfer on asset accumulation. Moreover, the threshold effect is also noted for the per capita healthcare and education expenditures, with the programs positively impacting healthcare expenditure among asset-poorer beneficiary households and positively impacting education expenditure among asset-wealthier beneficiaries. These findings on positive association between Takaful and Karamah social schemes' transfers and per capita healthcare and education expenditures for beneficiary households, and the variations of the nature and distribution of impacts across wealth levels, recommend a further inspection of the true inducing motives for the rise of these expenditures; whether or not these motives are truly related to the programs' conditionalities, and identify the need for impact studies to carefully examine the heterogeneous effects of these social interventions.

A final note to reference would be that the combination of the findings from the applied models might lead to conclude that there is not a sufficient evidence to establish a significant impact of Takaful and Karamah social schemes' transfers on the per capita consumption of the beneficiary household in order to conclude that a causal effect exists. However, these findings should be interpreted with caution especially when employed for drafting policy recommendations.

It should be underscored, in conclusion, that access to base-line data from governmental sources posed a serious and challenging limitation to the conduction and outcomes of this study; access to reliable public data is necessary for a broadened evidence base and accountable research findings. In addition, further empirical research is needed to investigate the impacts of cash-based transfers in different settings in order to analyze their outcomes and assess the pathways by which their effects are achieved.

## References

- Aakvik, Arild. 2001. Bounding a Matching Estimator: The Case of a Norwegian Training Program. *Oxford Bulletin of Economics and Statistics*, Vol. 63 (1): 115–143.
- Abadie, Alberto and Guido W. Imbens. 2002. Simple and Bias-Corrected Matching Estimators for Average Treatment Effects. The National Bureau of Economic Research (NBER) Technical Working Paper Series, Working Paper No. 283.
- . 2006. Large Sample Properties of Matching Estimators for Average Treatment Effects. *Econometrica*, Vol. 74 (1): 235–267.
- . 2011. Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business and Economic Statistics*, Vol. 29 (1): 1–11. DOI: 10.1198/jbes.2009.07333.
- . 2016. Matching on the Estimated Propensity Score. *Econometrica*, Vol. 84 (2): 781–807. DOI: 10.3982/ECTA11293.
- Ahmed, Akhter U., Agnes R. Quisumbing, Mahbuba Nasreen, John F. Hoddinott, and Elizabeth Bryan. 2009. Comparing Food and Cash Transfers to the Ultra Poor in Bangladesh. International Food Policy Research Institute (IFPRI) research monograph No. 163. DOI: 10.2499/9780896291737RR163.
- Angrist, Joshua D. and Guido W. Imbens. 1995. Identification and Estimation of Local Average Treatment Effect. National Bureau of Economic Research (NBER) Technical Working Paper No. 118.
- Augurzky, Boris and Christoph M. Schmidt. 2001. The Propensity Score: A Means to an End. IZA Discussion Paper No. 271, March 2001.
- Austin, Peter C. 2011. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research*, Vol. 46: 399–424.
- Banerjee, Abhijit V. and Esther Duflo. 2004. Growth Theory through the Lens of Development Economics. Massachusetts Institute of Technology (MIT), Department of Economics, Working Paper No. 05–01. <http://dx.doi.org/10.2139/ssrn.651483>.
- . 2009. The Experimental Approach to Development Economics. *Annual Review of Economics*, Annual Reviews, Vol. 1 (1): 151–178.
- Baser, Onur. 2006. Too Much Ado about Propensity Score models: Comparing Methods of Propensity Score Matching. In *Value in health: The journal of the International Society for Pharmacoeconomics and Outcomes Research*, Vol. 9 (6): 377–385. DOI: 10.1111/j.1524-4733.2006.00130.x.
- Becker, Sascha O. and Marco Caliendo. 2007. “mhbounds” – Sensitivity Analysis for Average Treatment Effects. IZA Discussion Paper No. 2542, January 2007.
- Bryson, Alex, Richard Dorsett and Susan Purdon. 2002. The Use of Propensity Score Matching in the Evaluation of Labour Market Policies. London School of Economics, Department for Work and Pensions, Working Paper No. 4. <http://eprints.lse.ac.uk/id/eprint/4993>.
- Caliendo, Marco and Sabine Kopeining. 2008. Some Practical Guidance for The Implementation of Propensity Score Matching. *Journal of economic surveys*, Vol. 22 (1): 31–72.
- Cameron, A. Colin and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press. ISBN: 0521848059.
- Central Agency for Public Mobilization and Statistics (CAPMAS). 2014. Newsletter for Poverty Indicators According to Household, Income, Expenditure, and Consumption Survey (HIECS) 2012/2013.
- . 2016. Newsletter for Poverty Indicators According to Household, Income, Expenditure, and Consumption Survey (HIECS) 2015.
- . 2019. Newsletter for Poverty Indicators According to Household, Income, Expenditure, and Consumption Survey (HIECS) 2017/2018.
- Dehejia, Rajeev and Sadek Wahba, 1997. Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs. In Rajeev Dehejia “Econometric Methods for Program

