



Munich Personal RePEc Archive

**Evaluating time-varying treatment effects  
in latent Markov models: An application  
to the effect of remittances on poverty  
dynamics**

Tullio, Federico and Bartolucci, Francesco

University of Rome Tor Vergata, University of Perugia

14 January 2019

Online at <https://mpra.ub.uni-muenchen.de/91459/>

MPRA Paper No. 91459, posted 17 Jan 2019 07:43 UTC

# Evaluating time-varying treatment effects in latent Markov models: An application to the effect of remittances on poverty dynamics

Federico Tullio <sup>1</sup>

Francesco Bartolucci <sup>2</sup>

## Abstract

To assess the effectiveness of remittances on the poverty level of recipient households, we propose a causal inference approach that may be applied with longitudinal data and time-varying treatments. The method relies on the integration of a propensity score based technique, the inverse propensity weighting, with a general Latent Markov (LM) framework. It is particularly useful when the interest is in an individual characteristic that is not directly observable and the analysis is focused on: (i) clustering individuals in a finite number of classes according to this latent characteristic and (ii) modelling its evolution across time depending on the received treatment. Parameter estimation is based on a two-step procedure in which individual weights are computed for each time period based on predetermined covariates and a weighted version of the standard LM model likelihood based on such weights is maximised by means of an expectation-maximisation algorithm. Finite-sample properties of the estimator are studied by simulation. The application is focused on the effect of remittances on the poverty status of Ugandan households, based on a longitudinal survey spanning the period 2009-2014 and where response variables are indicators of deprivation.

Keywords: Causal inference; Expectation-maximisation algorithm; Potential outcomes; Weighted Maximum Likelihood

JEL: I32, C33

## 1 Introduction

In applications involving longitudinal data, Latent Markov (LM) models (Bartolucci et al., 2013), also known as Hidden Markov Models (HMMs) for longitudinal data (Zucchini and MacDonald, 2009), are widely used to analyse the evolution across time of an

---

<sup>1</sup>University of Rome Tor Vergata, email: federico.tullio@alumni.uniroma2.eu

<sup>2</sup>University of Perugia, email: francesco.bartolucci@unipg.it

individual characteristic that is not directly observable and then is latent. These models are particularly useful when the latent characteristic is measured on the basis of a series of indicators observed at each time occasion, which typically correspond to categorical response variables. The main output is the clustering of the units of observation in a specified number of groups, named latent classes or latent states, at each time period, permitting the analysis of the transition between these clusters. The aggregation of the sample units in different groups is common to the Latent Class (LC) analysis (Lazarsfeld and Henry, 1969), with the main difference that in an LM approach these units are allowed to move between the classes across time. The typical assumption of LC and LM models is that the underlying heterogeneity of subjects is explained by the latent grouping variable, and every individual is clustered depending on the probability of belonging to each group given the observed data.

In its basic formulation, for all the observational units the LM model assumes the same (time-homogenous) measurement model, corresponding to the conditional distribution of the response variables given the latent states, and the same latent model, corresponding to the distribution of the latent variables (Bartolucci et al., 2013, Chapter 3). A straightforward extension of the basic LM model is for the inclusion of individual covariates that may be time varying or time fixed and be included either in the measurement or in the latent model. The focus is on the latter specification whenever the main interest is in modelling the effect of the covariates on the latent variable distribution (Bartolucci et al., 2014). In this case, the covariates are adopted as predictors of the latent states. Other possible extensions are the multilevel LM model for clustered sample units (Bartolucci et al., 2011) and LM models with the inclusion of random effects, also known as Mixed HMMs (Altman, 2007; Maruotti, 2011).

Recent works integrate propensity score methods (Rosenbaum and Rubin, 1983) with static and dynamic latent variable analysis to make causal inference on the effect of the covariates on the latent variable distribution. Lanza et al. (2013) first proposed a propensity score based approach to estimate such a causal effect in an LC framework. The approach resorts to matching (Rosenbaum and Rubin, 1983; Heckman et al., 1997) and Inverse-Propensity Weighting (IPW; Robins et al., 1995, 2000) on the basis of the scores obtained conditioning the probability of treatment or exposure to predetermined covariates. Their research question was whether college enrolment has a long-term effect on adult substance use. To address this question, they used the Average Treatment Effect (ATE), corresponding to the expected “gain” in terms of dependent variable for a randomly selected unit from the population, which is measured on the logit scale in terms of relative probability of belonging to a certain latent class with respect to a baseline category<sup>1</sup>. Bartolucci et al. (2016) incorporated the IPW setting described in Lanza et al. (2013) in a longitudinal framework and with multiple treatments. They proposed an LM model formulated with a potential outcome notation (Rubin, 1974), and thus obtained estimates of the ATEs in terms of initial probabilities as well as transition probabilities, among latent classes again on the logit scale. The multiple treatments correspond to different college degrees and are compared each other in terms of human capital as proxied by different outcomes in the labor market. In both Lanza et al. (2013) and Bartolucci et al. (2016) the distribution of the latent variable is thus conditioned on a time-fixed treatment, such as past college enrolment.

The strategies described above are not appropriate to deal with treatments that can

---

<sup>1</sup>The authors extended the analysis also to the estimation of the average treatment effect on the treated, that is the average gain from treatment for those who actually were treated.

be received at multiple timepoints, as it is often the case in longitudinal settings. Motivated by an application about the casual effect of remittances on the poverty dynamics in Uganda, we extend the model in Bartolucci et al. (2016) allowing the treatment to vary over time. The framework is the standard LM model with covariates in the latent distribution, where the self-selection into the treatment level is a possible source of bias in the parameter estimation. Selection into treatment is therefore occurring at every time period and its potential bias is addressed by the prediction of the individual time-specific probabilities of being treated conditional on a set of exogenous covariates. The estimation strategy is based on a two-step procedure. The first step consists in computing the individual probabilities of sorting into the treatment, namely the propensity scores. The second step consists in maximising a weighted version of the LM model with covariates in the latent component, where the weights depend on the scores obtained as step 1. As usual, the maximisation of the likelihood at the second step is based on the Expectation-Maximisation (EM) algorithm (Dempster et al., 1977). The standard errors for the parameter estimates are computed via a non-parametric bootstrap (Efron and Tibshirani, 1994). The finite-sample properties of the proposed estimator are assessed by Monte Carlo simulations.

The remainder of this article is organised as follows. In the next section we describe the motivating application of the proposed methodology. In Section 3 we outline the traditional LM framework and the causal LM model proposed by Bartolucci et al. (2016). In Section 4 we illustrate our proposal, linking the adopted assumptions to the previous literature and describing the estimation procedure. Section 5 illustrates the simulation study of the finite-sample properties of the estimator, while in Section 6 we present the empirical application based on the assessment of the effect of remittances on households' deprivation status in Uganda. Some concluding remarks are provided in the last section.

## 2 Motivation: effect of remittances on material deprivation

The determinants of migration decision have been studied in depth in the economic literature. In particular, according to the New Economics of Labor Migration (NELM; Stark and Bloom, 1985), migration is a risk-sharing behaviour of families or households rather than an individual choice. This is because, while remittances are ignored in neo-classical migration theory<sup>2</sup>, as maintained by NELM they are perceived as one of the most essential motives for migrating. Remittances are in this sense the fundamental mean through which migration turns out to be a social protection scheme. Devereux and Sabates-Wheeler (2004, p. 9) defined social protection as “(...) the set of all initiatives, both formal and informal, that provide: social assistance to extremely poor individuals and households; social services to groups who need special care or would otherwise be denied access to basic services; social insurance to protect people against the risks and consequences of livelihood shocks; and social equity to protect people against social risks such as discrimination or abuse”. The effectiveness of remittances is related to two crucial aspects of social protection: whether they operate as a social insurance strategy or as a social assistance strategy.

In general terms, there are several reasons to remit (see, e.g., Lucas and Stark, 1985; Rapoport and Docquier, 2006) and the studies on the relation between remittances and

---

<sup>2</sup>See De Haas (2010) for a broad review of migration theories.

social protection are numerous. Several studies using households survey data found that remittances reduce severity of poverty in many countries (see Ratha et al., 2011, and Adams, 2011, for a review), although the magnitude of the impact turns out to be limited. Dealing with the self-selectivity into treatment in evaluating the causal impact of remittances is a further matter of complication (McKenzie and Yang, 2012). It is worth noting that these studies on remittances are not conclusive on the overall impact of migration on the households left behind because of the lost labor earnings of the migrants, in particular for the most vulnerable households in the short term. If one is interested in the broad poverty-reducing impact of remittances, attention must be paid also to who is receiving remittances, whether rich or poor households (Démurger, 2015; McKenzie and Sasin, 2007).

Since the early works of Amartya Sen and Peter Townsend<sup>3</sup>, material deprivation has captured the interest of researchers and policymakers. There is today a widespread agreement that deprivation has a multidimensional nature (Atkinson, 2003), although this is traditionally ignored by headline money metric measures of poverty and analysis of the impact of remittances, which focus on one dimension at a time. Poverty may be defined as the lack of a combination of items depicting material living conditions<sup>4</sup>, thus material deprivation can be used to monitor progresses in the fight against poverty and social exclusion. In Sen's words, a way to identify the poor for a given set of *basic needs* is "(...) simply to check the set of people whose actual consumption baskets happen to leave some basic need unsatisfied. This we may call the *direct method* (...)" (Sen, 1981, p. 26). In contrast, the *indirect method* involves the use of the income notion and consists in setting a poverty line and identifying those whose actual incomes fall below the line.

