Statistical Methods and Applications manuscript No.
(will be inserted by the editor)

A weighted χ^2 test to detect the presence of a major change point in non-stationary Markov chains

Alessandra Micheletti · Giacomo Aletti · Giulia Ferrandi · Danilo Bertoni · Daniele Cavicchioli · Roberto Pretolani

Received: date / Accepted: date

Abstract The problem of detecting a major change point in a stochastic process is often of interest in applications, in particular when the effects of modifications of some external variables, on the process itself, must be identified. We here propose a modification of the classical Pearson χ^2 test to detect the presence of such major change point in the transition probabilities of an inhomogeneous discrete time Markov Chain, taking values in a finite space. The test can be applied also in presence of big identically distributed samples of the Markov Chain under study, which might not be necessarily independent. The test is based on the maximum likelihood estimate of the size of the 'right' experimental unit, i.e. the units that must be aggregated to filter out the small scale variability of the transition probabilities. We here apply our test both to simulated data and to a real dataset, to study the impact, on farmland uses, of the new Common Agricultural Policy, which entered into force in EU in 2015.

Keywords Weighted χ^2 test · Inhomogeneous discrete time Markov Chains · Nonparametric Inference

1 Introduction

In the study of time-dependent data, it is often of interest the detection of instants of major change in the process distribution, in order to quantify the effects of some external variables which have been modified to influence the considered process. This is of interest for example when some measures to limit the air pollution of an urban area enter into force (stop of polluting vehicles in specific days, enlargement

A. Micheletti, G. Aletti, D. Bertoni, D. Cavicchioli, R. Pretolani
Department of Environmental Science and Policy
Universitá degli Studi di Milano
E-mail: {alessandra.micheletti}{giacomo.aletti}{danilo.bertoni}{daniele.cavicchioli}{roberto.pretolani}@unimi.it

G.Ferrandi
Department of Mathematics and Computer Science
TU Eindhoven
E-mail: giulietta.ferrandi@gmail.com

of pedestrian zones, etc.) and we then want to detect if a major variation in the spatial distribution of the measured pollutants is observed and after how much time; or when a new economical policy is adopted by a state and the time when its effects on the GDP of different regions become relevant must be identified; etc. All these examples have in common a period in which the system under study is assumed to have a "standard behaviour", meaning that changes in the distribution of the system may happen but in a "smooth way" and with a small variance. This period precedes the modification of the external variable and can be used as a *reference period*.

We consider only discrete time dependent stochastic processes and we assume that the process can take values in a finite set of states, i.e. we consider the stochastic process

$$X: =\{X(t)\}_{t\in\mathbb{N}}$$

with $X(t)\in S$ with S having cardinality $k<\infty$. Further we assume that X is a Markov Chain, that is

$$E[X(t)|X(s),\ s<t]=E[X(t)|X(t-1)] \qquad \forall t\in\mathbb{N},$$

with transition probability matrix $P(t)=[p_{ij}(t)]$, where $p_{ij}(t)=P[X(t)=j|X(t-1)=i]$. We also assume that $p_{ij}(t)>0, \forall i,j\in S, \forall t\in\mathbb{N}$.

We say that the Markov Chain X is *stationary* if its transition probability matrix is constant in time, i.e. $p_{ij}(t)=p_{ij},\ \forall t\in\mathbb{N}$.

Methods to identify point of changes in the parameters of time series or of stochastic differential equations have been widely studied (see e.g. Moreno, Casella, and Garcia-Ferrer 2005; De Gregorio and Iacus 2008; Ringstad Olsen, Chaudhuri, and Godtliebsen 2008; Schütz and Holschneider 2011; Iacus and Yoshida 2012), but since they are applied to general stochastic processes, they don't exploit the markovianity that is characterizing our case study and they don't assume to have a sample (of big size) of copies of the considered process.

In order to test if any major change occurred in the distribution of the process X between time t and $t+1$ we will use a weighted χ^2 test. It is well known [Knoke, Bohrnstedt, and Potter Mee 2002; Eisinger and Chen 2017] that the Pearson χ^2 test for goodness of fit is very sensitive to the sample size N, since the Pearson χ^2 statistics, given by

$$C=\sum_{i=1}^{n_{rows}}\sum_{j=1}^{n_{columns}}\frac{(O_{ij}-E_{ij})^2}{E_{ij}},$$

(where O_{ij} and E_{ij} are the observed and expected frequencies, respectively) is directly proportional to N. In fact, multiplying by a factor F every cell frequency in a crosstabulation, will result in a C statistics multiplied by the same factor F, but with no changes in the degrees of freedom of the asymptotic χ^2 statistics, which remain $(n_{rows}-1)(n_{columns}-1)$. This implies that when N is large, even small deviations between the expected and the observed frequencies produce a C statistics bigger than the critical value, and are thus significant.

We will consider contexts where the sample size N is large and we will reduce it by aggregating the sample units into a new experimental unit, that will provide a C statistics insensitive to the variation of the process X in the reference period.

In general the process X is not assumed to be stationary, but in order to set up our test and the related estimators, we will assume the stationarity of X in the reference period. This corresponds to filter out the “natural variability” of the process.

