# Inference of a non-parametric covariate-adjusted variable importance measure of a continuous exposure

Cabral Amilcar Chanang Tondji

# Inference of a non-parametric covariate-adjusted variable importance measure of a continuous exposure

Cabral Amilcar Chanang Tondji
LAMA, Université Paris Est Marne La Vallée

June 23, 2016

**Abstract**

We consider a setting where a real-valued variable of cause $X$ affects a real-valued variable of effect $Y$ in the presence of a context variable $W$. The objective is to assess to what extent $(X, W)$ influences $Y$ while making as few assumptions as possible on the unknown distribution of $O = (W, X, Y)$. Based on a user-supplied marginal structural model, our new variable importance measure is non-parametric and context-adjusted. It generalizes the variable importance measure introduced by Chambaz et al. [4]. We show how to infer it by targeted minimum loss estimation (TMLE), conduct a simulation study and present an illustration of its use.

## 1 Introduction

**The setting.** Consider the situation where a real-valued variable of cause, $X \in \mathbb{R}$, affects a $[0, 1]$-valued variable of effect, $Y$, in the presence of a variable of context $W \in \mathcal{W}$. The objective is to assess to what extent $(X, W)$ influences $Y$ while making as few assumptions as possible on the unknown distribution $P_0$ of $O = (W, X, Y)$. This requires both the definition of a tailored statistical parameter and the elaboration of a semi-parametric inferential procedure to construct confidence intervals of a given asymptotic level based on independent copies of $O$ drawn from $P_0$.

Marginal structural models (MSMs) are very useful tools in this regard. Let $\{\mathrm{msm}_\beta : \beta \in B\}$ be such a class of functions mapping $\mathbb{R} \times \mathcal{W}$ to $\mathbb{R}$. It is associated with a parameter defined as a minimizer in $\beta$ of the real-valued criterion

$$E_{P_0}\left([Y - E_{P_0}(Y|X = x_0, W) - \mathrm{msm}_\beta(X, W)]^2\right),$$

where $x_0$ is a reference value for $X$ for which there exists $0 < c < 1/2$ such that $P_0(X \neq x_0|W) \in [c, 1 - c]$ $P_0$-almost surely. For instance, choosing a MSM with $B = \mathbb{R}$ and $\mathrm{msm}_\beta$ given by $\mathrm{msm}_\beta(X, W) = \beta(X - x_0)$ yields the non-parametric variable importance measure studied in [2, 4]. This statistical parameter measures the effect of $X$ on $Y$ accounting for $W$ (the conditional expectation in the definition of the parameter is conditional on $X$ and $W$) but averaging out $W$ eventually ($\mathrm{msm}_\beta(X, W)$ does not depend on $W$). For technical reasons, we focus on MSMs of the form

$$\{(X, W) \mapsto (X - x_0)f_\beta(W) : \beta \in B\}, \tag{1}$$

where $f_\beta$ is linear in $\beta$. The statistical parameter of interest is formally defined as

$$\psi_0 = \arg\min_{\beta \in B} E_{P_0}\left([Y - E_{P_0}(Y|X = x_0, W) - (X - x_0)f_\beta(W)]^2\right), \tag{2}$$

assuming that the minimizer exists and is unique. We interpret $(X - x_0)f_{\psi_0}(W)$ as the best approximation of the form $(X - x_0)f_\beta(W)$ to $(E_{P_0}(Y|X, W) - E_{P_0}(Y|X = x_0, W))$. It quantifies the influence of $X$ on $Y$, using $x_0$ as a reference value, while accounting for the covariates $W$ on a linear scale.

**Relevant literature.** Our main sources of inspiration are [2, 4]. These articles were motivated by an application to the analysis of the effect of DNA copy number variations on gene expression

accounting for DNA methylation. Similar to the parameter defined and studied in [2, 4], our parameter of interest (2) belongs to the family of variable importance measure which was introduced in [12].