Multidimensional poverty analysis is increasingly adopting latent variable models (Moisio, 2004; Whelan and Maitre, 2006; Krishnakumar, 2008; Machado et al., 2009; Dotto et al., 2018), grounding on the fact that observed heterogeneity in the population is due to unobservable components. For this analysis, an LM model is particularly suited in presence of longitudinal data (Dotto et al., 2018), as it implements the clustering of households in two subpopulations according to their current poverty status at each time period, depending on a set of household-level indicators of deprivation. Poverty is conceived as an unobservable status of the household, and is represented by a discrete latent variable, while the manifest (observable) variables are the deprivation item responses. Such a model also allows us to follow households' transition among classes over time, permitting to distinguish between persistently poor and currently poor households, depending on the number of subsequent time periods in which they are clustered as poor. The dynamics of the latent grouping variable can be modelled conditional on some variables of interest, as the reception of remittances. Such a strategy allows us to account for a heterogeneous treatment effect, as recipient households may be clustered either in the poor and in the nonpoor latent groups, and thus to test whether the treatment acts as a social insurance mechanism (i.e., reducing the probability of falling into poverty) and/or as a measure to alleviate poverty (i.e., increasing the probability of escaping from poverty).

---

<sup>3</sup>See, for example, Sen (1976, 1980, 1981) and Townsend (1979, 1987).

<sup>4</sup>E.g., Alkire and Foster (2011a,b) and UNDP (2016).

### 3 Available methodology

In the following we first illustrate the multivariate version of the basic LM model and its extension for the inclusion of individual covariates in the latent component. We also present the causal framework for multilevel treatments proposed by Bartolucci et al. (2016).

#### 3.1 Basic LM model

Let  $T$  denote the number of time occasions and  $n$  the sample size, and suppose that a realisation  $\mathbf{y}_{it}$  of the vector of  $r$  categorical response variables  $\mathbf{Y}_{it} = (Y_{i1t}, \dots, Y_{irt})'$  is observed for each individual  $i = 1, \dots, n$  and time occasion  $t = 1, \dots, T$ . Each variable  $Y_{ijt}$  in  $\mathbf{Y}_{it}$  has  $c_j$  categories labeled from 0 to  $c_j - 1$ , where  $j = 1, \dots, r$ . As mentioned in Section 1, the main interest of the model is in the distribution of the latent variables  $U_{it}$ . In particular, for every unit  $i$ , we assume that  $\mathbf{U}_i = (U_{i1}, \dots, U_{iT})'$  follows a first-order Markov chain with state space  $\{1, \dots, k\}$  and initial and transition probabilities common to the other units. Thus  $U_{it}$  is conditionally independent of  $U_{is}$  given  $U_{i,t-1}$  for  $s = 1, \dots, t - 2$  with  $t > 2$ , whereas for any  $i$  the initial probabilities are defined as

$$\lambda_u = P(U_{i1} = u), \quad u = 1, \dots, k,$$

and the transition probabilities, which are not constrained to be time homogenous, are defined as

$$\pi_{t,v|u} = P(U_{it} = v | U_{i,t-1} = u), \quad t = 1, \dots, T, \quad u, v = 1, \dots, k.$$

On the basis of the above parameters, the distribution of  $\mathbf{U}_i$  has probability mass function

$$p(\mathbf{u}_i) = P(\mathbf{U}_i = \mathbf{u}_i) = \lambda_{u_{i1}} \prod_{t=1}^T \pi_{t,u_{it}|u_{i,t-1}},$$

where  $\mathbf{u}_i = (u_{i1}, \dots, u_{iT})'$  denotes a realisation of  $\mathbf{U}_i$ .

A crucial assumption is that of *local independence*, according to which the response variables referred to the same individual are conditionally independence given the underlying latent variables. In practice, this means that if we knew the latent state  $U_{it}$  of a subject  $i$  at occasion  $t$ , the realisation of any response variable apart from  $Y_{ijt}$  would not help in predicting  $Y_{ijt}$ . This is because the latent variables represent the only factors affecting the response variables and implies that

$$p(\mathbf{y}_{it}|u_{it}) = P(\mathbf{Y}_{it} = \mathbf{y}_{it} | U_{it} = u_{it}) = \prod_{j=1}^r \phi_{jy_{ijt}|u_{it}},$$

where  $\mathbf{y}_{it} = (y_{i1t}, \dots, y_{irt})'$  and

$$\phi_{jy|u} = P(Y_{ijt} = y | U_{it} = u), \quad j = 1, \dots, r, \quad u = 1, \dots, k, \quad y = 0, \dots, c_j - 1,$$

are the conditional response probabilities, assumed to be individual and time constant, so to have a clear identification of the latent states. Denoting by  $\tilde{\mathbf{Y}}_i$  the set of all

response variables  $Y_{ijt}$  for the same individual  $i$ , which is composed by the vectors  $\mathbf{Y}_{it}$  for  $t = 1, \dots, T$ , we have that

$$p(\tilde{\mathbf{y}}_i | \mathbf{u}_i) = P\left(\tilde{\mathbf{Y}}_i = \tilde{\mathbf{y}}_i | \mathbf{U}_i = \mathbf{u}_i\right) = \prod_{j=1}^r \prod_{t=1}^T \phi_{jy_{ijt}|u_{it}},$$

where the realisation  $\tilde{\mathbf{y}}_i$  is made by the union of vectors  $\mathbf{y}_{it}$ . The previous expression is again based on the conditional independence assumption.

Finally, the *manifest distribution* of  $\tilde{\mathbf{Y}}_i$ , corresponding to the marginal distribution of all responses provided by the same individual  $i$  once the latent variables are integrated out, is given by

$$\begin{aligned} p(\tilde{\mathbf{y}}_i) &= P(\tilde{\mathbf{Y}}_i = \tilde{\mathbf{y}}_i) = \sum_{\mathbf{u}} p(\tilde{\mathbf{y}}_i | \mathbf{u}) p(\mathbf{u}) \\ &= \sum_{\mathbf{u}} \left( \lambda_{u_1} \prod_{t>1} \pi_{t,u_t|u_{t-1}} \right) \prod_{j=1}^r \prod_{t=1}^T \phi_{jy_{ijt}|u_{it}}, \end{aligned} \quad (1)$$

where the sum  $\sum_{\mathbf{u}}$  is extended to all the possible configurations of the latent process  $\mathbf{u} = (u_1, \dots, u_T)'$ . In order to avoid computational problems, this marginal distribution is computed in practice by using the Baum-Welch forward recursion (Baum et al., 1970; Welch, 2003).

Overall, the number of free parameters to be estimated is equal to  $k \sum_j (c_j - 1)$  in the measurement model ( $\phi_{jy|u}$ ),  $k - 1$  in the initial distribution ( $\lambda_u$ ), and  $(T - 1)k(k - 1)$  in the transition distribution ( $\pi_{t,v|u}$ ). Parameter estimation is carried out via Maximum Likelihood (ML). Using the above formulation and given the observed data corresponding to the vector  $\tilde{\mathbf{y}}_i$  for each  $i$ , the log-likelihood function has expression

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log p(\tilde{\mathbf{y}}_i),$$

where we assume that the observed sample consists on  $n$  independent units observed for  $T$  time periods and  $\boldsymbol{\theta}$  is the vector of all model parameters. Function  $\ell(\boldsymbol{\theta})$  can be maximised by means of the EM algorithm (Dempster et al., 1977) implemented on the basis of the Baum-Welch recursions; see Bartolucci et al. (2013, Chapter 3.5) for details. In general, the EM algorithm is a method for ML estimation in presence of missing data and it can be used to estimate the parameters of an LM model by assuming the latent states underlying the observations as missing. As usual the algorithm alternates two steps (E-step and M-step) until convergence.<sup>5</sup>

## 3.2 Covariates in the latent model

Suppose that for each individual we also observe the column vectors of  $q$  covariates  $\mathbf{x}_{it}$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ , and define  $\tilde{\mathbf{x}}_i$  as the vector of all such covariates stacked along the time dimension for the same individual  $i$ . These vectors are realisations of the random vectors  $\mathbf{X}_{it}$  and  $\tilde{\mathbf{X}}_i$ , respectively. As already mentioned in Section 1, individual covariates may be included either in the measurement model (corresponding to the conditional

---

<sup>5</sup>A description of the EM algorithm is provided in Section 4.2 in connection with the proposed approach.

distribution of the responses given the latent state) or in the latent model (corresponding to the distribution of the unobserved Markov chain). In the former case, the latent structure accounts for unobserved heterogeneity, while in the latter the main interest is in modelling the effect of the covariates on the latent variables. We are interested in the second specification, as we assume that observable outcomes indirectly measure a latent construct, in our application the poverty status of a household, which may evolve over time. In such a case, direct effects of the covariates on the responses are usually ruled out (Bartolucci et al., 2013).

Under the assumed formulation, the *initial* and *transition* probabilities are defined conditional on the realisation of the random vector of covariates as

$$\begin{aligned}\lambda_{i,u} &= P(U_{i1} = u | \mathbf{X}_{i1} = \mathbf{x}_{i1}), \quad u = 1, \dots, k, \\ \pi_{it,v|u} &= P(U_{it} = v | U_{i,t-1} = u, \mathbf{X}_{it} = \mathbf{x}_{it}), \quad u, v = 1, \dots, k,\end{aligned}$$

for  $i = 1, \dots, n$ . A typical way to model the dependence of the latent variable distribution on the individual covariates is to resort to different types of logit<sup>6</sup>. We focus on a simple multinomial logit parametrisation where the reference category is the first category in the initial distribution ( $t = 1$ ) and category  $u$  in the transition distribution ( $t > 1$ ):

$$\begin{aligned}\log \frac{\lambda_{i,u}}{\lambda_{i,1}} &= \mathbf{x}'_{i1} \boldsymbol{\beta}_u, \quad u = 2, \dots, k, \\ \log \frac{\pi_{it,v|u}}{\pi_{it,u|u}} &= \mathbf{x}'_{it} \boldsymbol{\gamma}_{v|u}, \quad u, v = 1, \dots, k, \quad v \neq u,\end{aligned}$$

where  $\boldsymbol{\beta}_u$  and  $\boldsymbol{\gamma}_{v|u}$  are two column vectors of dimension  $q+1$  including the constant term.