Evaluation” Ph.D. Dissertation, Harvard University, 1997, Chapter 1.

———. 1998. Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of the Training Programs. The National Bureau of Economic Research (NBER) Working Paper Series, Working Paper No. 6586, June 1998.

———. 2002. Propensity Score-Matching Methods for Non-experimental Causal Studies. *The Review of Economics and Statistics*, Vol. 84 (1): 151–161.

Diaz, Juan Jose and Sudhanshu Handa. 2005. An Assessment of Propensity Score Matching as a Non Experimental Impact Estimator: Evidence from Mexico’s PROGRESA Program. Office of Evaluation and Oversight Working Paper, OVE/WP-04/05, July 22, 2005.

DiPrete, Thomas A. and Markus Gangl. 2004. Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, Vol. 34(1): 271–310.

Djimeu, Eric W. and Deo-Gracias Houndolo. 2016. Power Calculation for Causal Inference in Social Science: Sample Size and Minimum Detectable Effect Determination. International Initiative for Impact Evaluation Working Paper No. 26, March 2016.

Filmer, Deon and Lant H. Pritchett. 2001. Estimating Wealth Effects without Expenditure Data or Tears: An Application to Educational Enrollments of India. *Demography*, 38 (1): 115–132.

Frölich, Markus. 2004. Finite-Sample Properties of Propensity-Score Matching and Weighting Estimators. *The Review of Economics and Statistics*, Vol. 86 (1): 77–90.

———. 2007. Nonparametric IV Estimation of Local Average Treatment Effects with Covariates. *Journal of Econometrics*, Vol. 139 (1): 35–75.

Härdle, Wolfgang. 1995. Applied Nonparametric Regression. Econometric Society Monographs, Cambridge: Cambridge University Press.

Heckman, James J. and Jeffrey A. Smith. 1995. Assessing the Case for Social Experiments. *The*

*Journal of Economic Perspectives*, Vol. 9 (2): 85–110.

Heckman, James J., Hidehiko Ichimura and Petra Todd. 1998. Matching As an Econometric Evaluations Estimator. *Review of Economic Studies*, Vol. 65 (2): 261–294.

———. 1997. Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies*, Vol. 64 (4): 605–654.

Heckman, James J., Hidehiko Ichimura, Jeffrey A. Smith and Petra Todd. 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica*, Vol. 66 (5): 1017–1098.

Heinrich, Carolyn, Alessandro Maffioli and Gonzalo Vazquez. 2010. A Primer for Applying Propensity-Score Matching: Impact-Evaluation Guidelines. Inter-American Development Bank, Office of Strategic Planning and Development Effectiveness, Technical Note No. IDB-TN-161, August 2010.

Hill, Jennifer L., Jerome P. Reiter and Elaine L. Zanutto. 2004. A Comparisons of Experimental and Observational Data Analyses. In “Applied Bayesian Modeling and Causal Inference From an Incomplete-Data Perspective,” edited by Andrew Gelman and Xiao-Li Meng, 44–56. Wiley.

Hirano, Keisuke and Guido W. Imbens. 2002. Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catherization. *Health Services and Outcomes Research Methodology*, Vol. 2: 259–278.

Hirano, Keisuke, Guido W. Imbens and Geert Ridder. 2003. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. *Econometrica*, Vol. 71 (4): 1161–1189.