The structure of the paper is the following: in Section 2 we introduce the modified χ^2 test, and we provide a maximum likelihood estimator of the experimental unit by which we rescale the countings in the test; we also recover the distribution of such estimator, and we exploit it to build confidence intervals for the estimated parameter. In Section 3 we test the effectiveness of the weighted χ^2 test on a simulated non-homogeneous Markov Chain, showing a jump in the transition probabilities at a given time instant. We show that our test is able to identify the jump, and also that by applying the χ^2 test without the proposed rescaling, all the transitions of the Markov Chain result significantly different from one another.

Finally in Section 4 we will apply our weighted χ^2 test to study the impact of the new Common Agricultural Policy (CAP), which entered into force in the European Union from 2015, on the distribution of farmland uses in southern Lombardy region, one of the most intensively cultivated areas in Italy. The new CAP has introduced a new funding policy, called *greening*, which is conditioned to the compliance of farmers with some ecological practices, with some mandatory eco-friendly farming practices. Such farming practices regard, and potentially influence, farmland allocation, particularly that of arable land and grassland [Solazzo, Donati, and Arfini 2015; Cortignani, Severini, and Dono 2017]. We will apply our test to a sample of 638,952 land parcels in Lombardy, occupying a total area of 743,072 hectares, whose land use has been observed in the years 2011-2016. The application of our test reveals a significant change in the land use starting from 2015, and this change could not have been identified without filtering out the natural variability of the process up to 2014. A detailed study of this application, from an agricultural economics point of view, can be found in Bertoni et al. 2018.

2 The weighted χ^2 test

We observe the values assumed by X in the integer time units $t = 0, \dots, m$. We assume to observe M copies (not necessarily independent), $X_1, \dots, X_M$, of the process X. We further assume that the variables $X_j(t)$ take values in the finite set of states $S = \{1, \dots, k\}$.

For any fixed t, from the given samples, we compute the transition frequencies, that is we compute $m_{i1}(t), \dots, m_{ik}(t), i \in S$, where

$$m_{ij}(t) = \#\{l : X_l(t-1) = i, X_l(t) = j\}.$$

We now assume that our process is stationary, i.e. the observed frequencies do not significantly vary from one instant $t-1$ to the next one t. We want then to rescale our experimental units, that is our unit of measure, so that the assumption of stationarity is satisfied. Thus our aim is to estimate the number $\{\mathcal{U}_i, i \in S\}$ of units, evolving from state i, that must be “aggregated” to filter out the small scale variability that violates the assumption of stationarity. Using the working example to which we will apply our test in Section 4, if the original units are the hectares of land cultivated with crop i, the new experimental units will be groups of $\mathcal{U}_i$ hectares of land cultivated with crop i.

For any state $i \in S$, we denote by $M_i(t) = \sum_j m_{ij}(t) = \#\{l : X_l(t-1) = i\}$ the frequency of state i immediately before time t. The rescaled transition frequencies will then be $n_{ij}(t) = \frac{m_{ij}(t)}{\mathcal{U}_i}$ = *amount of new experimental units in the sample which assume the values j at time t and i at time t−1*, and the total amount of new experimental units in state i before time t will be $N_i(t) := \frac{M_i(t)}{\mathcal{U}_i} = \frac{\sum_j m_{ij}(t)}{\mathcal{U}_i} = \sum_j n_{ij}(t)$. Note that the quantities $N_i(t)$ and $n_{ij}(t)$ are not any more necessarily integer numbers.

Consider now two consecutive time units, $t-1$ and t. The assumption of stationarity implies that the following set of hypotheses are satisfied

$$H_0 : p_{ij}(t-1) = p_{ij}(t) \qquad i, j \in S, t = 1, \dots, m.$$

For any time t, we consider the estimate $\hat{P}(t)$ of the transition matrix $P(t)$, with elements given by

$$\hat{p}_{ij}(t) = \frac{m_{ij}(t)}{M_i(t)} = \frac{n_{ij}(t)}{N_i(t)} \qquad i, j \in S.$$

For each starting state $i \in S$, we build the Pearson χ^2 statistics, given by

$$\begin{aligned} Q_i(t) &= \sum_{j=1}^{k} \frac{(n_{ij}(t) - N_i(t)\hat{p}_{ij}(t-1))^2}{N_i(t)\hat{p}_{ij}(t-1)} \\ &= \sum_{j=1}^{k} \frac{(\frac{m_{ij}(t)}{\mathcal{U}_i} - \frac{M_i(t)}{\mathcal{U}_i}\hat{p}_{ij}(t-1))^2}{\frac{M_i(t)}{\mathcal{U}_i}\hat{p}_{ij}(t-1)} \\ &= \frac{1}{\mathcal{U}_i} \sum_{j=1}^{k} \frac{(m_{ij}(t) - M_i(t)\hat{p}_{ij}(t-1))^2}{M_i(t)\hat{p}_{ij}(t-1)}. \end{aligned} \tag{1}$$

We assume that the $N_i(t)$ rescaled random processes $\tilde{X}_1(t), \dots, \tilde{X}_{N_i(t)}$, obtained by weighting the original sample with $\mathcal{U}_i$, are now independent. This implies that the statistics $Q_i(t)$ are asymptotically distributed as χ^2_{k-1} [Fisz 1963], when $M_i \to \infty, \forall i \in S$, since each vector $[n_{ij}(t), j \in S]$ asymptotically approaches a multinomial distribution, with parameters $Multinomial(N_i(t), [\hat{p}_{ij}(t-1), j \in S])$. In fact, when $M_i \to \infty$, the fractional parts of $n_{ij}(t)$ and $N_i(t)$ become negligible with respect to their integer part.