Overall, the number of free parameters to be estimated is equal to  $k \sum_r (c_r - 1)$  in the measurement model ( $\phi_{jy|u}$ ),  $(q+1)(k-1)$  in the initial distribution ( $\boldsymbol{\beta}_u$ ), and  $k(k-1)(q+1)$  in the transition distribution ( $\boldsymbol{\gamma}_{v|u}$ ). To estimate these parameters on the basis of the observed data,  $(\tilde{\mathbf{x}}_i, \tilde{\mathbf{y}}_i)$ ,  $i = 1, \dots, n$ , we maximise the log-likelihood function given by

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log p(\tilde{\mathbf{y}}_i | \tilde{\mathbf{x}}_i), \quad (2)$$

where the *manifest* distribution is defined as in equation (1), taking into account the dependence of the initial and transition probabilities on the covariates. The EM algorithm can again be used to maximise equation  $\ell(\boldsymbol{\theta})$ , with the inclusion of iterative algorithms to update the regression parameters involved in the initial and transition probabilities at the M-step (Bartolucci et al., 2014; Bartolucci et al., 2013, Chapter 5.6).

### 3.3 Causal LM model for time-fixed multiple treatments

Now suppose that a multiple treatment exists which is administered before the first longitudinal response and let  $Z_i$  be the corresponding indicator variable. Also suppose that the distribution of this variable depends on an  $s$ -dimensional vector of pretreatment covariates  $\mathbf{X}_i$ . Bartolucci et al. (2016) reformulated the LM model with covariates in the latent component from a causal perspective, in the spirit of Rubin (1974) potential outcomes

---

<sup>6</sup>Usual parametrisations are multinomial logits if latent states are assumed to be not ordered or cumulative/global logits if classes are assumed to be ordered (Bartolucci et al., 2014). For other possible parametrisations, see Agresti (2002) among others.

notation used to define causal models. In the following we outline the basic assumptions of this model using a notation that is more suitable for the following developments.

The approach of Bartolucci et al. (2016) is based on potential versions of the latent variables underlying the individual responses that we refer to as *potential latent variables* and are denoted by  $U_{it}^{(z)}$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ , and  $z = 0, \dots, d - 1$ , where  $d$  is the number of possible treatments starting from 0 that typically corresponds to the control. In practice,  $U_{it}^{(z)}$  correspond the latent state, taken in the set  $\{1, \dots, k\}$  if the treatment taken by individual  $i$  would be  $z$ . For the same individual, the vector collecting the sequence of latent variables is  $\mathbf{U}_i^{(z)} = (U_{i1}^{(z)}, \dots, U_{iT}^{(z)})$ . In such a context, only one of the possible outcomes  $\mathbf{U}_i^{(z)}$  is selected, the one corresponding to the taken treatment. Accordingly, we denote the initial and transition probabilities as follows:

$$\begin{aligned}\lambda_{i,u}^{(z)} &= P(U_{i1}^{(z)} = u), \\ \pi_{i,v|u}^{(z)} &= P(U_{it}^{(z)} = v | U_{i,t-1}^{(z)} = u),\end{aligned}$$

for  $i = 1, \dots, n$ ,  $t = 2, \dots, T$ ,  $u, v = 1, \dots, k$ ,  $u \neq v$ , and  $z = 0, \dots, d - 1$ . The initial probabilities are modelled by the multinomial logit parametrisation

$$\log \frac{\lambda_{i,u}^{(z)}}{\lambda_{i,1}^{(z)}} = \alpha_u + \beta_u^{(z)}, \quad u = 2, \dots, k,$$

where  $\alpha_u$  is the intercept corresponding to the effect of the control and  $\beta_u^{(z)}$ ,  $z = 1, \dots, d - 1$ , with  $\beta_u^{(0)} \equiv 0$ , is a measure of the causal differential effect of treatment  $z$  with respect to the control. Similarly, the transition probabilities are parametrised as follows:

$$\log \frac{\pi_{i,v|u}^{(z)}}{\pi_{i,u|u}^{(z)}} = \gamma_{v|u} + \delta_{v|u}^{(z)}, \quad u, v = 1, \dots, k, \quad u \neq v,$$

where  $\gamma_{v|u}$  is the intercept and  $\delta_{v|u}^{(z)}$ ,  $z = 1, \dots, d - 1$ , with  $\delta_{v|u}^{(0)} \equiv 0$ , is again a measure of causal differential effect of treatment  $z$  with respect to the control.

As is already clear, parameters  $\beta_u^{(z)}$  and  $\gamma_{v|u}^{(z)}$ ,  $z = 1, \dots, d - 1$ , correspond to ATEs measured on the logit scale and referred to the latent state  $u$ , when we consider the first time occasion, and the transition from latent state  $u$  to latent state  $v$ , for the following occasions. Alternatively, we can directly express the ATEs on the probability rather than on the logit scale by direct comparisons of the type  $\lambda_{i,u}^{(z)} - \lambda_{i,u}^{(0)}$  or  $\pi_{i,v|u}^{(z)} - \pi_{i,v|u}^{(0)}$ , again for  $z = 1, \dots, d - 1$ . More details about this second possible formulation are provided in the following section. In any case, further assumptions are required in order to identify these causal effects. In fact, individuals with different levels of the dependent variable may be self-selecting into the different levels of the treatment group, and no counterfactual is observed at the individual level. In our setting, being the variable of interest (i.e., the latent variable) not directly observable is a matter of further complication in the estimation procedure.

Bartolucci et al. (2016) proposed to adopt a propensity score method, the IPW, to estimate the causal effects. The propensity score is the conditional probability of assignment to a particular treatment given a vector of observed covariates. According to Rosenbaum and Rubin (1983), a pseudo random assignment environment can be replicated when exposures are not randomised by adjusting the pre-treatment covariates for

the propensity scores, thus removing all the bias associated with differences in the pre-treatment variables. *Strong ignorability* is the key assumption (Rosenbaum and Rubin, 1983) and it consists of the ignorability (unconfoundedness) assumption and the overlap assumption.

Regarding the treatment selection mechanism, Bartolucci et al. (2016) formulated the following assumptions, which are rather common in the causal inference literature when the aim is removing confounding due to observable variables:

**A.1**  $U_{it} = U_{it}^{(z_i)}$ , where  $z_i$  is the observed treatment level of individual  $i$ ;

**A.2**  $0 < P(Z_i = z | \mathbf{X}_i) < 1$ ,  $z = 0, \dots, d - 1$  (*Overlap*);

**A.3**  $Z_i \perp\!\!\!\perp \mathbf{U}_i^{(z)} | \mathbf{X}_i$ ,  $z = 0, \dots, d - 1$  (*Ignorability*).

In words, each individual has a positive probability of taking each treatment, and the treatment is independent of the potential latent variables conditional on the set of pre-treatment covariates.

The estimation approach proposed by Bartolucci et al. (2016) is based on a two-step procedure. First, a model for the probability of taking a certain type of treatment is estimated based on a multinomial model (McCaffrey et al., 2013). Next, a weighted version of the LM model likelihood described above is estimated. As this procedure is closely related to our extension, we propose a detailed discussion in Section 4.2.

## 4 Proposed extension

In this section we reformulate the causal LM model described in Section 3.3 for the case of a time-varying treatment variable. First, we state the model assumptions and we link them to the previous sections. Then, we propose an estimation strategy based on a two-step procedure.

### 4.1 The model

Let  $Z_{it}$  be the treatment variable with levels  $0, \dots, d - 1$  and let  $\mathbf{X}_{it}$  be an  $s$ -dimensional vector of covariates affecting  $Z_{it}$  for individual  $i$  at time  $t$ . Let  $\mathbf{Z}_i$  and  $\tilde{\mathbf{X}}_i$  be the vector of treatments and covariates stacked along the time dimension. The potential versions of the latent variable are indicated by  $U_{it}^{(z_t)}$ ,  $i = 1, \dots, n$ ,  $t = 1, \dots, T$ , and  $z_t = 0, \dots, d - 1$ . Note that when the treatment varies at each time period  $t$ , there will be  $d^T$  possible treatment configurations  $\mathbf{z}$  for the whole observation period<sup>7</sup> and the vector of latent variables is defined as  $\mathbf{U}_i^{(\mathbf{z})} = (U_{i1}^{(z_1)}, \dots, U_{iT}^{(z_T)})$ .

In terms of measurement model, we rely on the *local independence* assumption discussed in Section 3.1, namely the conditional independence of the response variables given the latent states. Regarding the initial and transition probabilities, we assume a baseline-category logit model that does not require any ordering of the latent states. Let

$$\begin{aligned} \lambda_{i,u}^{(z_1)} &= P(U_{i1}^{(z_1)} = u), \\ \pi_{i,v|u}^{(z_t|z_{t-1})} &= P(U_{it}^{(z_t)} = v | U_{i,t-1}^{(z_{t-1})} = u), \end{aligned}$$

---

<sup>7</sup>On the other hand, in Bartolucci et al. (2016) the number of possible configurations is equal to the number of the treatment levels.

for  $i = 1, \dots, n$ ,  $t = 2, \dots, T$ ,  $u, v = 1, \dots, k$ ,  $u \neq v$ , and  $z_t = 0, \dots, d - 1$ . We also assume that the transition at time  $t$  only depends on the current level of the treatment ( $\pi_{i,v|u}^{(z_t|z_{t-1})} = \pi_{i,v|u}^{(z_t)}$ ); we discuss this assumption in Section 6.1 for our application. The following specifications for the initial and transition probabilities are assumed:

$$\log \frac{\lambda_{i,u}^{(z_1)}}{\lambda_{i,1}^{(z_1)}} = \alpha_u + \beta_u^{(z_1)}, \quad u = 2, \dots, k \quad (3)$$

$$\log \frac{\pi_{i,v|u}^{(z_t)}}{\pi_{i,u|u}^{(z_t)}} = \gamma_{v|u} + \delta_{v|u}^{(z_t)}, \quad u, v = 1, \dots, k, u \neq v, \quad (4)$$

where  $\alpha_u$  and  $\gamma_{v|u}$  are the intercepts corresponding to the effect of the control and  $\beta_u^{(z_1)}$  and  $\delta_{v|u}^{(z_t)}$  are the differential effect parameters of treatment  $z$ , for  $z = 1, \dots, d - 1$ .