Hujer, Reinhard, Marco Caliendo and Stephan Thompsen. 2003. New Evidence on the Effects of Job Creation Schemes in Germany - A Matching Approach with Threefold Heterogeneity. IZA Discussion Paper No. 750, April 2003.

Imbens, Guido W. 2004. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. *The Review of Economics and Statistics*, Vol. 86 (1): 4–29.

- . 2014. Matching Methods in Practice: Three Examples. IZA Discussion Paper No. 8049, March 2014.
- Juras, Randall, Alison Comfort and Ed Bein. 2016. How Study Design Influences Statistical Power in Community-Level Evaluations. Tier 1B Evaluation Technical Assistance Brief No.3, Office of Adolescent Health, USA.
- Keele, Luke. 2010. An Overview of “rbounds”: An R Package for Rosenbaum Bounds Sensitivity Analysis with Matched data. <https://pdfs.semanticscholar.org/c0b1/823186cf5e0869ee35657d6809f803c8e35c.pdf>
- Lalonde, Robert. 1986. Evaluating the Econometric Evaluations of Training Programs. *American Economic Review*, Vol. 76 (4): 604–620.
- Leuven, Edwin and Barbara Sianesi. 2003. PSMATCH2: STATA Module to Perform Full Mahalanobis and Propensity Score Matching, Common Support Graphing and Covariate Imbalance Testing.
- Liu, Weiwei, S. Janet Kuramoto and Elizabeth A. Stuart. 2013. An Introduction to Sensitivity Analysis for Unobserved Confounding in Non-Experimental Prevention Research. *Journal of Prevention Science*, Vol. 14 (6): 570–580. DOI:10.1007/s11121-012-0339-5.
- Lunt, Mark. 2013. Selecting an Appropriate Caliper Can Be Essential for Achieving Good Balance with Propensity Score Matching. *American Journal of Epidemiology*, Vol. 179 (2): 226–235. DOI: 10.1093/aje/kwt212.
- Ministry of Finance (MoF). 2015. Monthly Financial Bulletin. December 2015.
- Roberts, Chris and David J. Torgerson. 1999. Understanding Controlled Trials: Baseline Imbalance in Randomised Controlled Trials. *The British Medical Journal (BMJ)*, Vol. 319 (7203): 185.
- Rosenbaum, Paul R. 2005. Sensitivity Analysis in Observational Studies. In *Encyclopedia of Statistics in Behavioral Science*, edited by Brian S. Everitt and David C. Howell, Vol. 4: 1809–1814. Chichester: John Wiley and Sons, Ltd.
- . 2010. Design of Observational Studies. Springer Series in Statistics. Springer, New York, NY. [https://doi.org/10.1007/978-1-4419-1213-8\\_3](https://doi.org/10.1007/978-1-4419-1213-8_3)
- Rosenbaum, Paul R. and Donald B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, Vol. 70 (1): 41–55.
- . 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, Vol. 39 (1): 33–38.
- Rubin, Donald B. 1997. Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Annals of Internal Medicine*, Vol. 127 (8): 757–763.
- Smith, Jeffrey and Petra Todd. 2005. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? *Journal of Econometrics*, Vol. 125 (1-2): 305–353.
- Stürmer, Til and Hermann Brenner. 2001. Degree of Matching and Gain in Power and Efficiency in Case-Control Studies. *Journal of Epidemiology*, Vol. 12 (1): 101–108.
- Tembo, Gelson, Nicholas Freeland, Bernadette Chimai and Esther Schüring. 2014. Social Cash Transfers and Household Welfare: Evidence from Zambia's Oldest Scheme. *Applied Economics and Finance*, Vol. 1 (1): 13–26.
- Todd, Petra E. 2010. Matching Estimators. In “Microeconometrics,” edited by Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan, 2010.
- . 2003. Extending Social Security: Policies for Developing Countries. *International Labour Review*, Vol. 142 (3): 277–294. International Labor Organization (ILO), Geneva.
- White, Howard and Shagun Sabarwal. 2014. Quasi-experimental Design and Methods. Methodological Briefs: Impact Evaluation 8, UNICEF Office of Research, Florence.
- Zanutto, Elaine L. 2006. A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data. *Journal of Data Science*, Vol. 4 (2006): 67–91.