Contrary to [1, 5, 6, 11, 15], [2, 4] do not assume that the real-valued variable of cause $X$ is discrete (or do not discretize it), but rather exploit the fact that $X$ has a reference value $x_0$ and features a continuum of other values. We also avoid discretizing $X$ and make the same assumption on its conditional distribution given $W$. Contrary to [8, 9], [2, 4] do not assume a semiparametric model but rather exploit one [see 4, end of Section 1 for a discussion]. Following [2, 4], we do too exploit MSM (1) but do not assume that $E_{P_0}(Y|X,W) - E_{P_0}(Y|X = x_0, W)$ belongs to it.

Our main contribution is that our $\psi_0$ (2) points to an element of MSM (1) such that $E_{P_0}(Y|X,W) - E_{P_0}(Y|X = x_0, W)$ is best approximated by $(X - x_0)f_{\psi_0}(W)$. In words, $W$ is not averaged out completely like in [2, 4]. Instead, the effect of $X$ on $Y$ is quantified as $(X - x_0)$ times a function of the linear expression $f_{\psi_0}(W)$ of $W$. We later give a causal interpretation to $\psi_0$. Since the functional $\Psi$ defined ad hoc (3) so that $\psi_0$ be the value of $\Psi$ at $P_0$ is pathwise differentiable, we can carry out the inference of $\psi_0$ by targeted minimum loss estimation (TMLE) [2, 4, 13, 14].

**Organization.** Section 2 defines and studies the functional $\Psi$ (3) mentioned in the previous paragraph. Section 3 describes the inference procedure tailored to the construction of confidence intervals of a given asymptotic level for $\psi_0$. Section 4 presents the results of a simulation study. Section 5 gives an illustration based on real data on climate change. Relevant materials and proofs are gathered in the appendix.

## 2 Studying the parameter of interest

Without loss of generality, we assume from now on that $x_0 = 0$.

**Differentiability and robustness.** We denote $\dot{f} = \frac{\partial}{\partial\beta} f_\beta$ the gradient of $f_\beta$ which, by choice, does not depend on $\beta$. Denote $d$ the dimension of the space where $\dot{f}(W)$ lives. For every possible data-generating distribution $P$ of $O$, we denote $\theta(P)(X,W) = E_P(Y|X,W)$ and $g(P)(W) = P(X \neq 0|W)$. Assume that $P$ is chosen such that

1. $\mu(P)(W) = E_P(X\dot{f}(W)|W)$ and $\Sigma(P) = E_P[X^2\dot{f}(W)^\top\dot{f}(W)]$ are well-defined features of $P$;
2. $\Sigma(P)$ is invertible;
3. there exists $c \in ]0, 1/2[$ such that $g(P)(W) \in [c, 1-c]$ $P$-almost surely.

Conditions 1, 2, 3 concern the joint distribution of $(X, \dot{f}(W))$. The two first ones are met if $X\dot{f}(W)$ is a bounded random variable and if there is no deterministic linear combination of the components of $X\dot{f}(W)$ which equals 0 $P$-almost surely. For such a $P$, the equation

$$\Psi(P) = \underset{\beta\in B}{\arg\min}\, E_P\left([\theta(P)(X,W) - \theta(P)(0,W) - Xf_\beta(W)]^2\right) \tag{3}$$

uniquely characterizes a parameter of $P$ such that $\Psi(P_0) = \psi_0$, if $P_0$ meets the constraints, which we assume from now on to be true. It is easily seen that $\Psi(P)$ rewrites

$$\Psi(P) = \Sigma(P)^{-1}[E_P(X\dot{f}(W)(\theta(P)(X,W) - \theta(P)(0,W)))]. \tag{4}$$

The functional $\Psi$ is pathwise differentiable at $P$, with an efficient influence curve given by $D^\star(P) = D_1^\star(P) + D_2^\star(P)$ where $D_1^\star(P)$ and $D_2^\star(P)$ are two $L_0^2(P)$-orthogonal components characterized by

$$\begin{aligned} D_1^\star(P)(O) &= \Sigma(P)^{-1}\left[(\theta(P)(X,W) - \theta(P)(0,W) - Xf_\beta(W))\right]X\dot{f}(W), \text{ and} \\ D_2^\star(P)(O) &= \Sigma(P)^{-1}\left[(Y - \theta(P)(X,W))\left(X\dot{f}(W) - \frac{1_{\{X=0\}}}{g(P)(0|W)}\mu(P)(W)\right)\right]. \end{aligned}$$

