Economics Bulletin

**Volume 29, Issue 4**

# A note on the variance of average treatment effects estimators

Gabriel Montes-Rojas
*City University London*

## Abstract

We derive the variance of the Hirano, Imbens and Ridder (Econometrica 66, 315--31, 2003) average treatment effects estimator when the true propensity score is known. This variance is used in the derivation of the variance of a similar two-step estimator, where a M-estimator is used in the first step to estimate the propensity score.

I am in debt to Anil Bera and to an anonymous referee for valuable comments and suggestions.
**Citation:** Gabriel Montes-Rojas, (2009) "A note on the variance of average treatment effects estimators", *Economics Bulletin*, Vol. 29 no.4 pp. 2937-2943.
**Submitted:** Sep 04 2009. **Published:** November 24, 2009.

# 1 Introduction

Following the work of Hahn (1998) that derived the asymptotic semiparametric efficiency bound for average treatment effects (ATE) estimators, Hirano, Imbens and Ridder (2003) proved that weighting by the inverse of a nonparametric series estimator of the propensity score, rather than the true propensity score, lead to an efficient estimator of the ATE. Their estimator is rather simple when compared to alternative estimators in the literature, such as propensity score matching or treatment regression models. However, the fact that a particular nonparametric estimator of the propensity score is needed, which is rather computationally intensive, determine that their estimator has not been extensively used. For instance Wooldridge (2002, p.617) states that "as a practical matter, series estimation [of the propensity score] is not ideal, because for a binary response, it is identical to a linear probability model in functions of $\mathbf{x}$. Plus, it is difficult to estimate the asymptotic variance of the resulting estimators." Chen, Hong and Tarozzi (2008) prove that different combinations of nonparametric and parametric estimates of the propensity score have to be specifically derived to achieve the efficiency bounds. In practice, the propensity score is estimated using parametric models (such as logit or probit), and this estimate is used to construct other consistent estimators of ATE. Moreover, its variance is generally computed using bootstrap methods because, in many cases, an explicit derivation of the small sample or even asymptotic variance is difficult.

The goal of this paper is to study the Hirano *et al.* (2003) estimator when the propensity score is estimated by an M-estimator (e.g. maximum likelihood when the propensity score is correctly specififed) and to provide an explicit expression for the asymptotic variance of this estimator using the delta-method. This is a particular case of the Chen *et al.* (2008) inverse probability weighting based GMM estimators.

As an intermediate step we also derive the variance of the Hirano *et al.* (2003) estimator when the propensity score is known. Of course, in many empirical situations, the propensity score is not known. However, this has pedagogical importance, provided that the statement above seems paradoxical (Hahn 1998 proves that the propensity score is an ancilliary statistic to the ATE estimation). In fact, this answers the question: If the propensity score were known, what is the cost of using it in order to avoid its nonparametric series estimator?

Although the derivation developed in this paper is quite simple, its method can be applied to more complicated treatment effects estimators. For instance, similar steps can be used to obtain the asymptotic variance of the Firpo (2007) quantile treatment effects (QTE) estimator, where the conditional mean and variances are replaced by those of the influence function of the quantile functions, and the propensity score is estimated by a M-estimator instead of the nonparametric series estimator.

This paper is organized as follows. Section 2 discusses the assumptions used in the treatment effects literature. Section 3 derives the asymptotic variance of the Hirano *et al.* (2003) estimator when the true propensity score is known. Section 4 uses this variance to derive the asymptotic variance of an ATE estimator with a parametric estimate of the propensity score. Section 5 concludes.

# 2 Notation, definitions and assumptions

We follow the standard notation in Imbens (2004). Consider $N$ individuals indexed by $i = 1, 2, ..., N$ who may receive a "treatment", indicated by the binary variable $W_i = 0, 1$. Each individual has a pair of potential outcomes $(Y_{0i}, Y_{1i})$ that corresponds to the outcome with and without the treatment effect respectively. The fundamental problem, of course, is the inability to observe at the same time the same individual both with and without treatment effects. That is, we only observe $Y_i = W_i \times Y_{1i} + (1 - W_i) \times Y_{0i}$ and a set of exogenous variables $X_i$. Moreover define the propensity score as $p(X) = P[W = 1|X]$ and define $\mu_j(X) = E[Y_j|X]$ and $\sigma_j^2(X) = VAR[Y_j|X]$ for $j = 0, 1$.