## **Appendix A: Variables Included in the Construction of the Asset-Wealth Index**

### **A. Material assets owned by a household:**

- |                                    |                           |
|------------------------------------|---------------------------|
| 1. Car                             | 20. Air conditioner       |
| 2. Bicycle                         | 21. Electric fan          |
| 3. Motor bicycle                   | 22. Heater                |
| 4. Land line                       | 23. Iron                  |
| 5. Mobile phone                    | 24. Colored TV            |
| 6. Smart phone                     | 25. Black-and-white TV    |
| 7. Internet                        | 26. Video/DVD             |
| 8. Fridge                          | 27. Mp3/Mp4/Mp5           |
| 9. Deep-freezer                    | 28. Radio/cassette player |
| 10. Water cooler                   | 29. Satellite             |
| 11. Electric cooker                | 30. Computer              |
| 12. Gas cooker                     | 31. Tablet                |
| 13. Microwave/grill                | 32. Camera                |
| 14. Washing machine                | 33. Digital Camera        |
| 15. Half-automatic washing machine | 34. Water filter          |
| 16. Automatic washing machine      | 35. Blender               |
| 17. Dish washer                    | 36. Kitchen machine       |
| 18. Water heater                   | 37. Sewing machine        |
| 19. Vacuum cleaner                 | 38. Electric generator    |

### **B. Household characteristics:**

1. Type of dwelling: 0 “other” 1 “separate room or more” 2 “room or more in a building” 3 “rural house” 4 “apartment” 5 “more than one apartment” 6 “whole building” 7 “villa”
2. Number of household members per room
3. Wall material: 0 “other” 1 “tin” 2 “wood” 3 “mud bricks” 4 “stone” 5 “concrete” 6 “bricks and cement”
4. Floor material: 0 “other” 1 “cement/concrete” 2 “tiles” 3 “vinyl” 4 “ceramic” 5 “parquet” 6 “marble”
5. Roof material: 0 “none” 1 “cardboard” 2 “straw/palm fronds” 3 “mats” 4 “wood” 5 “metal” 6 “concrete”
6. Source of ventilation: 0 “none” 1 “hole in wall” 2 “more than one hole” 3 “window” 4 “more than one window”
7. Source of water: 0 “other” 1 “water springs” 2 “well” 3 “trumbah” 4 “public network”
8. Water piped into residence: 0 “no connection” 1 “water piped into building” 2 “water piped into dwelling”
9. Bathroom location: 0 “none” 1 “outside and shared” 2 “outside and private” 3 “inside and shared” 4 “inside and private”
10. Type of toilet facility: 0 “none” 1 “field” 2 “bucket” 3 “pit” 4 “traditional wo/flush” 5 “traditional w/flush” 6 “modern w/flush”

- 11. Kitchen: 0 “none” 1 “shared” 2 “private”
- 12. Sewage network: 0 “none” 1 “other” 2 “trunsh” 3 “private network” 4 “public sewage network”
- 13. Garbage disposal: 0 “other” 1 “throw in street” 2 “public containers” 3 “garbage collecting company” 4 “garbage collector”
- 14. Electricity: 0 “not electrified” 1 “electrified”
- 15. Cooking fuel: 0 “wood and logs” 1 “kerosene” 2 “LPG/natural gas” 3 “electricity”

**C. Agricultural assets owned by a member of the household:**

These include the net annual income from:

- 1. agricultural land
- 2. breeding animals
- 3. poultry
- 4. feeding animals
- 5. beehives
- 6. fish farms
- 7. agricultural machinery
- 8. other agricultural activities

**D. Annual net income accrued to the household from financial properties:**

These include the net annual income from:

- 1. dividends
- 2. bonds interests
- 3. deposits interests
- 4. banks savings and current accounts returns
- 5. postal savings fund
- 6. returns on investment certificates
- 7. returns on investment sharing
- 8. other financial properties

## Appendix B: Auxiliary Output

Figure B.1. Region of common support: distribution of propensity score by treatment status