The previous parameters are already ATEs measured on the logit scale. As already mentioned, these effects may be directly measured on the probability scale. The ATEs in terms of initial and transition probabilities will be given respectively by  $\lambda_{i,u}^{(z)} - \lambda_{i,u}^{(0)}$  for  $u = 2, \dots, k$ , and  $\pi_{i,v|u}^{(z)} - \pi_{i,v|u}^{(0)}$  for  $u, v = 1, \dots, k$ ,  $z = 1, \dots, d - 1$ . Given the parametrisation in (3) and (4), in terms of the regression coefficients, the ATEs in terms of initial probabilities may be computed as

$$\begin{aligned} ATE_u(z) &= \frac{1}{1 + \sum_{h=2}^k \exp(\alpha_h + \beta_h^{(z)})} - \frac{1}{1 + \sum_{h=2}^k \exp(\alpha_h)}, \quad u = 1, \\ ATE_u(z) &= \frac{\exp(\alpha_u + \beta_u^{(z)})}{1 + \sum_{h=2}^k \exp(\alpha_h + \beta_h^{(z)})} - \frac{\exp(\alpha_u)}{1 + \sum_{h=2}^k \exp(\alpha_h)}, \quad u = 2, \dots, k, \end{aligned}$$

and those in terms of transition probabilities as

$$\begin{aligned} ATE_{u|u}(z) &= \frac{1}{1 + \sum_{\substack{h=1 \\ h \neq u}}^k \exp(\gamma_{h|u} + \delta_{h|u}^{(z)})} - \frac{1}{1 + \sum_{\substack{h=1 \\ h \neq u}}^k \exp(\gamma_{h|u})}, \quad u = 1, \dots, k, \\ ATE_{v|u}(z) &= \frac{\exp(\gamma_{v|u} + \delta_{v|u}^{(z)})}{1 + \sum_{\substack{h=1 \\ h \neq u}}^k \exp(\gamma_{h|u} + \delta_{h|u}^{(z)})} - \frac{\exp(\gamma_{v|u})}{1 + \sum_{\substack{h=1 \\ h \neq u}}^k \exp(\gamma_{h|u})}, \quad u, v = 1, \dots, k, v \neq u, \end{aligned}$$

for  $z = 1, \dots, d - 1$ .

The assumptions regarding the treatment selection mechanism retrace those in Bartolucci et al. (2016). However, modifications of assumptions A.1-A.3 are required in order to account for the time-variation of the treatment. We therefore formulate the following assumptions:

**B.1**  $U_i = \mathbf{U}_i^{(\mathbf{z}_i)}$ , where  $\mathbf{z}_i = (z_{i1}, \dots, z_{iT})'$  is the observed treatment configuration of individual  $i$ ;

**B.2**  $0 < P(Z_{it} = z_t | \mathbf{X}_{it}) < 1$ ,  $z_t = 0, \dots, d - 1$  (*Overlap*);

**B.3**  $Z_{it} \perp\!\!\!\perp U_{it}^{(z_t)} | \mathbf{X}_{it}$ ,  $z_t = 0, \dots, d - 1$  (*Ignorability*).

These assumptions refer to time-specific treatment and vector of covariates and hold for every individual and time occasion. Note that the the *Ignorability* assumption here implies that the vector of treatments  $\mathbf{Z}_i$  is independent of all the possible configurations of latent variables  $\mathbf{U}_i^{(z)}$  conditional on the set of stacked covariates  $\tilde{\mathbf{X}}_i$ .

## 4.2 Weighted Maximum Likelihood Estimation

The proposed estimation strategy is based on a two-step procedure. The first step consists in estimating a multinomial logit model for the probability of being treated, given the individual vector of covariates. At the second step, a weighted version of the log-likelihood of the proposed LM model is maximised.

In more detail, propensity score estimates are obtained by fitting a multinomial logistic regression of the treatment variable  $Z_{it}$  on the set of  $s$  covariates  $\mathbf{X}_{it}$ ; see McCaffrey et al. (2013). This model assumes that

$$\log \frac{P(Z_{it} = z | \mathbf{X}_{it} = \mathbf{x}_{it})}{P(Z_{it} = 0 | \mathbf{X}_{it} = \mathbf{x}_{it})} = \mathbf{x}'_{it} \boldsymbol{\psi}_t^{(z)}, \quad z = 1, \dots, d-1,$$

where  $\boldsymbol{\psi}_t^{(z)}$  is a time-specific vector of dimension  $s+1$  including the intercept. The propensity score estimate for individual  $i$  is thus given by  $\hat{p}_{it} = \hat{P}(Z_{it} = z_{it} | \mathbf{X}_{it} = \mathbf{x}_{it})$ ,  $z_{it} = 0, \dots, d-1$ .

Among the different propensity score based methods to account for selection issues, the IPW is particularly suited in the present framework. The rationale behind IPW is that there is likely to be an under-representation of individuals with a low probability of being treated and are actually treated, and an over-representation of treated individuals with high propensity scores. In order to estimate the ATE, an up (down)-weighting of under (over)-represented individuals is needed. On the basis of the parameter estimates we can compute the individual weights:

$$\hat{w}_{it}^* = \frac{1}{\hat{p}_{it}}, \quad i = 1, \dots, n, t = 1, \dots, T.$$

Note that the above computation of individual weights is time-specific, as it relies on the distinct cross-sectional dataset for each time period. Allowing the treatment to vary over time requires the computation of the following weights:

$$\hat{w}_i^* = \prod_{t=1}^T \hat{w}_{it}^*, \quad (5)$$

which is the product of the cross-sectional weights and it corresponds to the inverse conditional probability that a subject had his/her own observed treatment history across time (Robins et al., 2000).

As Robins et al. (2000) pointed out, the weights may be inaccurate or unstable for subjects with a very low probability of receiving the treatment actually received. A common practice in such cases is to use a trimmed approach or truncation, where all the observations with weights exceeding a specified threshold are either discarded or their weights are set equal to the threshold (Elze et al., 2017). As there is no standard definition of a large weight, we here reduce all the weights  $\hat{w}_i^*$  greater than the 99th percentile of the distribution to that threshold, and we denote the new overall individual weights by  $\hat{w}_i$ .

Once the weights have been computed for all individuals, a weighted version of equation (2) can be maximised to obtain the parameter estimates. The target log-likelihood function is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \hat{w}_i \log p(\tilde{\mathbf{y}}_i | \mathbf{z}_i), \quad (6)$$

where  $p(\tilde{\mathbf{y}}_i | \mathbf{z}_i)$  is the manifest probability of an LM model based on assumptions

$$\begin{aligned} \log \frac{\lambda_{i,u}}{\lambda_{i,1}} &= \alpha_u + \beta_u^{(z_{i1})}, & u = 2, \dots, k, \\ \log \frac{\pi_{it,v|u}}{\pi_{it,u|u}} &= \gamma_{v|u} + \delta_{v|u}^{(z_{it})}, & u, v = 1, \dots, k, \quad u \neq v, \end{aligned}$$

where the initial and transition probabilities are defined as

$$\begin{aligned} \lambda_{i,u} &= P(U_{i1}^{(z_{i1})} = u), \\ \pi_{it,v|u} &= P(U_{it}^{(z_{it})} = v | U_{i,t-1}^{(z_{i,t-1})} = u), \end{aligned}$$

for  $i = 1, \dots, n$ ,  $t = 2, \dots, T$ ,  $u, v = 1, \dots, k$ ,  $u \neq v$ , and  $z = 0, \dots, d - 1$ . We recall that  $z_{it}$  is the observed treatment taken by individual  $i$  at occasion  $t$ .

In order to maximise equation (6), the EM algorithm makes use of the *complete data log-likelihood*:

$$\begin{aligned} \ell^*(\boldsymbol{\theta}) &= \sum_{u=1}^k \sum_{j=1}^r \sum_{t=1}^T \sum_{y=0}^{c_j-1} a_{jt,uy} \log \phi_{jy|u} + \sum_{i=1}^n \sum_{u=1}^k \hat{w}_i b_{i1,u} \log \lambda_{i,u} \\ &+ \sum_{i=1}^n \sum_{u=1}^k \sum_{v=1}^k \sum_{t=2}^T \hat{w}_i b_{it,uv} \log \pi_{it,v|u}, \end{aligned} \quad (7)$$

where  $a_{jt,uy} = \sum_{i=1}^n \hat{w}_i b_{it,u} I(y_{ijt} = y)$  is the weighted frequency of units responding by  $y$  to the  $j$ th response variable and belonging to latent state  $u$  at occasion  $t$ , with  $I(\cdot)$  being the indicator function equal to 1 if its argument is true and 0 otherwise. Moreover,  $b_{it,u}$  is an indicator variable for the unit  $i$  belonging to class  $u$  at occasion  $t$ , and  $b_{it,uv} = b_{i,t-1,u} b_{it,v}$  is an indicator variable equal to 1 if the same unit moves from state  $u$  to state  $v$  at occasion  $t$ . As the latent configuration is not known for each subject, to maximise equation (7) the EM algorithm alternates two steps until convergence:

**E-step** : it consists in computing the expected value of  $\ell^*(\boldsymbol{\theta})$  given the observed data and the current level of parameters. This requires the posterior distribution of the latent variables obtained by the Baum-Welch recursion.

**M-step** : it maximises the expected value of  $\ell^*(\boldsymbol{\theta})$  obtained above with respect to the model parameters  $\boldsymbol{\theta}$ . This is based on a Newton-Raphson algorithm having a structure similar to that for the maximisation of the weighted log-likelihood function of a multinomial logit model.

