This implies, finally, that:

$$\text{ATENT} = \text{E}\left\{ \frac{[w - p(\mathbf{x})]y}{p(w=0)\,p(\mathbf{x})]} \right\} \tag{6}$$

## 3. Sample estimation and standard errors for ATE, ATET and ATENT

Assuming that the propensity score is *correctly specified*, we can estimate previous parameters simply by using the "sample equivalent" of the population parameters, that is:

$$\hat{\text{ATE}} = \frac{1}{N}\sum_{i=1}^{N} \frac{[w_i - \hat{p}(\mathbf{x}_i)]y_i}{\hat{p}(\mathbf{x}_i)[1 - \hat{p}(\mathbf{x}_i)]}$$

$$\hat{\text{ATET}} = \frac{1}{N}\sum_{i=1}^{N} \frac{[w_i - \hat{p}(\mathbf{x}_i)]y_i}{\hat{p}(w=1)[1 - \hat{p}_i(\mathbf{x})]}$$

$$\hat{\text{ATENT}} = \frac{1}{N}\sum_{i=1}^{N} \frac{[w_i - \hat{p}_i(\mathbf{x}_i)]y_i}{\hat{p}(w=0)\,\hat{p}(\mathbf{x}_i)}$$

Estimation follows in *two steps*: (i) estimate the propensity score $p(\mathbf{x}_i)$ getting $\hat{p}(\mathbf{x}_i)$; (ii) substitute $\hat{p}(\mathbf{x}_i)$ into the formulas to get the parameter. Observe that *consistency* is guaranteed by the fact that these estimators are M-estimators. How to get the standard errors for previous estimations? We can exploit some results from the case in which the first step is a ML-estimation and the second step is a M-estimation. In our case, the first step is a ML based on logit or probit and the second step is a standard M-estimator. For case like this, Wooldridge (2007; 2010, p. 922-924) has proposed a straightforward procedure to get standard errors, provided that the propensity score is *correctly specified*. In what follows, we set out the Wooldridge's procedure and formulas for obtaining these (analytical) standard errors.

*(i) Standard Errors estimation for* ATE

*First*: define the estimated ML-score of the first step (Probit or Logit). It is, by definition equal to:

$$\hat{\mathbf{d}}_i = \hat{\mathbf{d}}(w_i, \mathbf{x}_i, \hat{\boldsymbol{\gamma}}) = \frac{[\nabla_{\gamma}\hat{p}(\mathbf{x}_i, \hat{\boldsymbol{\gamma}})]' \cdot [w_i - \hat{p}(\mathbf{x}_i, \hat{\boldsymbol{\gamma}})]}{\hat{p}(\mathbf{x}_i, \hat{\boldsymbol{\gamma}})[1 - \hat{p}(\mathbf{x}_i, \hat{\boldsymbol{\gamma}})]}$$

Observe that $\mathbf{d}$ is a row-vector of the $R$-1 parameters $\boldsymbol{\gamma}$ and is the gradient of the function $p(\mathbf{x}, \boldsymbol{\gamma})$.
*Second*: define the generic estimated summand of ATE as:

$$\hat{k}_i = \frac{[w_i - \hat{p}(\mathbf{x}_i)]y_i}{\hat{p}(\mathbf{x}_i)[1 - \hat{p}(\mathbf{x}_i)]}$$