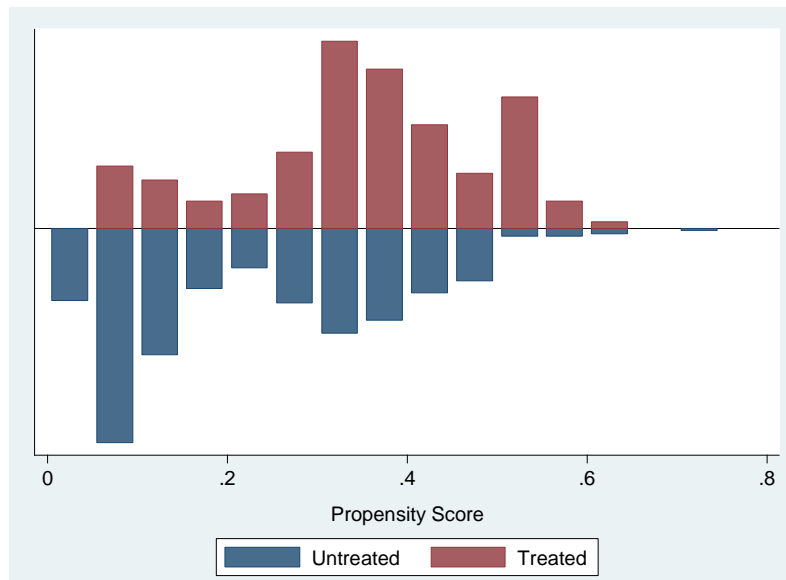
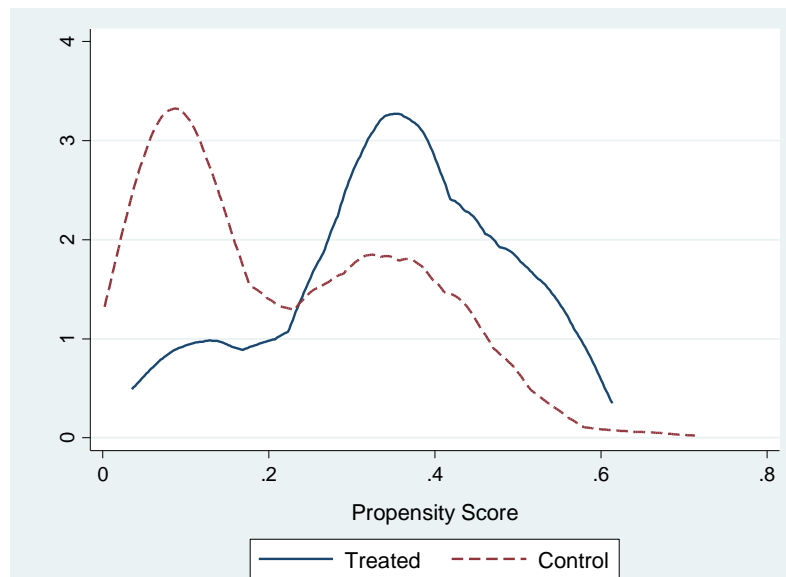
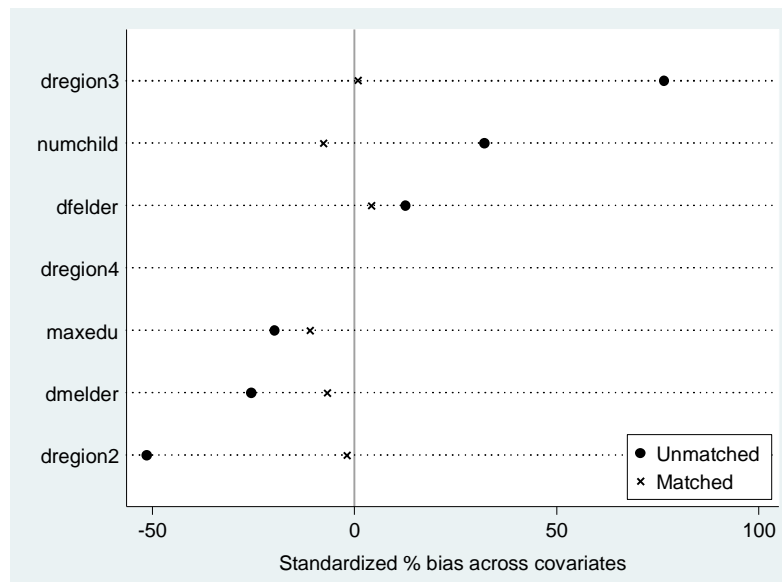