The posterior estimates obtained at the E-step are used to cluster units according to a modal assignment rule at each time occasion  $t$  (local decoding). Insights about LM modelling with covariates and details on the estimation procedure can be found in Bartolucci et al. (2013, Chapter 5) and Bartolucci et al. (2014).

## 5 Simulation study

The finite-sample properties of the estimator proposed in the previous section are assessed by a simulation study illustrated in the following. We set the number of latent classes

$k$  equal to 2 in connection with the empirical application, where the aim is to cluster households depending on their poverty status. At the same time, the treatment variable is assumed to be dichotomous ( $z_{it} = 0, 1$ ) and will be equal to 1 for remittances recipients. In the simulation, the estimates obtained with the proposed (weighted) approach are compared with the estimates obtained with the standard (unweighted) and possibly biased approach and with the standard estimator under random assignment.

## 5.1 Design

The following design closely traces the simulation method in Bartolucci et al. (2016). We assume the existence of a time-varying dichotomous treatment  $Z_{it}$ , with assignment mechanism of the treatment depending on a covariate  $X_{1it}$  following an AR(1) process with a standard normal error term and an autoregressive coefficient of 0.5. The endogenous treatment is generated by the logit model:

$$\log \frac{p_{it}}{1 - p_{it}} = x_{1it}, \quad (8)$$

where  $p_{it} = P(Z_{it} = 1 | X_{1it} = x_{1it})$  and  $x_{1it}$  is the observed value of  $X_{1it}$ . We also assume the existence of another covariate  $X_{2it}$  affecting the distribution of the latent variable, which is a discrete random variable assuming value  $-1$  or  $1$  with equal probability. The initial and transition probabilities are defined as follows:

$$\log \frac{\lambda_{i,2}}{\lambda_{i,1}} = \alpha^* + \beta^{(1)*} z_{i1} + \nu^* (x_{1i1} + x_{2i1}), \quad (9)$$

$$\log \frac{\pi_{it,v|u}}{\pi_{it,u|u}} = \gamma_{v|u}^* + \delta_{v|u}^{(1)*} z_{it} + \tau_{v|u}^* (x_{1it} + x_{2it}), \quad u, v = 1, 2, \quad u \neq v, \quad (10)$$

with  $\alpha^* = 1$ ,  $\beta^{(1)*} = -1$ ,  $\nu^* = -0.5$ ,  $\gamma_{2|1}^* = -1$ ,  $\delta_{2|1}^{(1)*} = 1.5$ ,  $\tau_{2|1}^* = 0.5$  and  $\gamma_{1|2}^* = 1$ ,  $\delta_{1|2}^{(1)*} = -1.5$ , and  $\tau_{1|2}^* = -0.5$ . Thus the treatment has a negative effect on the initial allocation to class 2 and opposite effects on the transitions between the two classes. The parameters are defined so to have a balance between the two groups along the time dimension. Note that we add an asterisk to the parameters in (9) and (10) as we are making explicit the relation between the latent variable  $U_{it}$  and the confounder  $X_{1it}$  in order to simulate data under endogeneity of the treatment. On the other hand, we do not formulate assumptions on the conditional distribution of  $U_{it}$  given the pretreatment covariate and we consequently estimate a marginal model in the sense of Robins et al. (2000) not including  $X_{1it}$  in the latent model.

The Data Generating Process (DGP) is based on the following steps:

1.  $X_{1it}$  and  $X_{2it}$  are drawn for each individual and time period;
2. the endogenous treatment  $Z_{it}$  is generated according to equation (8);
3. the individual latent states are determined for each time period according to models (9) and (10);
4. given the latent states, six dichotomous response variables  $Y_{ijt}$  are generated according to the probabilities of “success” reported in Table 1, for each individual and time period.

Table 1: Probability of success for the  $r = 6$  response variables

	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$	$j = 6$
$u = 1$	0.3	0.5	0.8	0.8	0.9	0.9
$u = 2$	0.1	0.2	0.3	0.4	0.3	0.6

The response probabilities are assumed to be time-fixed and state 2 has lower probability of success for each response, indicating high class separation. Steps 3 and 4 of the DGP and the estimation routine are carried out using the **R** package **LMest** (Bartolucci et al., 2017).

Once data are generated, the observational weights are estimated running a logistic regression of the model in step 2 (correct specification of the model). Individual weights are thus computed as in equation (5) for the case of a dichotomous treatment:

$$\hat{w}_i^* = \left( \prod_t \hat{p}_{it}^{z_{it}} (1 - \hat{p}_{it})^{1-z_{it}} \right)^{-1}.$$

The final weights  $\hat{w}_i$  are retrieved thresholding the distribution of  $\hat{w}_i^*$  to the 99th percentile.

Using the generated dataset (“endogenous” dataset), an LM model in its standard version (*naive* model) is estimated together with the version of this estimator based on weighting the likelihood contribution of each observation by  $\hat{w}_i$  (*proposed* model). It is useful to stress again that we are not incorporating  $X_{1it}$  in the structural model as the parameters we are estimating are different from the ones reported in equations (9) and (10) and are thus indicated by  $\alpha$ ,  $\beta^{(1)}$ ,  $\nu$ ,  $\gamma_{v|u}$ ,  $\delta_{v|u}^{(1)}$  and  $\tau_{v|u}$  for  $u, v = 1, 2$ ,  $u \neq v$ .

It is important noting that steps 3 and 4 of the DGP are repeated using randomly assigned treatments (i.e., the treatments are assigned with probability 0.5 to each individual). With this new set of variables (“exogenous” dataset) we estimate the *random* model (not making use of individual weights). The exogenous dataset differs from the endogenous one as the treatment is not endogenous anymore.

Finally note that the simulation design presented above encompasses a more general framework with respect to the model presented in Section 4, as we are including an exogenous covariate in both the initial and transition distribution. Thus, the ATE in terms of initial probability is computed averaging the individual partial effects of the treatment on the probability of belonging to class 2 in the first time period, namely

$$\widehat{ATE}_2 = n^{-1} \sum_{i=1}^n \left( \Lambda(\hat{\alpha} + \hat{\beta}^{(1)} + \hat{\nu}x_{2i1}) - \Lambda(\hat{\alpha} + \hat{\nu}x_{2i1}) \right),$$

where  $\Lambda(\cdot)$  is the standard logistic distribution function. Similarly, the ATEs in terms of transition probabilities are averaged across the individuals whose latent state are estimated to be equal to a generic state  $u$  at occasion  $t - 1$  under a modal assignment rule. In this case, all the individuals are pooled, and for  $t > 1$  we have

$$\widehat{ATE}_{v|u} = \left( \sum_{i=1}^n \sum_{t=2}^T \hat{b}_{i,t-1,u} \right)^{-1} \sum_{i=1}^n \sum_{t=2}^T \hat{b}_{i,t-1,u} \left( \Lambda(\hat{\gamma}_{v|u} + \hat{\delta}_{v|u}^{(1)} + \hat{\tau}_{v|u}x_{2it}) - \Lambda(\hat{\gamma}_{v|u} + \hat{\tau}_{v|u}x_{2it}) \right)$$

where  $\hat{b}_{i,t-1,u}$  is equal to 1 if subject  $i$  is assigned to latent state  $u$  at time  $t - 1$ , for  $u, v = 1, 2, u \neq v$ . The benchmark ATEs are computed using the “endogenous” dataset and the parameter estimates of the random model.

The sample size  $n$  is set equal to 1000 and 2000 and the time period  $T$  equal to 3 and 5. The overall procedure is iterated 500 times for each configuration of  $n$  and  $T$ . At each iteration, the standard errors of the estimates are computed by a non-parametric bootstrap procedure with 100 repetitions.

## 5.2 Results

The bias of the naive and proposed estimators are computed with respect to the estimated parameters of the random estimator, averaging the estimates of the models across the Monte Carlo replications. Tables 2 and 3 summarise the results in terms of estimated coefficients.

Table 2: Monte Carlo results: Parameter estimates with  $n = 1000$

Model			$\alpha$	$\beta^{(1)}$	$\nu$	$\gamma_{1 2}$	$\gamma_{1 2}^{(1)}$	$\tau_{1 2}$	$\gamma_{2 1}$	$\gamma_{2 1}^{(1)}$	$\tau_{2 1}$
T = 3	Random	mean	0.949	-0.958	-0.480	0.882	-1.417	-0.473	-0.988	1.422	0.479
		sd	0.146	0.174	0.088	0.180	0.229	0.105	0.155	0.190	0.096
	Naive	bias	0.219	-0.418	-0.001	0.237	-0.509	-0.011	-0.274	0.500	0.008
		sd	0.151	0.185	0.091	0.184	0.249	0.124	0.163	0.198	0.100
	Proposed	bias	0.044	-0.069	-0.004	0.055	-0.097	-0.014	-0.053	0.090	0.010
		sd	0.207	0.248	0.130	0.250	0.330	0.167	0.209	0.275	0.137
	se/sd	1.038	1.034	0.976	1.049	1.015	1.002	1.057	1.009	1.003	
T = 5	Random	mean	0.956	-0.957	-0.480	0.871	-1.400	-0.470	-0.990	1.422	0.473
		sd	0.142	0.173	0.092	0.119	0.151	0.076	0.106	0.127	0.068
	Naive	bias	0.206	-0.404	0.001	0.229	-0.511	-0.008	-0.280	0.492	0.001
		sd	0.152	0.188	0.088	0.131	0.161	0.083	0.115	0.144	0.069
	Proposed	bias	0.027	-0.070	-0.005	0.044	-0.086	-0.003	-0.045	0.083	0.001
		sd	0.241	0.297	0.145	0.207	0.251	0.123	0.188	0.231	0.110
	se/sd	1.000	0.999	1.031	1.019	1.067	1.074	0.994	0.994	1.039	

On the basis of the first simulation results, we observe that the bias of the naive estimator is large and doesn’t decrease in  $n$ . The treatment parameters show the largest biases in absolute term ( $\beta^{(1)}$ ,  $\delta_{2|1}^{(1)}$  and  $\delta_{1|2}^{(1)}$ ), as expected. The parameters related to the exogenous covariate are instead consistently estimated.