We are interested in estimating the ATE of the W-treatment, defined as $\delta^* = E[Y_1 - Y_0]$ and following Hirano *et al.* (2003) we propose the estimator:

$$\hat{\delta} = \frac{1}{N}\sum_{i=1}^{N}\frac{W_i}{p(X_i)}Y_i - \frac{1}{N}\sum_{i=1}^{N}\frac{1-W_i}{1-p(X_i)}Y_i. \quad (1)$$

The standard assumptions in the treatment effects literature are

**Assumption 1.** $(Y_0, Y_1) \perp W|X$

**Assumption 2.** *For* $c \in (0, 1)$, $c < p(X) < 1 - c$

**Assumption 3.** $E[Y_0^2] < \infty$, $E[Y_1^2] < \infty$

# 3 ATE estimator with known true propensity score

Under these assumptions, the unbiasedness of this estimator can be easily proved following the results of Hirano *et al.* (2003), and therefore we omit that proof. We are more interested in showing that this estimator does not achieve the semiparametric efficiency bound (defined in Hahn 1998) and we explicitly quantify the loss of efficiency.

First note that by the i.i.d. set-up,

$$\lim_{N\to\infty} VAR[\sqrt{N}(\hat{\delta}-\delta)] = VAR\left[\frac{WY}{p(X)} - \frac{(1-W)Y}{1-p(X)}\right].$$

Expressing the variance as the sum of the variance of the conditional expectation and the expectation of the conditional variance, and using the assumptions above, it becomes

$$E\left[\frac{\sigma_1^2(X)}{p(X)} + \frac{\sigma_0^2(X)}{1-p(X)} + (1-p(X))p(X)\left(\frac{\mu_1(X)}{p(X)} + \frac{\mu_0(X)}{1-p(X)}\right)^2\right]$$
$$+VAR\left[\mu_1(X) - \mu_0(X)\right].$$

Therefore, the asymptotic variance can be written as

$$\lim_{N\to\infty} VAR\left[\sqrt{N}(\hat{\delta}-\delta)\right] = B + L, \tag{2}$$

where

$$B = E[\frac{\sigma_1^2(X)}{p(X)} + \frac{\sigma_0^2(X)}{1-p(X)} + (\delta^*(X) - \delta^*)^2]$$

is the semiparametric efficiency bound in Hahn (1998), $\delta^*(X) = \mu_1(X) - \mu_0(X)$, and $L$ is the loss of efficiency for using the true propensity score and not a nonparametric series estimator,

$$L = E\left[\left(\sqrt{\frac{1-p(X)}{p(X)}}\mu_1(X) + \sqrt{\frac{p(X)}{1-p(X)}}\mu_0(X)\right)^2\right],$$

which is nonnegative. Note that even if all the variables are constants in $X$, i.e. $p(X) = p$ and $\mu_1(X) = \mu_0(X) = \mu$, $L = \mu^2\left(2 + \frac{(1-p)^2+p^2}{p(1-p)}\right) \geq 0$, which has a minimum for $p = 1/2$ with $L = 4\mu^2$. Therefore, the incurred loss is potentially big if $\mu$ is.

# 4 Two-step ATE estimator

Following Hirano *et al.* (2003), $\hat{\delta}$ can be seen as a M-estimator with estimating equation

$$\psi(Y, W, X, p(X); \delta) = \frac{WY}{p(X)} - \frac{(1-W)Y}{1-p(X)} - \delta, \tag{3}$$

and therefore,

$$\lim_{N\to\infty} VAR[\sqrt{N}(\hat{\delta} - \delta^*)] = E[\psi(Y, W, X, p(X); \delta^*)^2] \tag{4}$$

Of course, after some algebra, the variance in eq. (4) is the same as that in eq. (2).

Now assume that $p(X) = \Phi(X; \gamma^*)$ and that $\gamma^*$ is unknown to the econometrician. $\Phi$ is used to denote a distribution function (not necessarily the normal c.d.f.). Moreover, assume that a $\sqrt{N}$-consistent estimator of $\gamma$ is $\hat{\gamma}$, which satisfy

$$\sqrt{N}(\hat{\gamma} - \gamma^*) = N^{-1/2}\sum_{i=1}^{N} s(W_i, X_i; \gamma^*) + o_p(1).$$

Any M-estimator of $\gamma$ can be framed in these terms under standard regularity conditions, including, of course, the maximum likelihood estimator.