This means that, for any bounded $s \in L_0^2(P)$ taking values in $\mathbb{R}^d$ and $\varepsilon \in \mathbb{R}^d$ with $\|\varepsilon\|_\infty < \|s\|_\infty^{-1}$, if we characterize a data-generating distribution $P_\varepsilon$ of $O$ by setting

$$\frac{dP_\varepsilon}{dP}(O) = 1 + \varepsilon^\top s(O),$$

then for $\varepsilon$ small enough, $P_\varepsilon$ meets Conditions 1, 2, 3, hence $\Psi(P_\varepsilon)$ is well-defined, and moreover $\varepsilon \mapsto \Psi(P_\varepsilon)$ is differentiable at $\varepsilon = 0$ with a derivative satisfying

$$\lim_{\varepsilon \to 0} \frac{\Psi(P_\varepsilon) - \Psi(P)}{\varepsilon} = E_P[s(O)^\top D^\star(P)(O)].$$

The efficient influence curve $D^\star(P)$, of $\Psi$ at $P$, enjoys a remarkable "double-robustness" property: if $P, P'$ are two data-generating distributions of $O$ satisfying Conditions 1, 2, 3 and such that $E_P(D^\star(P')(O)) = 0$, then $\Psi(P) = \Psi(P')$ whenever $\theta(P')(0, \cdot) = \theta(P)(0, \cdot)$ or ($\mu(P) = \mu(P')$ and $g(P') = g(P)$).

The validity of all the statements we make in this section can be checked easily by adapting, *mutatis mutandis*, the proofs of similar statements in [4].

**Causal interpretation.** We now present a causal interpretation to $\Psi(P)$, which partly relies on untestable assumptions. Assume, in this section only, that there exists a collection $(Y_x)_{x \in \mathbb{R}}$ of random variables such that *(i)* $(Y_x)_{x \in \mathbb{R}} \perp X|W$ (randomization assumption), and *(ii)* $Y = Y_X$ (consistency assumption). The above holds for instance in the following structural equation model: *there exists three deterministic functions $f_W, f_X, f_Y$ and three independent random variables $U_W, U_X, U_Y$ such that $W = f_W(U_W), X = f_X(W, U_X)$ and $Y = f_Y(W, X, U_Y)$*. In addition, assume that the conditional laws of $X$ given $W$ are all dominated by a common measure $\mu$. Then, there exists a collection of conditional densities $\phi(\cdot|W)$ of $X$ given $W$, all with respect to $\mu$.

Let us denote by $\mathbb{P}$ the law of the full data $(W, X, (Y_x)_{x \in \mathbb{R}})$. It obviously holds that $E_P(Y|X = x, W) = E_{\mathbb{P}}(Y_x|X = x, W) = E_{\mathbb{P}}(Y_x|W)$, by independence of $Y_x$ and $X$. Furthermore, for each $\beta \in B$,

$$\begin{aligned} &E_P\{(E_P(Y|X, W) - E_P(Y|X = 0, W) - X f_\beta(W))^2\} \\ &\qquad = \int E_P\left[(E_{\mathbb{P}}(Y_x - Y_0 - x f_\beta(W)|W))^2 \phi(x|W)\right] \mu(dx). \quad (5) \end{aligned}$$

Thus, $\Psi(P)$ can be interpreted as the coefficient associated with the regression of $Y_x$ on $Y_0 + f_\beta(x, W)$ based on a weighted $L^2$-loss function.

# 3 Inference

We infer $\psi_0$ by TMLE. The first step consists in constructing initial estimators of some relevant features of $P_0$. The second step consists in iteratively updating them until convergence.

