Inverse probability weighting to estimate causal effects of sequential treatments: a latent class extension to deal with unobserved confounding

Leonardo Grilli

Department of Statistics

University of Florence

email: grilli@ds.unifi.it

Francesco Bartolucci, Luca Pieroni

Department of Economics, Finance and Statistics

University of Perugia

27th IWSM, July 16 - 20, 2012 - Prague

Outline

- *Marginal Structural Models* for causal inference in a longitudinal setting (sequential treatment)
- Estimation via *Inverse Probability-to-treatment Weighting* (IPW)
- Latent class extension to deal with unobserved confounding (LC-IPW)
- Simulation study: IPW vs LC-IPW
- Application: effect of wage subsidies on employment (Finnish firms)

The context

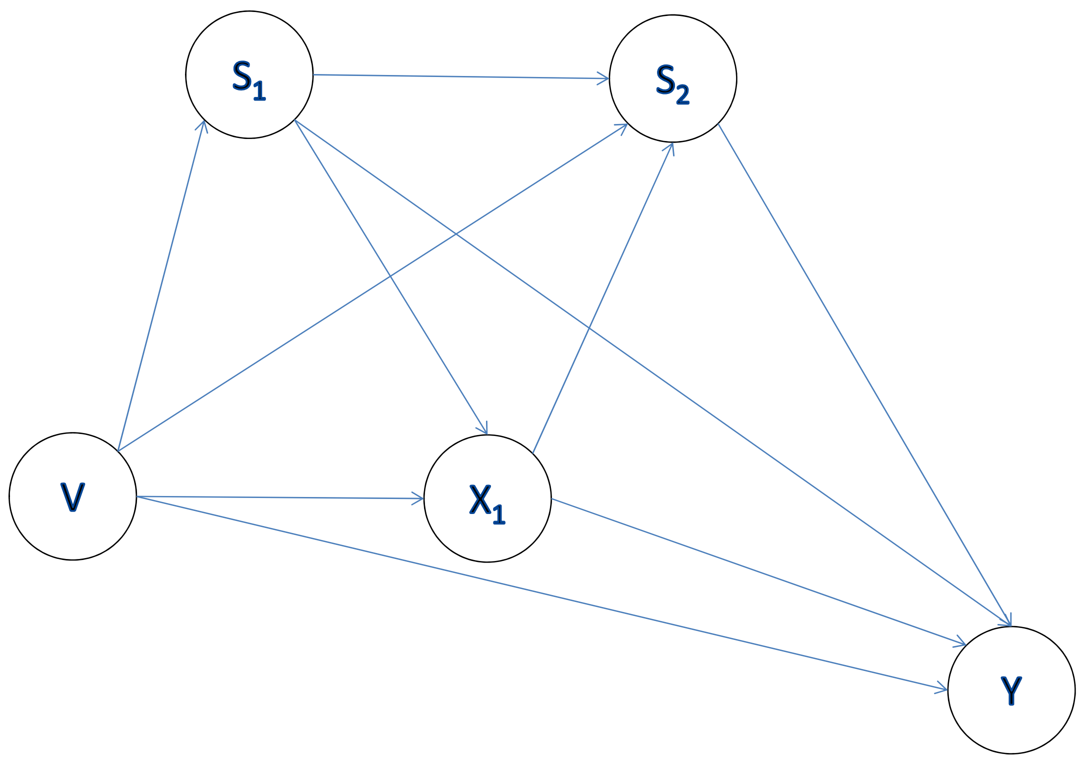
- Longitudinal data with several occasions (time points or intervals)
- Wish to assess the *causal effect* of a *sequential treatment* on an outcome measured at the end of the period
- Treatment assignment at a given occasion may depend on the sequence of previous assignments, as well as on *time-varying confounders* (i.e. variables affecting both treatment assignment and outcome)

Basic notation

- random sample of n subjects
- $t = 1, \ldots, T$ measurement occasions (time points or intervals)
- Y: outcome (measured after the last occasion)
- S_t: binary indicator of treatment at occasion t, with $\boldsymbol{S}_{1:t} = (S_1, \ldots, S_t)'$
- $\boldsymbol{V}$ column vector of pre-treatment covariates (measured before the first occasion)
- $\boldsymbol{X}_t$ column vector of time-varying covariates at occasion t, with $\boldsymbol{X}_{1:t} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_t')'$