Remark 1. Note then that our test is distribution free but is an asymptotic one, and will work only when the number of experimental units $M \to \infty$, exactly like the classical Pearson χ^2 test. The assumption that all the transition probabilities $p_{ij}(t)$ are strictly positive implies that when $M \to \infty$ also $M_i \to \infty, \forall i \in S$.

Remark 2. The assumption that the $N_i(t)$ rescaled random processes $\tilde{X}_1(t), \dots, \tilde{X}_{N_i(t)}$ are independent can be interpreted as follows, with reference to our working example in Section 4: hectares which are part of the same farm, or belong to neighbouring farms will in general not be independent, since each farmer will split the crops that she/he wants to cultivate in a specific way in the hectares of her/his farm, taking into account the characteristics of the ground, the behaviour of the neighbours, the presence of natural barriers (rivers, streets, etc.), and many other factors. What we assume is that there is a scale of aggregation of the experimental units (i.e. the hectares) at which the decision to change from one crop to another one in two subsequent years becomes independent from the behaviour of the other experimental units (i.e. from the behaviour of groups of hectares "far apart" from the considered rescaled experimental unit).

From our sequence of samples taken at time units $t = 0, \ldots, m$, we can then consider the m couples of consecutive time units and build a sample $Q_i(1), \ldots, Q_i(m)$ of Pearson χ^2 statistics. These statistics can be assumed asymptotically independent, because of the Markov property and the consistency of the sample estimators $\hat{p}_{ij}$ of the transition matrix.

Let us now assume that the $N_i(t)$ are sufficiently big to approximate the distribution of $Q_i(t)$ with a χ^2_{k-1}, for all i, t.

Reminding that the χ^2_{k-1} distribution can be revisited as a $Gamma(\frac{k-1}{2}, 2)$ distribution, we will exploit the following two properties of the Gamma distribution (see e.g. Fisz 1963) to find the maximum likelihood estimator of the unknown parameter $\mathcal{U}_i$.

Property 1. If X is a random variable distributed as a $Gamma(a, b)$ and c is a constant, then cX is distributed as a $Gamma(a, cb)$.

Property 2. If $X_1, \ldots, X_m$ are independent random variables distributed as $X_t \sim Gamma(a_t, b)$, then their sum $X_1 + \cdots + X_m$ is distributed as a $Gamma(\sum_{t=1}^m a_t, b)$.

2.1 Maximum likelihood estimator of each scale parameter $\mathcal{U}_i$

Let us fix the starting state $i \in S$ from which our processes evolve. We apply Property 1 to the Pearson statistics $Q_i(t)$, by defining $\tilde{Q}_i(t) := \mathcal{U}_i Q_i(t)$, and thus obtaining that $(\tilde{Q}_i(1), \ldots, \tilde{Q}_i(m))$ is a sample of independent and identically distributed random variables, all distributed as $Gamma(\frac{k-1}{2}, 2\mathcal{U}_i)$. Their likelihood and log-likelihood functions are

$$L(\mathbf{q}, \mathcal{U}_i) = \prod_{t=1}^{m} \frac{1}{\Gamma(\frac{k-1}{2})(2\mathcal{U}_i)^{\frac{k-1}{2}}} q_t^{\frac{k-3}{2}} \exp(-\frac{q_t}{2\mathcal{U}_i}),$$

$$\log L(\mathbf{q}, \mathcal{U}_i) = -m \log(\Gamma(\frac{k-1}{2})) - m\frac{k-1}{2} \log(2\mathcal{U}_i) + \frac{k-3}{2} \sum_{t=1}^{m} \log q_t - \frac{1}{2\mathcal{U}_i} \sum_{t=1}^{m} q_t,$$

where $\mathbf{q} = [q_1, \ldots, q_m]$. By differentiating and equalizing to zero $\log L$ we find

$$\frac{d \log L}{d\mathcal{U}_i} = -\frac{m(k-1)}{2\mathcal{U}_i} + \frac{1}{2{\mathcal{U}_i}^2} \sum_{t=1}^{m} q_t = 0$$

from which we obtain the maximum likelihood estimator

$$\widehat{\mathcal{U}_i} = \frac{\sum_{t=1}^m \tilde{Q}_i(t)}{m(k-1)}. \qquad (2)$$

2.2 Distribution of the estimator and confidence interval

Since we know that $\tilde{Q}_i(t) \sim Gamma(\frac{k-1}{2}, 2\mathcal{U}_i), \forall t = 1, \ldots, m$ and they are independent, by Property 2 we have that

$$\sum_{t=1}^{m} \tilde{Q}_i(t) \sim Gamma(\frac{m(k-1)}{2}, 2\mathcal{U}_i)$$

and thus, by Property 1,

$$\widehat{\mathcal{U}_i} = \frac{1}{m(k-1)} \sum_{t=1}^{m} \tilde{Q}_i(t) \sim Gamma(\frac{m(k-1)}{2}, \frac{2\mathcal{U}_i}{m(k-1)}).$$

Then $E[\widehat{\mathcal{U}_i}] = \mathcal{U}_i$ and our estimator (2) is unbiased.