**Initialization.** The initialization consists in estimating the following features of $P_0$: marginal distribution of $W$, $\mu(P_0)$, $g(P_0)$, $\theta(P_0)$, $\Sigma(P_0)$ and, for each of them, a companion feature required to update them at the next step [see 4, Lemma 1]. We denote $P_n^0$ a data-generating distribution chosen such that *(i)* each estimator $\eta_n$ of a feature $\eta(P_0)$ among the above features of interest can be rewritten $\eta_n = \eta(P_n^0)$, and *(ii)* we can sample $(W, X)$ from $P_n^0$. As soon as we have built estimators of the marginal distribution of $W$, $\mu(P_0)$, $g(P_0)$, $\theta(P_0)$ and $\Sigma(P_0)$, we can also estimate $\psi_0$ and $D^\star(P_0)$. The estimation of $\psi_0$ is performed by Monte-Carlo simulation: we simulate $B$ independent random variables $(W^{(0,b)}, X^{(0.b)})$ from the marginal joint distribution of $(W, X)$ under $P_n^0$, then compute

$$\psi_n^0 = B^{-1} \sum_{b=1}^{B} \Sigma(P_n^0)^{-1} \left[ X^{(0,b)} \dot{f}(W^{(0,b)}) \left( \theta(P_n^0)(X^{(0,b)}, W^{(0,b)}) - \theta(P_n^0)(0, W^{(0,b)} \right) \right].$$

**Iterative updating.** Say we have built $(k-1)$ updates $P_n^1, \ldots, P_n^{k-1}$ of $P_n^0$. The $k$th update goes as follows. Set $0 < \rho < 1$ a constant close to 1, for instance $\rho = 0.99$ and, for each $\varepsilon \in \mathbb{R}^d$,

| MSM | $f_\beta(W_1, W_2) = \beta_1 W_1 + \beta_2 W_2$ | | $f'_\beta(W_1, W_2) = \beta_1 W_1 + \beta_2 W_1^2$ | |
|---|---|---|---|---|
| $n$ | $\psi_{0,1} = 0.56$ | $\psi_{0,2} = 0$ | $\psi_{0,1} = 1.53$ | $\psi_{0,2} = -1.42$ |
| 1000 | 94.6% | 95.0% | 95.9% | 94.5% |
| 2000 | 93.9% | 93.9% | 95.6% | 93.7% |

Table 1: Summary of the results of the simulation study. The values of the true parameters $\psi_0$ and $\psi'_0$ are reported in the second row. The third and fourth row give the empirical coverage of the regions of confidence for each coordinate of $\beta$ and each sample size $n$.

$\|\varepsilon\|_\infty \leq \rho \|D^\star(P_n^{k-1})\|_\infty$, introduce $P_n^{k-1}(\varepsilon)$ given by

$$\frac{dP_n^{k-1}(\varepsilon)}{dP_n^{k-1}}(O) = 1 + \varepsilon^\top D^\star(P_n^{k-1})(O)$$

where $D^\star(P_n^{k-1})(O)$ is the current estimator of the efficient influence curve. This defines a $d$-dimensional parametric model through $P_n^{k-1}$ fluctuating it in the direction of $D^\star(P_n^{k-1})$. We let $\varepsilon_n^{k-1}$ be the maximum likelihood estimator of $\varepsilon$ in this model and characterize the $k$th update as $P_n^k = P_n^{k-1}(\varepsilon_n^{k-1})$. This yields updated estimators of the features of interest in the spirit of [4, Lemma 1]. The corresponding $k$th update of $\psi_n^0$ is obtained by simulating $B$ independent random variables $(W^{(k,b)}, X^{(k,b)})$ from the marginal joint distribution of $(W, X)$ under $P_n^k$ then computing

$$\psi_n^k = B^{-1} \sum_{b=1}^{B} \Sigma(P_n^k)^{-1} \left[ X^{(k,b)} \dot{f}(W^{(k,b)}) \left( \theta(P_n^k)(X^{(k,b)}, W^{(k,b)}) - \theta(P_n^k)(0, W^{(k,b)}) \right) \right]. \quad (6)$$