On the other hand, the bias of the proposed estimator is negligible and overall tends to decrease as  $n$  grows, although thresholding the weights’ distribution seems to introduce a small bias in the estimator. However, this also produces reasonable standard deviations ( $sd$ ), that are moderately larger than in the naive model and decrease at a rate close to  $\sqrt{2}$  as  $n$  doubles, as expected.

Overall, the proposed estimator sensibly improves the estimation of the ATEs, as seen in Table 4. For each combination of  $n$  and  $T$ , the gain in bias reduction in terms

Table 3: Monte Carlo results: Parameter estimates with  $n = 2000$ 

Model			$\alpha$	$\beta^{(1)}$	$\nu$	$\gamma_{1 2}$	$\gamma_{1 2}^{(1)}$	$\tau_{1 2}$	$\gamma_{2 1}$	$\gamma_{2 1}^{(1)}$	$\tau_{2 1}$	
T = 3	Random	mean	0.943	-0.943	-0.476	0.892	-1.425	-0.473	-0.974	1.391	0.469	
		sd	0.102	0.126	0.062	0.119	0.159	0.078	0.102	0.126	0.063	
	Naive	bias	0.218	-0.416	-0.005	0.207	-0.481	-0.008	-0.274	0.509	0.006	
		sd	0.102	0.128	0.062	0.120	0.158	0.085	0.112	0.139	0.072	
	Proposed	bias	0.024	-0.047	-0.009	0.053	-0.073	-0.002	-0.030	0.053	0.002	
		sd	0.144	0.175	0.084	0.182	0.219	0.120	0.146	0.189	0.093	
		se/sd	1.029	0.972	1.034	1.083	1.080	1.015	1.015	1.022	0.999	
	T = 5	Random	mean	0.942	-0.950	-0.478	0.870	-1.405	-0.469	-0.985	1.413	0.469
			sd	0.102	0.127	0.058	0.089	0.111	0.052	0.080	0.096	0.046
Naive		bias	0.211	-0.407	-0.005	0.222	-0.504	-0.010	-0.282	0.495	0.008	
		sd	0.110	0.131	0.065	0.091	0.115	0.056	0.078	0.095	0.049	
Proposed		bias	0.038	-0.052	-0.009	0.038	-0.084	-0.006	-0.041	0.074	0.003	
		sd	0.165	0.194	0.105	0.137	0.173	0.090	0.127	0.146	0.078	
		se/sd	1.030	1.009	0.997	1.034	1.058	1.026	1.039	1.061	1.023	

of effect on initial probabilities (column  $ATE_2$ ) and transition probabilities (columns  $ATE_{2|1}$  and  $ATE_{1|2}$ ) overcomes the higher variability of the estimator. This is reflected in the considerably lower root mean-square errors ( $rmse$ ) and in the relative efficiency ( $eff$ ) of the proposed with respect to the naive estimator.

## 6 Empirical application

This section describes the application of the proposed methodology, where we estimate the effect of receiving remittances on the poverty status of the households in Uganda. Although since the early 1990s Uganda's progresses in reducing poverty have been remarkable, in 2013 more than a third of its population lived below the international extreme poverty line of USD 1.90 a day (The World Bank, 2016).

In the following, we first describe the available dataset and the context of the study. Then, we show and comment the main results.

### 6.1 Data description

The Uganda National Panel Survey (UNPS) implemented by the Uganda Bureau of Statistics is part of the survey project Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) run by the World Bank. The first round of data collection ran from September 2009 through August 2010. The current work plan spans a period of 5 years, consisting of four waves (2009/2010, 2010/2011, 2011/2012 and 2013/14) and it is representative at the national level.

The UNPS is composed of four questionnaires: Household, Agriculture, Community and Women. Our analysis is based on the Household questionnaire, that provides infor-

Table 4: Monte Carlo results: Average Treatment Effects

			n=1000			n=2000		
			$ATE_2$	$ATE_{2 1}$	$ATE_{1 2}$	$ATE_2$	$ATE_{2 1}$	$ATE_{1 2}$
T = 3	Random	mean	-0.212	-0.319	0.318	-0.209	-0.322	0.313
	Naive	rmse	0.095	0.112	0.106	0.091	0.102	0.105
	Proposed	rmse	0.052	0.067	0.056	0.039	0.044	0.041
		eff	0.549	0.594	0.534	0.429	0.431	0.390
T = 5	Random	mean	-0.211	-0.317	0.319	-0.211	-0.319	0.318
	Naive	rmse	0.093	0.108	0.101	0.089	0.105	0.100
	Proposed	rmse	0.063	0.053	0.048	0.064	0.053	0.048
		eff	0.676	0.492	0.476	0.667	0.491	0.457

mation on education, health, nutrition and income of each household members as well as consumption level and welfare indicators at the household level. The final sample of our analysis consists of 1,095 households interviewed in all the four waves of the survey. The LM model is estimated using the variables from the last three waves ( $T = 3$ ), as in the propensity score models we make use of two lagged variable. The following descriptives are thus based on the 2010/2011, 2011/2012, and 2013/2014 waves.

The treatment variable  $Z_{it}$  equals 1 if the household receives remittances or assistance either from abroad or locally (i.e., from elsewhere in the country) in the year prior to the interview. Although the effect of international and domestic remittances may differ, the low share of households reporting remittances from abroad (2.4% of the overall sample) and the relatively low amount of money of international remittances<sup>8</sup> justify such a pooling.

Almost 27% of the pooled sample belongs to the treatment group (i.e., 876 over 3,285 observations, see Table 7). The share of treated households is stable and varies from a minimum of 24.9% in 2010 to a maximum of 27.9% in 2011. Though many Ugandan households report receiving remittances, the amount of these transfers comprise a low share of the total consumption at the household level. Even if the proportion of treated households is almost the same in all the consumption classes, ranging from 24.6% in the third quartile to 29.4% in the fourth quartile, remittances are a slightly larger share for the upper classes, increasing from 3.9% in the lower quartile to 9.9% in the top quartile. Interviewed households were also asked what were the common uses for the remittances and assistance received<sup>9</sup>. On average in the three periods, goods and services consumption were indicated as the main use of the money received by 60% of the respondents, followed by other expenses (e.g., education, health, etc). Only a small fraction, corresponding to around 1.5%, reported investing in capital inputs or building materials. Overall, such a framework suggests that current consumption is the main use of the amount of money

<sup>8</sup>On average and for the whole sample, almost 35,000 Ugandan Shillings at 2010 prices. On the other hand, the average amount of internal remittances is almost 7 times higher and corresponds to 242,250 Ugandan Shillings at 2010 prices.

<sup>9</sup>Attention must be paid to the interpretation these types of question. As McKenzie and Sasin (2007) pointed out, money is fungible and thus distinguishing its use according to the different type of income sources is not trivial.

received and justify the assumptions of contemporaneous effect of remittances on the status of poverty in the adopted models based on assumptions (3) and (4).

Table 5: Response variables: Deprivation indicators

Indicator	Deprived if ...
Shoes	At least one household member hasn't got ... ... a pair of shoes
Clothes	... two pair of clothes
Assets	The household doesn't have at least one asset related to access to information and at least one asset related to mobility
FCS	Poor/Boarderline/Adequate food consumption in the past week

The  $r = 4$  selected response variables are the Food Consumption Score (FCS) threshold variable as defined by the World Food Programme (WFP, 2008) and three indicators of material deprivation as listed in Table 5.

For each household, we compute the FCS as a composite score on the basis of dietary diversity, food frequency, and relative nutritional importance of different food groups, based on the household consumption in the week prior to the interview. The scores are then classified in three categories. Labels of the variable are 0 for “poor consumption”, 1 and 2 for “borderline” and “adequate consumption”, respectively (WFP, 2008). *Shoes* and *Clothes* are dichotomous variables from the Welfare section of the questionnaire (GSEC 17/17A), equal to 1 if at least one household member has not got one pair of shoes or two pair of clothes respectively. Finally, *Assets* is a binary variable inspired by the multidimensional poverty index (Alkire et al., 2018) equal to 1 if the household does not have at least one asset related to access to information (TVs, radios or mobile phones) and one asset related to mobility (bicycles, motorcycles, other motor vehicles and boats).

Table 6 reports summary statistics of the response variables from the  $T$  time periods of analysis. As we can see, the percentage of households being deprived decreases over

Table 6: Response variables' summary statistics

Wave	10-11		11-12		13-14		min	max
	mean	sd	mean	sd	mean	sd		
Assets	0.10	0.30	0.09	0.29	0.09	0.28	0	1
Clothes	0.11	0.32	0.09	0.28	0.05	0.22	0	1
Shoes	0.48	0.50	0.44	0.50	0.38	0.49	0	1
FCS	1.82	0.44	1.90	0.35	1.93	0.29	0	2
$n$	1095							

time for each indicator, with the mean of the item *Shoes* reducing the most. On the other

hand, the number of households with an adequate food consumption slightly increases.

The covariates included in the propensity score model, pooled from the three time periods, are listed in Table 7 along with their means for both recipient and nonrecipient households.

Table 7: Propensity score variables' summary statistics (pooled sample)

	Remittances		diff <sup>a</sup>
	Recipients	Non recipients	
Average education	6.87	6.51	***
Urban	0.26	0.20	***
Female HOH	0.49	0.18	***
Age HOH	54.09	45.05	***
Married HOH	0.56	0.84	***
Males	2.88	3.38	***
Child	1.07	1.53	***
Elderly	2.82	2.78	
Households	876	2409	

<sup>a</sup> Two-tailed tests for equal means/proportions.