Let us now apply again Property 1 to the ratio $\frac{\widehat{\mathcal{U}_i}}{\mathcal{U}_i}$, which is then distributed as a $Gamma(\frac{m(k-1)}{2}, \frac{2}{m(k-1)})$, independent from $\mathcal{U}_i$, and which can thus be used as pivotal statistics to build confidence intervals or to test statistical hypotheses.

Let us find, as an example, a right confidence interval for $\mathcal{U}_i$ at level $1-\alpha$:

$$\begin{aligned} 1-\alpha &= P(\frac{\widehat{\mathcal{U}_i}}{\mathcal{U}_i} \le b) \\ &= P(\mathcal{U}_i \ge \frac{\widehat{\mathcal{U}_i}}{b}). \end{aligned}$$

Thus the interval $I(\mathcal{U}_i) = [\frac{\widehat{\mathcal{U}_i}}{b}, +\infty)$ is the desired confidence interval, where b is the critical value at level $1-\alpha$ of a $Gamma(\frac{m(k-1)}{2}, \frac{2}{m(k-1)})$. If the aggregation of units is significantly needed to filter out the small scale variability, then the interval $I(\mathcal{U}_i)$ should not include 1.

3 Results on a simulated example

In order to test the performance of our weighted χ^2 test, we simulated a sample of 30,000 i.i.d. copies of a Markov Chain X, with state space $S = \{1, 2, 3\}$, having transition matrix

$$P(t) = \begin{bmatrix} p_{11}(t) & p_{12}(t) & p_{13}(t) \\ p_{21}(t) & p_{22}(t) & p_{23}(t) \\ p_{31}(t) & p_{32}(t) & p_{33}(t) \end{bmatrix}.$$

The transition probabilities are varying in time, according to the equations reported in Table 1 and the graphs reported in Figure 1.

As can be seen from the graphs, for the first 10 time steps the probability to observe transitions to the state 1 is higher than the others, leading thus to a prevalence of the sample of our Markov chains being in state 1. Starting from time step 11 we observe a sudden increase in the probability to move to state 3, which then becomes the new most observed state in the sample of Markov chains.

All the Markov chains have been initialized at time $t = 0$ with a random state, equally distributed on $S = \{1, 2, 3\}$. We then applied our weighted χ^2 test

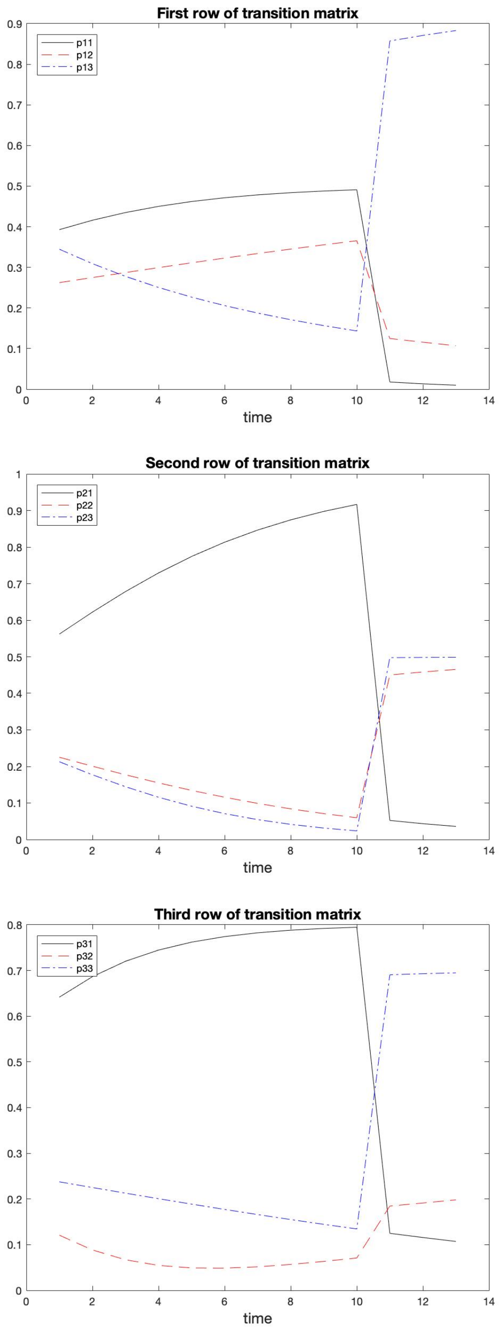

Fig. 1 Plots of the evolution of the transition probabilities in the matrix P during time