**Central limit theorem.** Suppose that performing $k_n$ iterations of the updating procedure guarantees that $\|P_n D^\star(P_n^{k_n})\|_\infty = o_P(1/\sqrt{n})$. Suppose moreover that there exists a function $f_1$ with $P_0 f_1 = 0$ such that $\|P_0(D^\star(P_n^{k_n}) - f_1)(D^\star(P_n^{k_n}) - f_1)^\top\|_\infty = o_P(1)$, and that $\|\Psi(P_n^{k_n}) - \psi_0 - P_0 D^\star(P_n^{k_n})\|_\infty = o_P(1/\sqrt{n})$. In addition, suppose that $S_n$ estimates consistently $E_{P_0}[f_1(O) f_1(O)^\top]$. Then $\psi_n^* = \Psi(P_n^{k_n})$ satisfies $\sqrt{n}(\psi_n^\star - \psi_0) = (P_n - P_0) f_1 + o_P(1)$, hence $\sqrt{n} S_n^{-1/2}(\psi_n^\star - \psi_0)$ converges in law to the $d$-multivariate Gaussian law with zero mean and identity covariance matrix. We refer the reader to [4, appendix] for the proof of a similar result.

# 4 Simulation study

**Simulation scheme and implementation of procedure.** We essentially rely on the same synthetic data-generating distribution $P^s$ as in [4, Section 6.4]. A random variable drawn from $P^s$ takes the form $(W_1, X, Y)$ with $W_1 \in [0, 1]$. We enrich it by augmenting the baseline covariate with $W_2$ drawn from the standard normal distribution independently of $(W_1, X, Y)$. We write $O = (W = (W_1, W_2), X, Y)$ the resulting complete observation.

We adapt the `R` package [3]. This is possible because we imposed that $f_\beta(W)$ be linear in $\beta$.

**Results of the simulation study.** We actually consider two choices of MSM: one based on $f_\beta(W) = \beta_1 W_1 + \beta_2 W_2$, the other based on $f'_\beta(W) = \beta_1 W_1 + \beta_2 W_1^2$. The evaluation of the values of the corresponding true parameters $\psi_0$ and $\psi'_0$ (see the second row of Table 1) was performed by Monte-Carlo based on (4). For each choice of MSM, independently, we repeated $B = 1000$ times independently the simulation of a data set of sample size $n = 1000$ and the simulation of another data set of sample size $n = 2000$. We applied the TMLE procedure described in Section 3 to each data set, with the same choice of the fine-tune parameters as in [4] and with the option `flavor="learning"`.

The results are summarized in Table 1. The empirical coverage is satisfying.

# 5 Illustration

It is commly agreed today that human activity has a significant impact on climate. Among others, IPCC (Intergovernmental Panel on Climate Change) has been conducting an exhaustive study on