Moreover, $\hat{\delta}$ satisfies

$$\sqrt{N}(\hat{\delta} - \delta^*) = N^{-1/2}\sum_{i=1}^{N} \zeta(Y_i, W_i, X_i; \gamma^*, \delta^*) + o_p(1),$$

where $\psi(., p(.); \delta^*) = \zeta(.; \gamma^*, \delta^*)$. Therefore, the asymptotic variance of $\hat{\hat{\delta}}$, the two-step estimator of $\delta^*$ where $\hat{\gamma}$ is used to construct the propensity score instead of $\gamma^*$, can be obtained using the delta-method (see Wooldridge 2002, ch.12),

$$\lim_{N\to\infty} VAR[\sqrt{N}(\hat{\hat{\delta}}-\delta^*)] = VAR[\zeta(Y,W,X;\gamma^*,\delta^*)+H^*(W,X;\gamma^*,\delta^*)s(W,X;\gamma^*)],$$

where

$$H^*(\gamma^*,\delta^*) = E[\nabla_\gamma \zeta(Y,W,X;\gamma^*,\delta^*)] = -E\left[\left(\frac{\mu_1(X)}{\Phi(X;\gamma^*)} + \frac{\mu_0(X)}{1-\Phi(X;\gamma^*)}\right)\nabla_\gamma \Phi(X;\gamma^*)\right].$$

Finally, note that

$$\begin{aligned}
& VAR[\sqrt{N}(\hat{\hat{\delta}}-\delta^*)] \\
= \;& E[\zeta(Y,W,X;\gamma^*,\delta^*)^2] + H^*(\gamma^*,\delta^*)E[s(W,X;\gamma^*)s(W,X;\gamma^*)']H^*(\gamma^*,\delta^*)' \\
& +2\times COV[\zeta(Y,W,X;\gamma^*,\delta^*), H^*(\gamma^*,\delta^*)s(W,X;\gamma^*)] \\
= \;& VAR[\sqrt{N}(\hat{\delta}-\delta^*)] + H^*(\gamma^*,\delta^*)VAR[\sqrt{N}(\hat{\gamma}-\gamma^*)]H^*(\gamma^*,\delta^*)' \\
& +2\times COV[\zeta(Y,W,X;\gamma^*,\delta^*), H^*(\gamma^*,\delta^*)s(W,X;\gamma^*)].
\end{aligned}$$

Then, the asymptotic variance of the two-step estimator can be re-written as

$$VAR[\sqrt{N}(\hat{\hat{\delta}}-\delta^*)] = B + L + G, \tag{5}$$

where $B$ and $L$ were defined in Section 3 and $G = H^*(\gamma^*,\delta^*)VAR[\sqrt{N}(\hat{\gamma}-\gamma^*)]H^*(\gamma^*,\delta^*)' + 2\times COV[\zeta(Y,W,X;\gamma^*,\delta^*), H^*(\gamma^*,\delta^*)s(W,X;\gamma^*)]$. Therefore $L+G$ can be seen is the efficiency loss arising from using a two-step estimator of the propensity score instead of the series estimator.

In practice, a consistent estimator of this variance can be obtained by the OPG method:

$$Est.VAR[\sqrt{N}(\hat{\hat{\delta}}-\delta^*)] = \frac{1}{N}\sum_{i=1}^{N}\left(\zeta(Y_i,W_i,X_i;\hat{\gamma},\hat{\hat{\delta}}) + \hat{H}(\hat{\gamma},\hat{\hat{\delta}})s(W_i,X_i;\hat{\gamma})\right)^2,$$

where

$$\hat{H}(\hat{\gamma},\hat{\hat{\delta}}) = -\frac{1}{N}\sum_{i=1}^{N}\left(\left(\frac{W_iY_i}{\Phi^2(X_i;\hat{\gamma})} + \frac{(1-W_i)Y_i}{(1-\Phi(X_i;\hat{\gamma}))^2}\right)\nabla_\gamma \Phi(X_i;\hat{\gamma})\right).$$

## 5 Conclusion

This note derives the variance of a simple ATE estimator using the known true propensity score. Moreover, it also derives its variance if a M-estimator is used to estimate the propensity score. This ATE estimator, where the propensity score is estimated by probit or logit models, is widely used in the empirical literature on treatment effects. However, despite its simple derivation, no explicit formulation of the asymptotic variance was given elsewhere in the literature.

## References


Chen, X., H. Hong and A. Tarozzi (2008) "Semiparametric efficiency in GMM models with auxiliary data" *Annals of Statistics* **36**, 808–43.

Firpo, S. (2007) "Efficient semiparametric estimation of quantile treatment effects" *Econometrica* **75**, 259-76.

Hahn, J. (1998) "On the role of the propensity score in efficient semiparametric estimation of average treatment effects" *Econometrica* **66**, 315-31.

Hirano, K., G.W. Imbens and G. Ridder (2003) "Efficient estimation of average treatment effects using the estimated propensity score" *Econometrica* **71**, 1161-90.

Imbens, G.W. (2004) "Nonparametric estimation of average treatment effects under exogeneity: a review" *Review of Economics and Statistics* **86**, 4-29.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press: Cambridge, Massachusetts.