	$\mathbf{t = 1, \ldots, 10}$	$\mathbf{t = 11, \ldots, 13}$
$p_{11}(t) =$	$\frac{1}{2}\exp(1 + 0.3t)/(1 + \exp(1 + 0.3t))$	$\frac{1}{2}\exp(-0.3t)/(1 + \exp(-0.3t))$
$p_{12}(t) =$	$\frac{1}{2}\exp(0.1t)/(1 + \exp(0.1t))$	$\frac{1}{2}\exp(-0.1t)/(1 + \exp(-0.1t))$
$p_{13}(t) =$	$1 - p_{11}(t) - p_{12}(t)$	$1 - p_{11}(t) - p_{12}(t)$
$p_{21}(t) =$	$1 - p_{22}(t) - p_{23}(t)$	$1 - p_{22}(t) - p_{23}(t)$
$p_{22}(t) =$	$\frac{1}{2}\exp(-0.2t)/(1 + \exp(-0.2t))$	$\frac{1}{2}\exp(0.2t)/(1 + \exp(0.2t))$
$p_{23}(t) =$	$\frac{1}{2}\exp(-0.3t)/(1 + \exp(-0.3t))$	$\frac{1}{2}\exp(2 + 0.3t)/(1 + \exp(2 + 0.3t))$
$p_{31}(t) =$	$0.8\exp(1 + 0.4t)/(1 + \exp(1 + 0.4t))$	$\frac{1}{2}\exp(-0.1t)/(1 + \exp(-0.1t))$
$p_{32}(t) =$	$1 - p_{31}(t) - p_{33}(t)$	$1 - p_{31}(t) - p_{33}(t)$
$p_{33}(t) =$	$\frac{1}{2}\exp(-0.1t)/(1 + \exp(-0.1t))$	$0.7\exp(1 + 0.3t)/(1 + \exp(1 + 0.3t))$

Table 1 Evolution of the transition probabilities in the simulated Markov chains. The values have been fixed so that the rows of P sum to one.

Starting state	**1**	**2**	**3**
$\hat{\mathcal{U}}_i$	9.49	12.5	3.89
95% CI	$[6.04, +\infty)$	$[7.96, +\infty)$	$[2.47, +\infty)$

Table 2 Estimate of the new experimental units from the simulated data and corresponding right 95% confidence intervals

using the time steps from 1 to 10 to compute the maximum likelihood estimators $\hat{\mathcal{U}}_i, i \in S$, obtaining the results reported in Table 2. Note that all the computed right 95% confidence intervals of the $\hat{\mathcal{U}}_i$ are not including 1, showing thus that the units must be aggregated to filter out the small scale variations. We then applied the test, rescaling by $\hat{\mathcal{U}}_i, i \in S$ the countings, and testing the set of null hypotheses:

$$H_0^{i,t} : p_{ij}(t) = p_{ij}(t+1), \forall j \in S \tag{3}$$

for $i = 1, 2, 3$ and $t = 2, \ldots, 12$. The p-values of the tests are reported in Table 3. Using a significance level $\alpha = 0.05$ we observe a change in the distribution of the sample of Markov chains only starting from the comparison of the transitions $p_{ij}(10)$ with respect to $p_{ij}(11)$, as was expected. We also tested the same set of hypotheses (3) using a "traditional" χ^2 test, without any counting rescaling. The results are also reported in Table 3 and show that at all time steps before the major change there is a significant difference in the distribution at level $\alpha = 0.05$. This confirms that our rescaling is strictly needed in order to allow the χ^2 test to detect only the major changes in the distribution of the studied process.

4 Application to detect the effects of "greening" in the new CAP

In this section we will apply our weighted χ^2 test to study the impact of the new Common Agricultural Policy (CAP), which entered into force in the European Union from 2015, on the distribution of farmland uses in Lombardy, one of the most intensively cultivated regions in Italy. The new CAP has introduced a new funding policy, called greening, for which farm subsidies are conditioned to the compliance of farmers with some "agricultural practices beneficial for the climate and the environment" (Regulation EU 1307/2013), namely i) arable crops diversification, ii) maintenance of permanent grassland and iii) ecological focus areas (EFA). Such farm practices regard, and potentially influence, farmland allocation,

p-values of the weighted χ^2 test			
	Starting state		
	1	**2**	**3**
Transitions			
$t = 1 \to 2$ vs. $t = 2 \to 3$	0.055	0.185	0.0397
$t = 2 \to 3$ vs. $t = 3 \to 4$	0.269	0.131	0.151
$t = 3 \to 4$ vs. $t = 4 \to 5$	0.237	0.305	0.19
$t = 4 \to 5$ vs. $t = 5 \to 6$	0.321	0.417	0.4
$t = 5 \to 6$ vs. $t = 6 \to 7$	0.588	0.175	0.883
$t = 6 \to 7$ vs. $t = 7 \to 8$	0.240	0.621	0.239
$t = 7 \to 8$ vs. $t = 8 \to 9$	0.478	0.413	0.798
$t = 8 \to 9$ vs. $t = 9 \to 10$	0.587	0.326	0.588
$t = 9 \to 10$ vs. $t = 10 \to 11$	0	2.31e-196	5.63e-182
$t = 10 \to 11$ vs. $t = 11 \to 12$	0.799	0.833	0.451
$t = 11 \to 12$ vs. $t = 12 \to 13$	0.951	0.617	0.332
p-values of the not weighted χ^2 test			
	Starting state		
	1	**2**	**3**
Transitions			
$t = 1 \to 2$ vs. $t = 2 \to 3$	1.2e-12	7.02e-10	3.57e-06
$t = 2 \to 3$ vs. $t = 3 \to 4$	3.96e-06	9.30e-12	6.49e-04
$t = 3 \to 4$ vs. $t = 4 \to 5$	1.16e-06	3.60e-07	1.59e-03
$t = 4 \to 5$ vs. $t = 5 \to 6$	2.07e-05	1.78e-05	2.84e-02
$t = 5 \to 6$ vs. $t = 6 \to 7$	6.46e-03	3.6e-10	6.15e-01
$t = 6 \to 7$ vs. $t = 7 \to 8$	1.35e-06	2.62e-03	3.8e-03
$t = 7 \to 8$ vs. $t = 8 \to 9$	9.16e-04	1.56e-05	4.15e-01
$t = 8 \to 9$ vs. $t = 9 \to 10$	6.33e-03	8.36e-07	1.26e-01
$t = 9 \to 10$ vs. $t = 10 \to 11$	0	0	0
$t = 10 \to 11$ vs. $t = 11 \to 12$	1.18e-01	1.02e-01	4.54e-02
$t = 11 \to 12$ vs. $t = 12 \to 13$	6.21e-01	2.41e-03	1.37e-02