Causal DAG

Two occasions with a pre-treatment observed confounder V and a time-varying confounder X_1



PROBLEM: *should condition* on X_1 because it is a confounder, *should not condition* on X_1 because it is a post-treatment variable

Marginal structural models (MSM)

- A solution to adjust for (observed) time-varying confounders: *Marginal Structural Models (MSM)* + *Inverse Probability-to-treatment Weighting (IPW)* (Robins, Hernan and Brumback, 2000)

- The framework is based on *potential outcomes* $Y^{(\boldsymbol{s}_{1:T})}$ (with Y denoting the $observed\ outcome$)

  T binary treatments $\Rightarrow 2^T$ potential outcomes

- A natural specification of a MSM is

$$E(Y^{(\boldsymbol{s}_{1:T})}) = \beta_0 + \boldsymbol{g}(\boldsymbol{s}_{1:T})'\boldsymbol{\beta}_1$$

  ▷ For example, $\boldsymbol{g}(\boldsymbol{s}_{1:T}) = s_+ = \sum_t s_t$, in this case a single parameter β_1 represents the *average causal effect of the treatment*

Inverse probability-to-treatment weighting (IPW)

- The causal parameters of a MSM can be consistently estimated using a *weighted regression* (IPW package in R)

- Each subject i is weighted by the inverse of the probability of its observed treatment sequence:

$$w_i = \frac{1}{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i)}$$

- Probabilities estimated through a pooled logistic regression (a standard logistic regression applied to the subject-occasion dataset)

- Higher efficiency is obtained with *stabilized weights*:

$$sw_i = \frac{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1})}{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i)}$$

Identification assumptions

- *Stable Unit Treatment Value Assumption (SUTVA)* $\Rightarrow$ no interference among units
- *Positivity* or *Random assignment*: the conditional probability of being assigned to treatment is neither zero nor one
- *Sequential Ignorability Assumption (SIA)*: conditionally on the observed history up to occasion $t-1$, the treatment assignment at occasion t is independent of the potential outcomes

$$S_t \perp \boldsymbol{Y}^{(all)} \mid \boldsymbol{S}_{1:t-1}, \boldsymbol{X}_{1:t-1}, \boldsymbol{V} \qquad t = 1, \ldots, T.$$

Unobserved confounding

- Often some of the confounders are unobserved
- The IPW estimator is no more consistent in case of unobserved confounders due to violation of the Sequential Ignorability Assumption (SIA)
- We extend the IPW method to derive a consistent estimator of causal effects in the presence of a *pre-treatment unobserved confounder U*
- We assume that U is a discrete variable with values $c = 1, \ldots, k$ corresponding to *latent classes*
- The number of latent classes k and their probabilities $\pi_c = Pr(U = c)$ are parameters to be estimated $\Rightarrow$ the approach is

flexible enough to satisfactorily approximate also continuous unobserved confounders

- We relax the ignorability assumption (SIA) by requiring that the independence holds within the latent classes induced by the unobserved confounder $U \Rightarrow$ *Latent Class Sequential Ignorability Assumption (LC-SIA)*:

$$S_t \perp \boldsymbol{Y}^{(all)} \mid \boldsymbol{S}_{1:t-1}, \boldsymbol{X}_{1:t-1}, \boldsymbol{V}, U \qquad t = 1, \ldots, T.$$

- Under LC-SIA the standard IPW estimator may be biased, but it is possible to correct it by computing the weights using probabilities conditioned on U:

$$Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i, U_i = c_i).$$

LC-IPW: a new estimator to account for unobserved confounding

We propose a two-step estimation procedure:

1. fit an auxiliary latent class model to assign subjects to latent classes
2. fit the MSM using weights computed with the latent-class-specific probabilities

We have written a `MATLAB` code, but estimation could be carried out by existing software (step 1: latent class (mixture) modelling; step 2: weighted logistic regression)