Significance levels: \* 0.10 \*\* 0.05 \*\*\* 0.01

*Average education* corresponds the mean years of schooling among the family members who are not students. *Urban* and *Female HOH* are dummy variables equal to 1 if the household lives in an urban area and the household head is a woman. *Age* and *Married HOH* are the age of the household head and a dummy variable for head being married respectively. *Males* is the number of male members of the household, *Child* and *Elderly* are the numbers of children aged 0-6 and adults aged more than 62.

The last column of Table 7 reports the results of testing the null hypothesis of equality of the variable means (proportions for dummy variables) between the treated and control groups, on the basis of the  $p$ -value of the tests. The null is rejected for all the variables, except *Elderly*, at the 1% level, indicating the presence of selectivity into the treatment. Similar results, that are here not shown, hold when conducting the tests cross-sectionally.

## 6.2 Results

We first report the results of the first step of the estimation process, aimed to obtain the propensity score weights, and then those of the second step, finalised to estimate the ATEs.

### 6.2.1 Propensity score weighting and covariates balancing

Table 8 shows the results from the propensity score logistic regressions used at the first step. As two lagged variables are included in the model specification, the last three waves

of the survey are used to estimate the LM models, and the probability of exposure to the treatment is conditioned on the variables of current and previous waves.

Table 8: Parameter estimates of the propensity score model

	Wave		
	10-11	11-12	13-14
Constant	-3.895***	-4.867***	-4.675***
Average Education	0.013	0.061**	0.068**
Urban	0.125	-0.031	-0.031
Female HOH	0.992***	1.222***	0.725***
Age HOH	0.032***	0.045***	0.047***
Married HOH	0.041	0.011	-0.370
Males	-0.105*	-0.136**	0.079
Child	-0.048	0.095	-0.071
Elderly	-0.002	0.128*	-0.129*
Main income ( $t - 1$ )	1.368***	1.159***	0.626
$Z_{t-1}$	1.514***	1.093***	0.615***
Pseudo R-sq	0.2561	0.2371	0.1894
n	1095	1095	1095

Region dummies included as controls.

Significance levels: \* 0.10 \*\* 0.05 \*\*\* 0.01

The two lagged variables are  $Z_{t-1}$ , that is the treatment variable lagged by one period and *Main income*, a dummy equal to 1 if remittances were the household's main source of income in the previous year.

Households with a female head are more likely to be recipients, just as households with older heads. Living in an urban area and having children of 0-6 years old are not significant determinants of being recipients. On the other hand, households with more male members are less likely to receive remittances. Other important positive determinants are instead the fact of having had remittances in the previous year, and that past remittances were the main source of income, indicating that remittances are time varying but also quite persistent.

Table 9 presents the standardised mean differences of the covariates among the treated and control groups, raw (unweighted), and weighted. The balance between the two groups is considerably higher after the propensity score adjustment based on inverse propensity weights. In particular, the balancing property is satisfied for each variable after weighting the data<sup>10</sup>, while prior to the adjustment it is not satisfied for all the variables except for

<sup>10</sup>Covariates balancing is achieved if the standardised mean differences are less than the usual rule of thumb threshold of 0.1 in absolute values.

Table 9: Standardized mean difference for each covariate in the propensity score model

	10-11		11-12		13-14	
	R	W	R	W	R	W
Main income (t-1)	.52	.03	.48	.02	.40	.01
Average Education	.06	.06	.16	.00	.12	.03
Urban	.19	.04	.17	.05	.07	.04
Female HOH	.79	.01	.73	.09	.58	.02
Age HOH	.57	.03	.70	.08	.66	.01
Married HOH	.68	.03	.64	.08	.61	.01
Males	.34	.01	.31	.03	.17	.05
Child	.43	.02	.35	.09	.36	.07
Elderly	.00	.00	.09	.03	.00	.02
$Z_{t-1}$	1.09	.01	.89	.03	.65	.02

Absolute values. R = raw; W = weighted.

the number of elderly men, as in Table 7.

### 6.2.2 Model fitting

As already mentioned, the number of latent classes is set to  $k = 2$  for a matter of interpretability of the estimated parameters. The households are thus clustered in two groups, ideally identifying a class of materially deprived and a class of non-deprived households. The number of free parameters to be estimated is equal to 10 in the measurement model, 2 in the initial distribution, and 4 in the transition distribution.

From the estimated response probabilities, denoted by  $\hat{\phi}_{ryr|u}$ , it is clear that the second class identifies poor households. As shown in Table 10, the probability of being deprived in each item is substantially higher for households clustered in the second group. The bigger difference is found in the variable *Shoes*, the most discriminating item, and it amounts to a 71 percentage points spread, while for the indicators *Clothes* and *Assets* the difference consists in 19 and 26 points respectively. In particular, households clustered in the second group have a 28% probability of being deprived in the *Assets* dimension, and the probability that at least one member lacks two pair of clothes is 0.21. Both these probabilities reduce to 2 percentage points for households belonging to class one. Also the probability of having an adequate food consumption is 17 percentage points lower in class two, and it amounts to 0.77. On the other hand, having a poor and borderline food consumption is less likely for households clustered in group 1 (1% and 5% respectively).

Households are grouped on the basis of the posterior probability of belonging to the latent class of poverty, by the means of a modal assignment rule. The number of households classified as materially deprived decreases from approximately 40% at the beginning of the period to 28% by the end of it. Among the recipients of remittances,

Table 10: Conditional response probabilities estimates

	$c_j$	u = 1	u = 2
FCS	0	0.01	0.03
	1	0.05	0.19
	2	0.94	0.77
Assets	0	0.98	0.72
	1	0.02	0.28
Shoes	0	0.82	0.11
	1	0.18	0.89
Clothes	0	0.98	0.79
	1	0.02	0.21

only 1 in 4 households is identified as belonging to group 2 in the whole sample. In the treated group, the share of households clustered in class 2 is 29% at the beginning of the period and 22% by the end of it, roughly 26% on average in the three time periods. Once we weigh the observations with the propensity weights of step 1, the share of treated households in class 2 rises to 60% and it is stable along the time dimension. We conclude that a better balancing among the groups is achieved after propensity weighting.

A general tendency observed throughout the time span of the analysis is a strong persistency of the latent states, with a substantially larger shift from the class of materially deprived households to the class of non material deprivation. The average estimated transition matrix is given by <sup>11</sup>:

$$\Pi = \begin{bmatrix} \bar{\pi}_{1|1} & \bar{\pi}_{2|1} \\ \bar{\pi}_{1|2} & \bar{\pi}_{2|2} \end{bmatrix} = \begin{bmatrix} 0.962 & 0.038 \\ 0.209 & 0.791 \end{bmatrix}.$$

Such a persistence tendency is reflected on the negative sign of the estimated intercepts in the transition distributions of Table 11 (Panel A, rows 3 and 5).

From the second row of Table 11 (Panel A), the treatment has a negative impact on the probability of belonging to the group of materially deprived households. The coefficient is negative ( $\hat{\beta}_2 = -0.549$ ) and statistically significant almost at the 5% level and it corresponds to a negative effect of 13.1% (Panel B, first row).

On the other hand, a significant effect is observed in terms of transition from state 1 to 2. The negative coefficient ( $\hat{\gamma}_{2|2,1} = -10.62$ ) is large and significant at the 5% level and it determines a negative partial effects of 5.5 percentage points, shrinking the probability of moving to state 2 almost to zero. However, the treatment plays no role in transitioning from class 2 to 1. As we see in the last row of Table 11, Panel A, the treatment coefficient is counterintuitively negative, though not statistically different from zero.

In this sense remittances act as a preventive measure, that is, as a measure to avert deprivation, rather than a measure aiming at escaping from poverty. Our findings are con-

<sup>11</sup>The transition matrix is obtained averaging across all the individuals' transition probabilities and time occasions.

Table 11: Initial and transition distributions' estimates

Panel A: Parameter Estimates				
	Estimate	S.E.	z-stat.	p-value
Initial distr.				
$\alpha_2$	-0.121	0.242	-0.500	0.617
$\beta_2^{(1)}$	-0.549	0.281	-1.951	0.051
Transition distr.				
$\gamma_{2 1}$	-2.892	1.818	-1.590	0.112
$\delta_{2 1}^{(1)}$	-10.365	4.796	-2.161	0.031
$\gamma_{1 2}$	-1.285	0.305	-4.213	0.000
$\delta_{1 2}^{(1)}$	-0.177	0.470	-0.337	0.706
Panel B: Treatment effects				
	Estimate	S.E.	z-stat.	p-value
$ATE_2$	-0.131	0.066	-1.995	0.046
$ATE_{2 1}$	-0.053	0.031	-1.695	0.090
$ATE_{1 2}$	-0.029	0.068	-0.424	0.672

sistent with the strand of literature in which migration is considered as an informal social insurance mechanism and remittances help stabilising consumption over time (Kurosaki, 2006; Amuedo-Dorantes and Pozo, 2011).

In order to assess how these conclusions are sensitive to the initially selected set of covariates, the same analysis was carried out restricting the set of exogenous covariates in the propensity score model to the significant determinants of the treatment, omitting variables *Urban*, *Married HOH*, and *Child* (see Table 8). As a further robustness specification, the original model was estimated including the 2013 longitudinal sample weights associated to each household<sup>12</sup>. The final weights are computed by taking the product of the inverse propensity weights and the sampling weights. The estimated parameters of both these versions of the model are almost identical to those presented, leading to the same conclusions in terms of effect of the treatment of interest.

## 7 Conclusions

In order to evaluate the impact of remittances on poverty, we propose a novel methodology for the assessment of a time-varying treatment effect on a latent variable in a longitudinal framework, when selection into treatment is a possible source of bias of the causal effect. Our method combines a propensity score based technique, the inverse propensity weighting, with a dynamic latent class model, namely the Latent Markov model (LM;

<sup>12</sup>The sample size  $n$  here reduces from 1095 to 1042 households due to missing weights.

Bartolucci et al., 2013), where the individual-specific latent process is assumed to follow a first-order Markov chain, possibly time-inhomogeneous. We set the number of latent classes equal to two, clustering observations into a group of deprived and a group of non-deprived households.