Table 3 p-values of the hypotheses (3) tested both with the weighted χ^2 test, with countings rescaled by $\hat{\mathcal{U}_i}, i \in S$, and with the not weighted χ^2 test, where countings have not been rescaled.

particularly that of arable land and grassland [Solazzo, Donati, and Arfini 2015; Cortignani, Severini, and Dono 2017]. Farmland evolution has already been modelled in literature using Markov Chains [Sang et al. 2011; Fu, Wang, and Yang 2018]. We will apply our weighted χ^2 test to detect if any major change in the farmland use of Lombardy occurred starting from 2015.

Our dataset is composed by a sample of 638,952 farmland parcels in Lombardy, occupying a total area of 743,072 hectares, whose land use has been observed in the years 2011-2016. The farmland uses have been aggregated into 23 classes, which are reported in Table 4, with the corresponding evolution of the extension of each cultivation across time.

We model the system as a sample of Markov chains, one for each cultivated hectar, which evolve across time in one of the different 23 cultivation classes. The assumption of markovianity, i.e. the fact that the land use of next year depends only on the present land destination and not from the past, is reasonable, since rotational crops are usually alternated on an annual or seasonal basis, and the majority of decisions of change of cultivation are based on current (and expected) prices of crops, annual environmental conditions and other factors, which are considered mainly on an annual basis.

Code	Farm land use	hectares 2011	hectares 2012	hectares 2013	hectares 2014	hectares 2015	hectares 2016
10	Maize	246,873	242,553	229,257	218,559	185,849	166,590
20	Maize for silage	57,484	60,801	67,278	73,305	66,004	68,045
30	Rotation ryegrass + maize for silage	31,728	34,025	34,907	37,141	42,943	48,003
40	Wheat	45,407	56,133	63,507	57,192	66,341	79,298
50	Barley	13,715	14,966	16,139	14,065	18,892	19,536
60	Triticale-other cereals	10,247	13,374	14,437	15,677	14,385	12,704
100	Rice	106,059	99,175	88,319	90,850	96,894	101,648
160	Soybean	22,160	15,680	24,574	27,253	39,290	32,325
190	Pulses	800	751	832	834	1,264	1,612
260	Horticulture	16,251	15,252	14,097	15,712	16,703	17,263
270	Flowers	4,231	4,158	4,081	3,952	3,892	3,902
320	Other arable crops	9,633	9,533	8,373	9,255	7,403	7,184
321	Ryegrass	1,514	787	760	898	4,024	5,315
322	Grass herbages	7,969	9,363	11,187	10,657	6,389	6,838
323	Legume herbages	238	172	116	142	1,214	909
325	Mixed herbages	4,458	3,928	4,892	4,629	3,674	4,435
330	Alfalfa	54,996	53,830	50,342	53,087	58,784	58,396
350	Other temporary grassland	49,632	49,575	49,833	49,910	48,261	47,373
360	Permanent grassland	8,831	8,979	8,797	8,700	8,871	8,941
414	Permanent crops	26,665	26,525	26,652	26,495	26,614	27,096
501	Wood-landscape	10,391	10,961	11,170	11,432	6,394	6,876
961	Fallow land	3,484	3,241	5,304	4,520	8,158	8,927
990	Non-eligible surfaces	10,292	9,295	8,205	8,797	10,818	9,845
	TOTAL BALANCED FARM LAND	743,072	743,072	743,072	743,072	743,072	743,072

Table 4 The 23 categories into which have been classified the different land uses in Lombardy and hectares occupied by each crop in different years. The considered sample of parcels has been selected so that the total farmland is constant across time. The code is an univocal (administrative) number which identifies the type of crop

Thus, coming back to the notations introduced in Section 2, in our application $k = 23$ is the number of different farmland uses, the times t span the years from 2011 to 2016, where the years from 2011 to 2014 are assumed to have a "standard behaviour", and will be used to estimate the experimental unit $\mathcal{U}_i$ which allows to filter out the small scale variation.

First we applied the Anderson-Goodman test for stationarity [Anderson and Goodman 1957], which is based on the maximum likelihood ratio, to the transition probabilities $p_{ij}(t)$, for t varying from 2011 to 2014, in order to check if they may be assumed constant in time, and thus if our sample of Markov chains can be assumed homogeneous, before the application of greening policy. The null hypothesis of homogeneity was rejected with a p-value < 0.0001. Thus we are exactly in a situation in which our proposed weighted χ^2 test should be applied.

We first considered the distribution of hectares into the 23 crop classes, and we applied our χ^2 test to check the hypotheses

H_0:the proportion of hectares in each of the 23 crop classes is the same in year $t-1$ and year t ($t = 2012, \ldots, 2016$).