Step 1: auxiliary latent class model

- In order to assign subjects to latent classes, we fit a latent class model for the treatment indicators and the observed covariates
- The joint distribution of the observed variables is written as a finite mixture over the latent classes $(c = 1, \ldots, k)$ and each component of the mixture is recursively factorized
- $f(s_t \mid \mathbf{S}_{1:t-1}, \mathbf{X}_{1:t-1}, \mathbf{V}, c) \Rightarrow$ logistic regression model with specific parameters for every combination of occasion t and latent class c
- $f(\mathbf{V}|c)$ and $f(\mathbf{X}_t \mid \mathbf{S}_{1:t}, \mathbf{X}_{1:t-1}, \mathbf{V}, c) \Rightarrow$ modeled according to the nature of the variables (e.g. for continuous variables we can use a multivariate normal regression model)

- The parameters of the auxiliary latent class model are estimated with maximum likelihood using an EM algorithm; the number of support points k is chosen by a fit index, e.g. the Normalized Entropy Criterion (NEC) of Celeux and Soromenho (1996)

- Once the parameters have been estimated, every subject is assigned to the latent class with the highest posterior probability

Step 2: weighted regression

- The second step of the proposed LC-IPW method entails fitting the MSM with a modified IPW procedure where the weight of each subject is computed conditionally on the assigned latent class

- The (stabilized) weights are

$$sw_{i,\hat{c}_i} = \frac{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, U_i = \hat{c}_i)}{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i, U_i = \hat{c}_i)}$$

- The probabilities are estimated using logistic models after assigning the latent classes

- Standard errors and confidence intervals for the parameters of the MSM are obtained via non-parametric bootstrap

Simulation study: design

- Model (for $T = 4$ or $T = 8$ occasions)
  - ▷ continuous outcome Y
  - ▷ sequential binary treatment S_t
  - ▷ pre-treatment continuous covariate V (confounder if and only if $\phi_2 \neq 0$)
  - ▷ time-varying continuous covariate X_t (confounder if and only if $\phi_2 \neq 0$)
  - ▷ unobserved pre-treatment covariate U (confounder if and only if $\phi_1 \neq 0$)

$$
\begin{aligned}
logitPr(S_{it} = 1) &= \begin{cases} -1 + u_i \phi_1 (4/T) + v_i \phi_2 (4/T), & t = 1, \\ -1 + u_i \phi_1 (4/T) + x_{i,t-1} \phi_2 (4/T) - s_{i,t-1}, & t = 2, \dots, T, \end{cases} \\
X_{it} &= \begin{cases} -0.25 + u_i/2 + v_i + s_{it} + \varepsilon_{it}, & t = 1, \\ -0.25 + u_i/2 + x_{i,t-1} + s_{it} + \varepsilon_{it}, & t = 2, \dots, T-1, \end{cases} \\
Y_i &= u_i/2 + x_{i,T-1} + s_{iT} - 0.25 + \varepsilon_{iT},
\end{aligned}
$$

where ε_{it} are iid $N(0, 0.25)$ and V_i are iid $N(0, 1)$.

- Parameters for confounding: $\phi_1 \in \{-0.5, 0, 0.5\}$, $\phi_2 \in \{-0.5, 0, 0.5\}$

- Alternative distributions of the unobserved pre-treatment covariate U:
  - ▷ LC2: U_i discrete Uniform on -1,1
  - ▷ LC3-type1: U_i discrete Uniform on $-\sqrt{1.5}$,0,$\sqrt{1.5}$
  - ▷ LC3-type2: U_i discrete Uniform on -2,0,2
  - ▷ Normal: U_i standard Normal
  - ▷ Uniform: U_i continuous Uniform in the interval $[-\sqrt{3}, \sqrt{3}]$

  (distributions with mean 0 and variance 1, except LC3-type2 with variance $8/3$)

- Regardless of the distribution of U_i, the MSM for the outcome is

$$E(Y^{(\boldsymbol{s}_{1:T})}) = \beta_0 + s_+\beta_1, \quad \text{where } \beta_1 = 1 \text{ for any } T \in \{4, 8\}$$