The treatment effect of interest is analysed in terms of initial distribution of the latent grouping variable, and in terms of transition among the latent classes. The estimation is based on a two-step procedure. The first step consists in estimating propensity score weights at the household level. Then, a weighted version of the LM model likelihood is maximised with respect to the model parameters. In this model the only covariate considered is the indicator variable for the treatment, which affects the initial and transition probabilities by a logit parametrisation. The finite-sample properties of the estimator are investigated via Monte Carlo simulations. Weighting the contribution of each sample unit to the likelihood function allows us to obtain substantially unbiased parameter estimates and allows for a proper causal interpretation of the treatment partial effect.

The model is estimated using data from the Household questionnaire of the Uganda National Panel Survey. Treatment confounders are derived from the migration literature and consist in a set of households' characteristics. The latent variable identifies the poverty status of the households as proxied by 4 response variables that identify different dimensions of material deprivation. Clustering households into two groups permits to disentangle whether remittances recipients are classified in the deprived or in the non deprived group. From the estimated response probability, we have a clear indication of class 2 as the group of deprived households. Our findings show that most of the recipients belong to the nonpoors. Accounting for the selectivity into treatment, we find a negative effect in terms of initial probability of belonging to the second group. While we see no impact in reference to transition from group 2 to 1, a positive effect is estimated for the opposite transition. We conclude that remittances act at the household level as a protective/preventive measure rather than a measure aiming at mitigating poverty. However, further work is needed in order to exploit the channels through which remittances operate as such measures. We also want to stress that our conclusions depend on the choice of the material deprivation indicators used as response variables (see Table 5). The role of remittances may be reconsidered when taking into consideration other aspects of the social life of individuals and households, such as the level of education, the health status, the participation in social life, the social capital, etc. However, we consider other dimensions as less appropriate for our study, given the short time span of the analysis. Moreover, the inclusion of more indicators would be less compatible with the assumption of two latent classes and setting  $k$  greater than 2 would result in a less straightforward interpretation of the parameters. Finally, a limitation of our approach is that, as with any propensity score analysis, we assumed that all of the confounders were measured and included in the first step of our estimation strategy.

## References

- Adams, R. H., J.  
2011. Evaluating the economic impact of international remittances on developing countries using household surveys: A literature review. *The Journal of Development Studies*, 47:809–828.
- Agresti, A.  
2002. *Categorical data analysis*, 2nd edition. Hoboken, N.J. : John Wiley & Sons.
- Alkire, S. and J. Foster  
2011a. Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95:476–487.
- Alkire, S. and J. Foster  
2011b. Understandings and misunderstandings of multidimensional poverty measurement. *Journal of Economic Inequality*, 9:289–314.
- Alkire, S., U. Kanagaratnam, and N. Suppa  
2018. *The Global Multidimensional Poverty Index (MPI): 2018 revision*. OPHI MPI Methodological Notes 46, Oxford Poverty and Human Development Initiative, University of Oxford.
- Altman, R. M.  
2007. Mixed Hidden Markov Models. *Journal of the American Statistical Association*, 102:201–210.
- Amuedo-Dorantes, C. and S. Pozo  
2011. Remittances and Income Smoothing. *The American Economic Review*, 101:582–587.
- Atkinson, A. B.  
2003. Multidimensional deprivation: contrasting social welfare and counting approaches. *The Journal of Economic Inequality*, 1:51–65.
- Bartolucci, F., A. Farcomeni, and F. Pennoni  
2013. *Latent Markov Models for Longitudinal Data*. Boca Raton, FL: Chapman and Hall/CRC press.
- Bartolucci, F., A. Farcomeni, and F. Pennoni  
2014. Latent markov models: a review of a general framework for the analysis of longitudinal data with covariates. *Test*, 23:433–486.
- Bartolucci, F., S. Pandolfi, and F. Pennoni  
2017. LMest: An R Package for Latent Markov Models for Longitudinal Categorical Data. *Journal of Statistical Software*, 81:1–38.
- Bartolucci, F., F. Pennoni, and G. Vittadini  
2011. Assessment of School Performance Through a Multilevel Latent Markov Rasch Model. *Journal of Educational and Behavioral Statistics*, 36:491–522.

- Bartolucci, F., F. Pennoni, and G. Vittadini  
 2016. Causal Latent Markov Model for the Comparison of Multiple Treatments in Observational Longitudinal Studies. *Journal of Educational and Behavioral Statistics*, 41:146–179.
- Baum, L., T. Petrie, G. Soules, and N. Weiss  
 1970. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41:164–171.
- De Haas, H.  
 2010. Migration and Development: A Theoretical Perspective. *International Migration Review*, 44:227–264.
- Dempster, A. P., N. M. Laird, and D. B. Rubin  
 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39:1–38.
- Démurger, S.  
 2015. Migration and families left behind. *IZA World of Labor*, 144.
- Devereux, S. and R. Sabates-Wheeler  
 2004. Transformative social protection. *IDS Working Paper 232*.
- Dotto, F., A. Farcomeni, M. G. Pittau, and R. Zelli  
 2018. A dynamic inhomogeneous latent state model for measuring material deprivation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
- Efron, B. and R. J. Tibshirani  
 1994. *An introduction to the bootstrap*. Boca Raton, FL: CRC press.
- Elze, M. C., J. Gregson, U. Baber, E. Williamson, S. Sartori, R. Mehran, M. Nichols, G. W. Stone, and S. J. Pocock  
 2017. Comparison of Propensity Score Methods and Covariate Adjustment: Evaluation in 4 Cardiovascular Studies. *Journal of the American College of Cardiology*, 69:345–357.
- Heckman, J. J., H. Ichimura, and P. E. Todd  
 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 6:605–654.
- Krishnakumar, J.  
 2008. Multidimensional Measures of Poverty and Well-being Based on Latent Variable Models. In *Quantitative Approaches to Multidimensional Poverty Measurement*, N. Kakwani and J. Silver, eds., chapter 7, Pp. 118–134. London: Palgrave Macmillan.
- Kurosaki, T.  
 2006. Consumption vulnerability to risk in rural Pakistan. *The Journal of Development Studies*, 42:70–89.
- Lanza, S. T., D. L. Coffman, and S. Xu  
 2013. Causal inference in latent class analysis. *Structural Equation Modeling: a Multidisciplinary Journal*, 20:361–383.

- Lazarsfeld, P. F. and N. W. Henry  
1969. *Latent Structure Analysis*. Boston, MA: Houghton Mifflin.
- Lucas, R. E. and O. Stark  
1985. Motivations to Remit: Evidence from Botswana. *Journal of Political Economy*, 93:901–918.
- Machado, C., C. D. Paulino, and F. Nunes  
2009. Deprivation analysis based on Bayesian latent class models. *Journal of Applied Statistics*, 36:871–891.
- Maruotti, A.  
2011. Mixed Hidden Markov Models for Longitudinal Data: An Overview. *International Statistical Review*, 79:427–454.
- McCaffrey, D. F., B. A. Griffin, D. Almirall, M. E. Slaughter, R. Ramchand, and L. F. Burgette  
2013. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, 32:3388–3414.
- McKenzie, D. and M. J. Sasin  
2007. Migration, remittances, poverty, and human capital: conceptual and empirical challenges. *The World Bank Policy Research Working Paper 4272*.
- McKenzie, D. and D. Yang  
2012. Experimental Approaches in Migration Studies. In *Handbook of Research Methods in Migration*, C. Vargas-Silva, ed., chapter 12. Edward Elgar Publishing.
- Moisio, P.  
2004. A Latent Class Application to the Multidimensional Measurement of Poverty. *Quality & Quantity*, 38:703–717.
- Rapoport, H. and F. Docquier  
2006. The Economics of Migrants’ Remittances. In *Handbook of the Economics of Giving, Altruism and Reciprocity*, S. Kolm and J. M. Ythier, eds., volume 1, chapter 17, Pp. 1135–1198. Elsevier.
- Ratha, D., S. Mohapatra, and E. Scheja  
2011. Impact of migration on economic and social development: A review of evidence and emerging issues. *The World Bank Policy Research Working Paper 5558*.
- Robins, J. M., M. A. Hernan, and B. Brumback  
2000. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560.
- Robins, J. M., A. Rotnitzky, and L. P. Zhao  
1995. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, 90:106–121.
- Rosenbaum, P. R. and D. B. Rubin  
1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

- Rubin, D. B.  
1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.
- Sen, A. K.  
1976. Poverty: An Ordinal Approach to Measurement. *Econometrica*, 44:219–231.
- Sen, A. K.  
1980. Equality of What? In *The Tanner Lecture on Human Values*, Pp. 197–220. Cambridge: Cambridge University Press.
- Sen, A. K.  
1981. *Poverty and Famines: An Essay on Entitlement and Deprivation*. Oxford: Clarendon Press.
- Stark, O. and D. E. Bloom  
1985. The New Economics of Labor Migration. *The American Economic Review*, 75:173–178.
- The World Bank  
2016. *The Uganda Poverty Assessment Report 2016*. Washington: The World Bank Group.
- Townsend, P.  
1979. *Poverty in the United Kingdom*. London: Allen Lane and Penguin Books.
- Townsend, P.  
1987. Deprivation. *Journal of the Social Policy*, 16:125–146.
- UNDP  
2016. *Human Development Report*. New York, NY: United Nations Development Programme.
- Welch, L. R.  
2003. Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter*, 53:1–13.
- WFP  
2008. *Technical Guidance Sheet - Food Consumption Analysis: Calculation and Use of the Food Consumption Score in Food Security Analysis*. Rome, IT: World Food Programme.
- Whelan, C. T. and B. Maitre  
2006. Comparing poverty and deprivation dynamics: Issues of reliability and validity. *The Journal of Economic Inequality*, 4:303–323.
- Zucchini, W. and I. L. MacDonald  
2009. *Hidden Markov models for time series: An introduction using R*. Boca Raton, FL: Chapman and Hall/CRC press.