The results are reported in Table 5. They show that a major discontinuity in the land use distribution has been observed exactly passing from 2014 to 2015, that is immediately after the introduction of greening policy. For this first test, we considered a unique $\mathcal{U} = \mathcal{U}_i$ for all i. The estimated rescaling factor here is $\hat{\mathcal{U}} = 225.47$.

We also applied the weighted χ^2 test to check if the transition probabilities of specific types of crops have changed significantly across time. This further set of tests was aimed to identify the main causes of the observed discontinuity on land use distribution on 2015, since we expected that some types of crops were not much affected by greening (like e.g. rice, wood-landscape, permanent grassland, etc.), since they already satisfied the ecological requirements of the new CAP,

	Q	DF	p-value	$freq < 5$	n(t-1)	n(t)
Comparison 2011-12	16.246	22	0.80350	0.087	3295.6	3295.6
Comparison 2012-13	21.901	22	0.46579	0.130	3295.6	3295.6
Comparison 2013-14	7.139	22	0.99881	0.130	3295.6	3295.6
Comparison 2014-15	64.714	22	0.00000	0.087	3295.6	3295.6
Comparison 2015-16	17.295	22	0.74685	0.043	3295.6	3295.6

Table 5 Comparison of the hectares distribution across couples of subsequent years. Here Q is the value of the rescaled χ^2 statistics, DF are its degrees of freedom, p-value is the p-value of the test, $freq < 5$ is the proportion of cells of the considered contingency table having expected rescaled frequencies lower than 5 (this is the heuristic condition that should be satisfied in order to apply asymptotic approximations with the χ^2 distribution), n(t-1) and n(t) are the number of total rescaled hectares in the two considered years

MAIZE						
Transitions	Q	DF	p-value	$freq < 5$	n(t-1)	n(t)
11-12/12-13	8.052	9	0.52887	0	569.9	558
12-13/13-14	9.948	9	0.35475	0	558	528.5
13-14/14-15	11.381	9	0.25048	0	528.5	503.9
13-14/15-16	14.718	9	0.09897	0	528.5	428.4
MAIZE FOR SILAGE						
Transitions	Q	DF	p-value	$freq < 5$	n(t-1)	n(t)
11-12/12-13	4.913	8	0.76687	0.111	630.9	669.5
12-13/13-14	11.087	8	0.1968	0.056	669.5	737.3
13-14/14-15	37.151	8	0.00001	0	737.3	789.6
13-14/15-16	31.102	8	0.00013	0	737.3	711.6
WHEAT						
Transitions	Q	DF	p-value	$freq < 5$	n(t-1)	n(t)
11-12/12-13	9.889	10	0.45031	0	738.7	911
12-13/13-14	10.111	10	0.43079	0	911	1020.1
13-14/14-15	722.043	10	0.01489	0	1020.1	914.6
13-14/15-16	42.91	10	0.00001	0	1020.1	1061.1
SOYBEAN						
Transitions	Q	DF	p-value	$freq < 5$	n(t-1)	n(t)
11-12/12-13	8.212	9	0.51292	0.25	433.2	298.9
12-13/13-14	9.788	9	0.36793	0.15	298.9	476.4
13-14/14-15	17.515	9	0.04124	0.05	476.4	528.1
13-14/15-16	32.822	9	0.00014	0	476.4	747.7
RICE						
Transitions	Q	DF	p-value	$freq < 5$	n(t-1)	n(t)
11-12/12-13	3.695	3	0.29628	0.250	325.6	303.2
12-13/13-14	7.669	3	0.05336	0.250	303.2	270.8
13-14/14-15	0.635	3	0.88834	0.250	270.8	278.5
13-14/15-16	2.751	3	0.43161	0.250	270.8	296.8

Table 6 Results of the weighted χ^2 test for specific crops to check if a significant change in the transitions before (2013/14) and after (2014/15 and 2015/2016) the greening introduction at level $\alpha = 0.05$ occurred. For the meaning of the columns see the caption to Figure 5

while other types of crops, like maize, maize for silage, wheat should show a bigger alternation with less pollutant crops, like soybean, starting from 2015, and thus should show a significant change in the transition probabilities. The results are reported in Table 6.

We observe that all results fit with our expectations, apart from maize, that doesn't show a significant change in the transition probabilities at level $\alpha = 0.05$.

MAIZE - CENTRAL LOMBARDY						
Transitions	Q	DF	p-value	$freq < 5$	n(t-1)	n(t)
11-12/12-13	5.536	8	0.69908	0.22	578.9	558.3
12-13/13-14	10.464	8	0.23394	0.11	558.3	524.2
13-14/14-15	21.421	8	0.00611	0	524.2	504.1
13-14/15-16	24.109	8	0.00219	0.05	524.2	431.1

Table 7 Results of the weighted χ^2 test on transition probabilities of maize in central Lombardy. For the meaning of the columns see the caption to Figure 5

After a deeper study, based on the computation of a *geographical Gini Index* [Bertoni et al. 2018], to highlight the local degree of crop differentiation, we realized that changes in the cultivation of maize happened, but mainly in the central part of Lombardy, which has a major livestock tradition, characterized by dairy farms based on on-farm feed production (particularly maize). Therefore, the weighted χ^2 test was applied only to the data located in the provinces of Bergamo, Brescia, Lodi, Cremona, representing the core of the livestock district. The results of the test are reported in Table 7. The small p-value in the comparisons 13/14-14/15 and 13/14-14/15 shows a significant change in the transition probabilities when greening was introduced, confirming that in this part of Lombardy a significant change in maize diffusion and in alternation with other crops occurred. For further comments and results see also Bertoni et al. 2018.