- Number of scenarios: 36 ($2 \times 2 \times 3 \times 3$ values of n, T, ϕ_1, ϕ_2)

- Sample size $n = 1000$ or $n = 4000$ Number of simulated samples: 1000

- Estimation methods: (i) OLS (unweighted) regression, (ii) IPW regression, (iii) proposed LC-IPW with a number of latent classes k chosen by NEC

Median Bias and MAE for U discrete (LC2)

			Median Bias				MAE			
			$T=4$		$T=8$		$T=4$		$T=8$	
ϕ_1	ϕ_2	Method	1000	4000	1000	4000	1000	4000	1000	4000
-0.5	-0.5	IPW	-0.585	-0.536	-0.486	-0.364	0.590	0.538	0.520	0.396
		LC-IPW	-0.155	-0.095	-0.185	-0.079	0.232	0.148	0.290	0.176
-0.5	0.0	IPW	-0.546	-0.541	-0.411	-0.409	0.546	0.541	0.411	0.409
		LC-IPW	-0.015	-0.011	-0.005	0.000	0.044	0.023	0.038	0.018
-0.5	0.5	IPW	-0.525	-0.527	-0.491	-0.502	0.525	0.527	0.491	0.502
		LC-IPW	0.006	0.001	0.014	0.008	0.060	0.030	0.088	0.044
0.0	-0.5	IPW	-0.052	-0.023	-0.116	-0.040	0.122	0.074	0.202	0.110
		LC-IPW	-0.066	-0.028	-0.106	-0.037	0.127	0.071	0.199	0.109
0.0	0.0	IPW	0.005	0.001	0.005	-0.002	0.045	0.022	0.053	0.027
		LC-IPW	0.000	0.001	0.004	0.001	0.029	0.014	0.028	0.014
0.0	0.5	IPW	0.027	0.018	0.025	0.012	0.108	0.055	0.143	0.078
		LC-IPW	0.043	0.021	0.047	0.016	0.091	0.052	0.135	0.071
0.5	-0.5	IPW	0.455	0.454	0.316	0.271	0.455	0.454	0.316	0.271
		LC-IPW	-0.013	-0.003	-0.051	-0.022	0.075	0.039	0.127	0.065
0.5	0.0	IPW	0.489	0.484	0.405	0.404	0.489	0.484	0.405	0.404
		LC-IPW	0.024	0.012	0.002	0.003	0.053	0.026	0.046	0.020
0.5	0.5	IPW	0.531	0.495	0.576	0.509	0.533	0.496	0.584	0.511
		LC-IPW	0.161	0.107	0.169	0.091	0.209	0.140	0.244	0.140

Simulation study: main findings

- The LC-IPW estimator *outperforms* IPW essentially in all cases:
  - ▷ As sample size n increases $\Rightarrow$ IPW stable, LC-IPW improves
  - ▷ As number of occasions T increases $\Rightarrow$ no monotone pattern (worse or better depending on type of confounding ϕ_1, ϕ_2)
- In terms of MAE, LC-IPW is slightly better than IPW even when U is not a $confounder$ but a $pure\ predictor\ of\ outcome$ $(\phi_1 = 0)$, consistently with results on over-adjustment in inverse probability weighting by Rotnitzky, Li and Li (2010) and other simulations by Lefebvre, Delaney and Platt (2008)
- Results are confirmed for alternative distributions of U both discrete and continuous

Application to wage subsidies

- Dataset about $n = 1640$ *Finnish firms* (manufactures and services) between 20 and 200 employees that applied for wage subsidies in the period 1995-2002 ($T = 8$ occasions)

- The *aim of the policy* is to fill the gap between the wage that the firm is willing to pay and the unionized wage level

- Observations were extracted from the registers compiled by he *Finnish Tax Authority*

- Wage subsidies are the *most common type of subsidy* (required at least once by 65% of the firms in the sample)