Figure B.2. Region of common support: Kernel density functions for treated and control groups



**Figure B.3. Standardized percentage bias across covariates before and after matching**



**Table B.1. Results of a t-test on the full set of chosen covariates after matching**

Variable	Matched/ Unmatched	Mean		%Bias	%Reduction in  bias	t-test	
		Treated	Control			t	p> t
Education level of most educated member	U	2.293	2.493	-19.700		-2.030	0.043
	M	2.293	2.405	-11.100	43.900	-0.910	0.364
children (14 years or younger)	U	3.421	2.987	32.100		3.130	0.002
	M	3.421	3.526	-7.800	75.800	-0.620	0.536
Male elders (61+ years) dummy	U	0.015	0.065	-25.500		-2.230	0.026
	M	0.015	0.028	-6.800	73.600	-0.730	0.463
Female elders (61+ years) dummy	U	0.083	0.051	12.600		1.320	0.189
	M	0.083	0.072	4.100	67.300	0.310	0.755
Region (2) dummy	U	0.098	0.297	-51.500		-4.670	0.000
	M	0.098	0.105	-1.800	96.400	-0.190	0.848
Region (3) dummy	U	0.865	0.536	76.600		7.010	0.000
	M	0.865	0.861	0.800	98.900	0.080	0.934
Region (4) dummy	U	0.000	0.000	0.000		.	.
	M	0.000	0.000	0.000	.	.	.

**Table B.2. Results of Rosenbaum (2002) bounds sensitivity test: total consumption expenditures**

( $\gamma$ ) value*	p-upper	p-lower	H-L+	H-L-	CI+	CI-
1	0.4955	0.4955	-0.0002	-0.0002	-0.0335	0.0283
1.05	0.3961	0.5952	-0.0040	0.0034	-0.0378	0.0324
1.1	0.3071	0.6850	-0.0082	0.0069	-0.0424	0.0361
1.15	0.2314	0.7617	-0.0116	0.0104	-0.0466	0.0392

1.2	0.1698	0.8245	-0.0156	0.0138	-0.0510	0.0420
1.25	0.1216	0.8738	-0.0191	0.0168	-0.0551	0.0445
1.3	0.0851	0.9113	-0.0224	0.0196	-0.0595	0.0474
1.35	0.0584	0.9389	-0.0259	0.0220	-0.0637	0.0501
1.4	0.0393	0.9587	-0.0294	0.0246	-0.0679	0.0526
1.45	0.0260	0.9726	-0.0324	0.0274	-0.0714	0.0551
1.5	0.0169	0.9821	-0.0352	0.0304	-0.0748	0.0575
1.55	0.0109	0.9885	-0.0381	0.0327	-0.0781	0.0598
1.6	0.0069	0.9927	-0.0410	0.0349	-0.0823	0.0624
1.65	0.0043	0.9954	-0.0440	0.0373	-0.0855	0.0641
1.7	0.0027	0.9971	-0.0467	0.0393	-0.0889	0.0663
1.75	0.0016	0.9982	-0.0494	0.0410	-0.0926	0.0679
1.8	0.0010	0.9989	-0.0525	0.0427	-0.0959	0.0697
1.85	0.0006	0.9994	-0.0547	0.0444	-0.0987	0.0717
1.9	0.0004	0.9996	-0.0577	0.0463	-0.1017	0.0732
1.95	0.0002	0.9998	-0.0602	0.0480	-0.1049	0.0747
2	0.0001	0.9999	-0.0630	0.0496	-0.1076	0.0767

\*( $\gamma$ ) is the log odds of differential assignment to treatment due to the existence of unobserved confounding factors.