Acknowledgements This study has been supported by Fondazione Cariplo, within the research project "Evaluation of CAP 2015-2020 and taking action – CAPTION" (Project Id: 2017-2513). Funding has also been received from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie project BIGMATH, Grant Agreement No 812912

References

Anderson, T. W. and Leo A. Goodman (1957). "Statistical Inference about Markov Chains". In: *Ann. Math. Statist.* 28.1, pp. 89–110. DOI: 10.1214/aoms/1177707039. URL: https://doi.org/10.1214/aoms/1177707039.

Fisz, M. (1963). *Probability Theory and Mathematical Statistics.* New York, London, Sydney: John Wiley & Sons.

Knoke, David, George W. Bohrnstedt, and Alisa Potter Mee (2002). *Statistics for Social Data Analysis.* F.E.Peacock Publishers. ISBN: 0-87581-448-4.

Moreno, Elias, George Casella, and Antonio Garcia-Ferrer (2005). "An objective Bayesian analysis of the change point problem". In: *Stochastic Environmental Research and Risk Assessment* 19, 191–204. DOI: {10.1007/s00477-004-0224-2}.

De Gregorio, Alessandro and Stefano M. Iacus (2008). "Least Squares Volatility Change Point Estimation for Partially Observed Diffusion Processes". In: *Communications in Statistics - Theory and Methods* 37.15, pp. 2342–2357. DOI: 10.1080/03610920801919692. eprint: https://doi.org/10.1080/03610920801919692. URL: https://doi.org/10.1080/03610920801919692.

Ringstad Olsen, Lena, Probal Chaudhuri, and Fred Godtliebsen (2008). “Multi-scale spectral analysis for detecting short and long range change points in time series”. In: *Computational Statistics & Data Analysis* 52, pp. 3310–3330. DOI: 10.1016/j.csda.2007.10.027.

Sang, Lingling et al. (2011). “Simulation of land use spatial pattern of towns and villages based on CA-Markov model”. In: *Mathematical and Computer Modelling* 54.3. Mathematical and Computer Modeling in agriculture (CCTA 2010), pp. 938 –943. ISSN: 0895-7177. DOI: https://doi.org/10.1016/j.mcm.2010.11.019. URL: http://www.sciencedirect.com/science/article/pii/S0895717710005108.

Schütz, Nadine and Matthias Holschneider (2011). “Detection of trend changes in time series using Bayesian inference”. In: *Phys. Rev. E* 84 (2), p. 021120. DOI: 10.1103/PhysRevE.84.021120. URL: https://link.aps.org/doi/10.1103/PhysRevE.84.021120.

Iacus, Stefano M. and Nakahiro Yoshida (2012). “Estimation for the change point of volatility in a stochastic differential equation”. In: *Stochastic Processes and their Applications* 122.3, pp. 1068 –1092. ISSN: 0304-4149. DOI: https://doi.org/10.1016/j.spa.2011.11.005. URL: http://www.sciencedirect.com/science/article/pii/S0304414911002808.

Solazzo, Roberto, Michele Donati, and Filippo Arfini (2015). “Impact Assessment of Greening and the Issue of Nitrogen-Fixing Crops: Evidence from Northern Italy”. In: *Outlook on Agriculture* 44.3, pp. 215–222. DOI: 10.5367/oa.2015.0215. eprint: https://doi.org/10.5367/oa.2015.0215. URL: https://doi.org/10.5367/oa.2015.0215.

Cortignani, Raffaele, Simone Severini, and Gabriele Dono (2017). “Complying with greening practices in the new CAP direct payments: An application on Italian specialized arable farms”. In: *Land Use Policy* 61, pp. 265 –275. ISSN: 0264-8377. DOI: https://doi.org/10.1016/j.landusepol.2016.11.026. URL: http://www.sciencedirect.com/science/article/pii/S026483771631239X.

Eisinger, Robert D. and Yuguo Chen (2017). “Sampling for Conditional Inference on Contingency Tables”. In: *Journal of Computational and Graphical Statistics* 26.1, pp. 79–87. DOI: 10.1080/10618600.2016.1153478. eprint: https://doi.org/10.1080/10618600.2016.1153478. URL: https://doi.org/10.1080/10618600.2016.1153478.

Bertoni, Danilo et al. (2018). “Farmland Use Transitions After the CAP Greening: a Preliminary Analysis Using Markov Chains Approach”. In: *Land Use Policy* 79, pp. 789 –800. ISSN: 0264-8377. DOI: https://doi.org/10.1016/j.landusepol.2018.09.012. URL: http://www.sciencedirect.com/science/article/pii/S0264837718308676.

Fu, Xin, Xinhao Wang, and Y. Jeffrey Yang (2018). “Deriving suitability factors for CA-Markov land use simulation model based on local historical data”. In: *Journal of Environmental Management* 206, pp. 10 –19. ISSN: 0301-4797. DOI: https://doi.org/10.1016/j.jenvman.2017.10.012. URL: http://www.sciencedirect.com/science/article/pii/S0301479717309908.