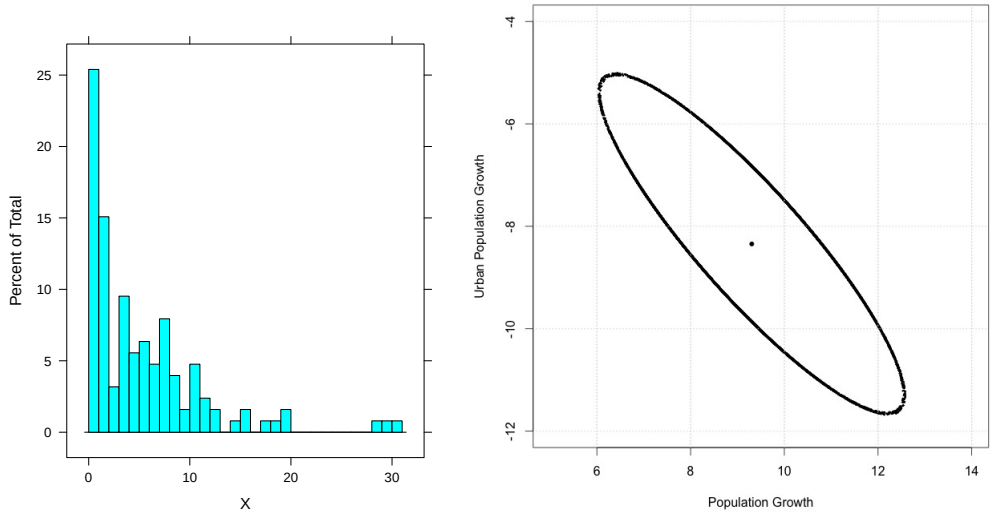


Figure 1: Left: Histogram of the variable $X$. Right: Confidence region of asymptotic level 95% for parameter $(\psi_{0,2}, \psi_{0,3})$.

the topic for decades. In particular, the effect of $CO_2$ emissions on climate change is now well understood [7, 10]. However, one of the major remaining challenge is to understand which factors are driving climate change. We illustrate the interest of our parameter and its inference in this setting.

We exploit a publicly available data set of the World Bank[1]. We extract from it our data set. It consists of $n = 126$ observed data-structures $O_1, \ldots, O_i = (W_i, X_i, Y_i), \ldots, O_n$ where, for the $i$th country,

- $W_i$ gathers its under-five mortality rate, population growth, urban population growth, $CO_2$ emissions per unit of Gross Domestic Product (GDP), energy use per unit of GDP, energy use per capita for the year 1998;
- $X_i$ is a thresholded version of total amount of $CO_2$ emissions per capita for the year 1998;
- $Y_i$ is the 10%-quantile of the projected annual temperature change for the period 2045–2065.

Under-five mortality rate is a reliable indicator of poverty. Population growth and urban population growth are relevant indicators of economical development. $CO_2$ emissions per unit of GDP is an indicator of industrialization and reliance on fossil fuel. Finally, energy use per unit of GDP and per capita reveal patterns of energy comsumption by the industry and by the country's inhabitants.

All the $X_i$ are non-negative. The empirical distribution is represented in the left plot of Figure 1. We set to $x_0 = 0$ exactly all the $X_i$s smaller than 0.99, the 25%-quantile of the empirical distribution of $X$.

We assume that $O_1, \ldots, O_n$ are independently drawn from a common distribution $P_0$. We infer $\psi_0 = \Psi(P_0)$ given by (4) for the MSM $\{(X, W) \mapsto X f_\beta(W) : \beta \in \mathbb{R}^6\}$ with $f_\beta(W) = \beta^\top W$.

Using the asymptotic normality of the TMLE $\psi_n^*$, we carry out Student tests of "$\psi_{0,k} = 0$" against "$\psi_{0,k} \neq 0$" for $k = 1, \ldots, 6$. We reject the null for its alternative at level 5% only for $k = 2, 3$, *i.e.*, for population growth and urban population growth, with $p$-values respectively equal to $3.69 \times 10^{-10}$ and $2.01 \times 10^{-8}$. The corresponding estimates are $\psi_{n,2}^* = 9.30 \pm 1.33$ and $\psi_{n,3}^* = -8.34 \pm 1.35$, see also the right plot in Figure 1. In other words, we estimate $f_{\psi_0}(W)$ with $f_{\psi_n^*}(W) \approx 9.30 \times W_2 - 8.34 \times W_3$.

[1] http://data.worldbank.org/data-catalog/climate-change

We conclude that there is a significant effect of population growth and urban population growth on the relationship between climate change and $CO_2$ emissions per capita. This does not come as a surprise. The greater the population, especially in urban areas, the more energy (very possibly fossil energy with high level of $CO_2$ emissions) is produced, hence contributing intensively to the overall climate change. Urban population growth is a considerable factor as well, since it is directly linked to the above point.

**Remark.** We have carried out the same study with $(W, X)$ corresponding to the years 1990 to 1997. The results of inference and subsequent conclusions were very similar to those presented here (results not shown).

**Acknowledgements**

I acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-13-BS01-0005 (project SPADRO). I would like to thank Antoine Chambaz (Modal'X) and Hulda Nana for their support.