**Table B.3. Results of Rosenbaum (2002) bounds sensitivity test: food consumption expenditures**

( $\gamma$ ) value*	p-upper	p-lower	H-L+	H-L-	CI+	CI-
1	0.0019	0.0019	-0.0626	-0.0626	-0.1053	-0.0214
1.05	0.0008	0.0041	-0.0672	-0.0569	-0.1107	-0.0161
1.1	0.0004	0.0081	-0.0739	-0.0526	-0.1156	-0.0105
1.15	0.0001	0.0147	-0.0782	-0.0482	-0.1200	-0.0052
1.2	0.0001	0.0249	-0.0831	-0.0434	-0.1247	0.0000
1.25	0.0000	0.0396	-0.0877	-0.0394	-0.1288	0.0042
1.3	0.0000	0.0597	-0.0919	-0.0345	-0.1339	0.0091
1.35	0.0000	0.0856	-0.0962	-0.0301	-0.1379	0.0133
1.4	0.0000	0.1177	-0.0999	-0.0265	-0.1417	0.0172
1.45	0.0000	0.1560	-0.1034	-0.0225	-0.1459	0.0214
1.5	0.0000	0.1998	-0.1082	-0.0191	-0.1492	0.0248
1.55	0.0000	0.2486	-0.1110	-0.0158	-0.1523	0.0289
1.6	0.0000	0.3011	-0.1139	-0.0122	-0.1562	0.0330
1.65	0.0000	0.3563	-0.1172	-0.0087	-0.1587	0.0364
1.7	0.0000	0.4129	-0.1200	-0.0051	-0.1626	0.0396
1.75	0.0000	0.4696	-0.1236	-0.0019	-0.1656	0.0428
1.8	0.0000	0.5254	-0.1260	0.0014	-0.1705	0.0456
1.85	0.0000	0.5791	-0.1284	0.0041	-0.1730	0.0495
1.9	0.0000	0.6300	-0.1319	0.0067	-0.1758	0.0524
1.95	0.0000	0.6776	-0.1346	0.0102	-0.1793	0.0556

2                      0.0000                      0.7213                      -0.1371                      0.0127                      -0.1818                      0.0581

\*( $\gamma$ ) is the log odds of differential assignment to treatment due to the existence of unobserved confounding factors.

**Table B.4. Results of Rosenbaum (2002) bounds sensitivity test: non-food consumption expenditures**

( $\gamma$ ) value*	p-upper	p-lower	H-L+	H-L-	CI+	CI-
1	0.0076	0.0076	0.0561	0.0561	0.0120	0.0982
1.05	0.0147	0.0037	0.0520	0.0618	0.0059	0.1035
1.1	0.0263	0.0017	0.0458	0.0669	-0.0008	0.1089
1.15	0.0435	0.0008	0.0403	0.0721	-0.0075	0.1142
1.2	0.0675	0.0004	0.0353	0.0772	-0.0130	0.1190
1.25	0.0989	0.0002	0.0295	0.0813	-0.0176	0.1233
1.3	0.1380	0.0001	0.0253	0.0849	-0.0243	0.1266
1.35	0.1844	0.0000	0.0208	0.0901	-0.0286	0.1307
1.4	0.2372	0.0000	0.0171	0.0935	-0.0333	0.1346
1.45	0.2950	0.0000	0.0132	0.0971	-0.0380	0.1374
1.5	0.3563	0.0000	0.0096	0.1002	-0.0421	0.1404
1.55	0.4194	0.0000	0.0056	0.1041	-0.0477	0.1435
1.6	0.4825	0.0000	0.0012	0.1077	-0.0526	0.1469
1.65	0.5441	0.0000	-0.0030	0.1108	-0.0556	0.1494
1.7	0.6028	0.0000	-0.0077	0.1143	-0.0602	0.1520
1.75	0.6578	0.0000	-0.0109	0.1177	-0.0654	0.1553
1.8	0.7082	0.0000	-0.0143	0.1203	-0.0698	0.1591
1.85	0.7537	0.0000	-0.0175	0.1229	-0.0745	0.1617
1.9	0.7940	0.0000	-0.0214	0.1253	-0.0799	0.1641
1.95	0.8293	0.0000	-0.0251	0.1274	-0.0842	0.1667
2	0.8598	0.0000	-0.0276	0.1299	-0.0886	0.1693

\*( $\gamma$ ) is the log odds of differential assignment to treatment due to the existence of unobserved confounding factors.