- Available variables (measured at every year):
  - employment (number of employees)
  - wage (total and per employee)
  - fixed capital
  - sales
  - profit
- *Treatment variable* S_t: indicator taking the value 1 if the firm receives a wage subsidy in year t
- *Outcome* Y: employment at the end of the period
- *Potential confounders* $\boldsymbol{X}_t$: all the variables observed at end of year t (possibly including lagged values)

Descriptive statistics

- Sample distribution of the subsidies:

year	# firms	%
1995	582	35.49
1996	448	27.32
1997	491	29.94
1998	450	27.44
1999	383	23.35
2000	293	17.87
2001	242	14.76
2002	232	14.15

#subsidies	% firms	% cum.
0	34.94	34.94
1	18.54	53.48
2	15.18	68.66
3	10.24	78.90
4	7.44	86.34
5	6.16	92.50
6	4.09	96.59
7	1.71	98.29
8	1.71	100.00

- We considered several specifications for the MSM, here are the results of the following:

$$E(Y^{(\boldsymbol{s}_{1:8})}) = \beta_0 + s_+\beta_1$$

  where

  - $Y^{(\boldsymbol{s}_{1:8})}$ = number of employees at the end of the period
  - s_+ = number of years receiving subsidy (0,1,. . . ,8)
  - β_1 = causal effect (average change in employment for each year receiving a subsidy)

- To compute the weights for the standard estimator (IPW), the treatment indicators S_t are modeled by a *logistic regression* with a time dummy for each year and several covariates at $t-1$ and $t-2$ (i.e. we added lagged values)

- Covariates:
  - ▷ treatment indicator (wage subsidy)
  - ▷ log(employment)
  - ▷ log(wage per employee)
  - ▷ log(fixed capital)
  - ▷ log(sales)
  - ▷ $sign$(profit)$|$profit$|^{0.25}$

- To compute the weights for the proposed estimator (LC-IPW), the treatment indicators S_t are modeled by a *logistic regression* as before with the addition of *latent classes*
  - ▷ the latent class is assigned to each subject through an auxiliary model for the confounders (latent class *multivariate normal regression* with a common variance-covariance matrix)

- Results for the IPW estimator

Parameter	Estimate	95% Conf.	interval
β_0	67.958	63.306	72.365
β_1	3.932	2.207	6.052

(confidence intervals based on non-parametric bootstrap)

- Results for the LC-IPW estimator (number of classes $k = 4$ chosen by the NEC criterion):

Parameter	Estimate	95% Conf.	interval
β_0	70.280	65.032	75.385
β_1	2.156	0.257	4.499

Final remarks

- Compared to standard IPW, the proposed LC-IPW method has
  - ▷ *higher complexity*: it requires to formulate a latent class auxiliary model which also involves the distribution of the confounders
  - ▷ *better performance*: it properly corrects for unobserved confounding and it may be efficient even in case of no unobserved confounding
- Further developments:
  - ▷ Sensitivity of the parameter estimates on the *specification and estimation of the auxiliary model* (e.g. how to choose the number of classes)
  - ▷ Using the LC approach with other methods, e.g. *longitudinal propensity score* (Achy-Brou, Frangakis and Griswold 2010)
  - ▷ Accounting for *time-varying unobserved confounders* using a latent Markov model

References

Celeux, G. and Soromenho, G. (1996) An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13: 195-212.

Cole S. R. and Hernan M. A. (2008) Inverse probability weights for marginal structural models. *American Journal of Epidemiology*, 168: 656-664.

Lechner M. and Miquel R. (2010) Identification of the effects of dynamic treatments by sequential conditional independence assumptions. *Empirical Economics*, 39: 111-137.

Lefebvre G., Delaney J. A. C. and Platt R. W. (2008) Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Statistics in Medicine*, 27: 3629-3642.

Robins J. M., Hernan M. A., and Brumback B. (2000) Marginal structural models and causal inference, *Epidemiology*, 11: 550-560.

Rotnitzky A., Li L. and Li X. (2010) A note on overadjustment in inverse probability weighted estimation. *Biometrika*, 97: 997-1001.

van der Wal W. and Geskus R. (2011) ipw: An r package for inverse probability weighting. *Journal of Statistical Software*, 43: 1-23.

Thank you!