# A Appendix: proof of (5)

The following series of equalities proves (5), where the third one is a consequence of Fubini's theorem:

$$\begin{aligned}
&E_P\{(E_P(Y|X,W) - E_P(Y|X=0,W) - Xf_\beta(W))^2\} \\
&\quad = E_P\left\{E_P\left((E_P(Y|X,W) - E_P(Y|X=0,W) - Xf_\beta(W))^2|W\right)\right\} \\
&\quad = E_P\left\{\int (E_{\mathbb{P}}(Y_x|W) - E_{\mathbb{P}}(Y_0|W) - xf_\beta(W))^2\phi(x|W)\mu(dx)\right\} \\
&\quad = \int E_P\left[(E_{\mathbb{P}}(Y_x|W) - E_{\mathbb{P}}(Y_0|W) - xf_\beta(W))^2\phi(x|W)\right]\mu(dx) \\
&\quad = \int E_P\left[(E_{\mathbb{P}}(Y_x - Y_0 - xf_\beta(W)|W))^2\phi(x|W)\right]\mu(dx).
\end{aligned}$$

# References

[1] J. Andrews, W. Kennette, J. Pilon, A. Hodgson, A. B. Tuck, A.F. Chambers, , and D. I. Rodenhiser. Multi-platform whole-genome microarray analyse refine the epigenetic signature of breast cancer metastasis with gene expresion and copy number. *PLos ONE*, 5(1):e8665, 2010.

[2] A. Chambaz and P. Neuvial. tmle.npvi: targeted, integrative search of associations between DNA copy number and gene expression, accounting for DNA methylation. *Bioinformatics*, 31 (18):3054–3056, 2015.

[3] A. Chambaz and P. Neuvial. TMLE.NPVI. *http://CRAN.R-project.org/package=tmle.npvi*, 2015. R package 0.10.0.

[4] A. Chambaz, P. Neuvial, and M. J. van der Laan. Estimation of a non-parametric variable importance measure of a continuous exposure. *Electron. J. Stat.*, 6:1059–1099, 2012.

[5] R. Louhimo and S. Hautaniemi. CNAmet: an R package for integrating copy number, methylation and expression data. *Bioinformatics*, 27(6):887, 2011.

[6] J.R. Pollack, T. Sorlie, C. A. Rees, S.S. Jeffrey, P.E. Lonnin, R. Tbishirnani, D. Botstein, A.-L. Borresen-Dale, and P.O. Brown. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci USA*, 99(20):12963–12968, 2002.

[7] V. Ramanathan and G. Carmichael. Global and regional climate changes due to black carbon. *Nature Geoscience*, 1:221–227, 2008.

[8] J. M. Robins and A. Rotnitzky. Comment on inference for semiparametric models : some questions and an answer by Bickel, P. J. and Kwon, J. *Statistica Sinica*, 11:920–935, 2001.

[9] J. M. Robins, S. D. Mark, and W. K. Newey. Estimating exposue effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48(4):479–495, 1992.

[10] S. Solomon, G-K. Plattner, R. Knutti, and P. Friedlingstein. Irreversible climate change due to carbon dioxide emissions. *PNAS*, 106(6):1704–1709, 2008.

[11] Z. Sun, Y.W. Asmann, K.R. Kalari, B. Bot, J.E. Eckel-Passow, T.R. Baker, J.M. Carr, I. Khrebtukova, S. Luo, and L. et al. Zhan. Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. *PLoS One*, 6 (2):e17490, 2011.

[12] M. J. van der Laan. Statistical inference for variable importance. *Int. J. Biostat.*, 2, 2006. Article 2.

[13] M. J. van der Laan and D. Rubin. Targeted maximum likelihood learning. *Int. J. Biostat.*, 2, 2006. Article 11.

[14] M.J. van der Laan and S. Rose. *Targeted Learning : Causal Inference for Observational and Experimental Data.* Springer Series in Statistics. Springer Verlag, 2011.

[15] W. N. van Wieringen and M. A. van de Wiel. Non parametric testing for DNA copy number induced differential mRNA gene expression. *Biometrics*, 65(1):19–29, 2008.