**Table B.5. Results of Rosenbaum (2002) bounds sensitivity test: healthcare consumption expenditures**

( $\gamma$ ) value*	p-upper	p-lower	H-L+	H-L-	CI+	CI-
1	0.0157	0.0157	0.2041	0.2041	0.0168	0.3787
1.05	0.0285	0.0081	0.1767	0.2257	-0.0063	0.4028
1.1	0.0479	0.0041	0.1582	0.2480	-0.0256	0.4231
1.15	0.0749	0.0020	0.1364	0.2714	-0.0455	0.4427
1.2	0.1103	0.0010	0.1164	0.2911	-0.0624	0.4607
1.25	0.1542	0.0005	0.0980	0.3073	-0.0807	0.4818
1.3	0.2058	0.0002	0.0780	0.3248	-0.1033	0.5009
1.35	0.2640	0.0001	0.0599	0.3439	-0.1214	0.5185
1.4	0.3270	0.0000	0.0415	0.3587	-0.1378	0.5381
1.45	0.3928	0.0000	0.0242	0.3727	-0.1578	0.5536
1.5	0.4594	0.0000	0.0084	0.3877	-0.1766	0.5688

1.55	0.5250	0.0000	-0.0071	0.4032	-0.1943	0.5840
1.6	0.5878	0.0000	-0.0202	0.4144	-0.2103	0.5967
1.65	0.6467	0.0000	-0.0310	0.4287	-0.2238	0.6103
1.7	0.7007	0.0000	-0.0452	0.4424	-0.2378	0.6220
1.75	0.7493	0.0000	-0.0563	0.4538	-0.2544	0.6326
1.8	0.7922	0.0000	-0.0672	0.4653	-0.2669	0.6445
1.85	0.8294	0.0000	-0.0785	0.4787	-0.2805	0.6544
1.9	0.8614	0.0000	-0.0922	0.4909	-0.2920	0.6683
1.95	0.8883	0.0000	-0.1052	0.5039	-0.3055	0.6807
2	0.9108	0.0000	-0.1163	0.5147	-0.3178	0.6926

\*( $\gamma$ ) is the log odds of differential assignment to treatment due to the existence of unobserved confounding factors.

**Table B.6. Results of Rosenbaum (2002) bounds sensitivity test: education consumption expenditures**

( $\gamma$ ) value*	p-upper	p-lower	H-L+	H-L-	CI+	CI-
1	0.0558	0.0558	0.1263	0.1263	-0.0261	0.2975
1.05	0.0889	0.0332	0.1066	0.1495	-0.0425	0.3190
1.1	0.1322	0.0192	0.0907	0.1672	-0.0660	0.3384
1.15	0.1853	0.0109	0.0752	0.1855	-0.0883	0.3549
1.2	0.2469	0.0060	0.0575	0.2053	-0.1054	0.3742
1.25	0.3148	0.0033	0.0427	0.2246	-0.1204	0.3933
1.3	0.3865	0.0017	0.0252	0.2429	-0.1399	0.4085
1.35	0.4594	0.0009	0.0077	0.2602	-0.1541	0.4206
1.4	0.5309	0.0005	-0.0049	0.2753	-0.1694	0.4373
1.45	0.5991	0.0002	-0.0176	0.2892	-0.1880	0.4507
1.5	0.6622	0.0001	-0.0281	0.3011	-0.2028	0.4642
1.55	0.7193	0.0001	-0.0388	0.3156	-0.2165	0.4815
1.6	0.7697	0.0000	-0.0520	0.3295	-0.2315	0.4950
1.65	0.8134	0.0000	-0.0689	0.3395	-0.2429	0.5107
1.7	0.8505	0.0000	-0.0829	0.3505	-0.2584	0.5249
1.75	0.8815	0.0000	-0.0955	0.3618	-0.2733	0.5385
1.8	0.9071	0.0000	-0.1054	0.3739	-0.2861	0.5485
1.85	0.9278	0.0000	-0.1145	0.3837	-0.2982	0.5581
1.9	0.9444	0.0000	-0.1253	0.3981	-0.3127	0.5714
1.95	0.9576	0.0000	-0.1372	0.4069	-0.3239	0.5833
2	0.9678	0.0000	-0.1466	0.4135	-0.3349	0.5947

\*( $\gamma$ ) is the log odds of differential assignment to treatment due to the existence of unobserved confounding factors.