20th of October 2022

Modelling the impact of Covid-19 pandemics on health insurance associated services demand

A. Fernández-Fontelo[1], Pedro Puig[1], M. Guillen[2], David Moriña[2]

[1]Universitàt Autònoma de Barcelona, [2]Universitat de Barcelona

Plan

Introduction

Misreporting

Misreporting in data refers to some event responsible for reporting less (under-reporting) or more (over-reporting) than the actual level of data.

- Many consequences can derive from misreporting: e.g., inferences that emerge from misreported data might be dramatically biased, providing an unrealistic picture of the actual problem.
- There are many real-world examples where each of the two forms of misreporting are well-documented, i.e., several cancers, STIs, gender-based violence or ADHD.

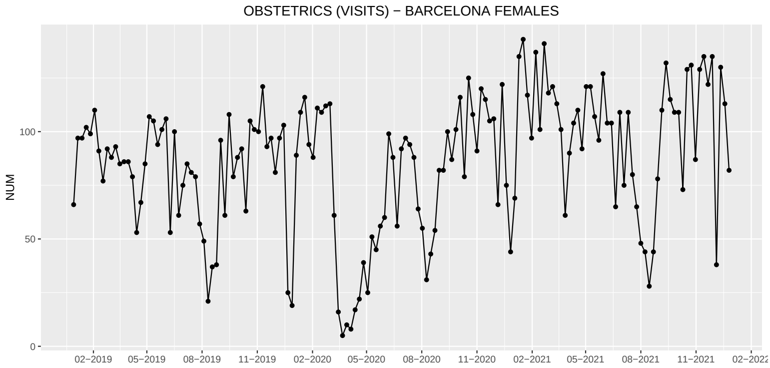
Private health insurance and Covid-19 pandemic

- More than 25% of the population has this type of coverage in Spain, exceeding 35% in some areas (UNESPA, 2020).
- The Covid-19 pandemic impacted the companies' claim rates in 2020 and 2021, especially for medical consultations and actions that were delayed for being (apparently) of low priority.
- **Hypothesis**: The number of company claims dramatically fell during the worst months of the Covid-19 pandemic, and many of the visits that should have been made at that time led to an increase in the number of visits in 2021 and 2022 (over-reporting).
  - $\Rightarrow$ Before Covid-19, we did not expect over-reporting but generally under-reporting for different health services.

Data

Data are the weekly number of visits to different health services of a given company from 2019 to 2021.

- **Geographic areas**: Spanish provinces.
- **Health services**: Urology, oncology, obstetrics, general medicine, osteopathy and cardiology.
- **Demographic variables**: Age and gender.


OBSTETRICS (VISITS) − BARCELONA FEMALES
NUM
100
50
0
02−2019
05−2019
08−2019
11−2019
02−2020
05−2020
08−2020
11−2020
02−2021
05−2021
08−2021
11−2021
02−2022

The model

Model I: Considering only under-reporting

Consider a **latent process** X_n with the following Poisson(λ)-INAR(1) structure:

$$X_n = \alpha \circ X_{n-1} + W_n(\lambda),$$

where $\alpha \in (0, 1)$. $\mathbb{E}(X_n) = \mathbb{V}(X_n) = \lambda/(1-\alpha) = \mu_X$. The operator $\circ$ is the binomial thinning such that: $\alpha \circ X_{n-1} = \sum_{i=1}^{X_{n-1}} Z_i$,, where Z_i are i.i.d Bernoulli(α).

Consider the following **observed and potentially under-reported process** Y_n: X_n with probability $1 - \omega$, or $q \circ X_n$ with probability ω.

Fernández-Fontelo, A., Cabaña, A., Puig, P. and Moriña, D. (2016). Under-reported data analysis with INAR-hidden Markov chains. Statistics in Medicine, 35(26): 4875-4890.

Model II: Considering both under-reporting and over-reporting

Still consider that the latent process X_n is a Poisson(λ)-INAR(1) model, but now the processes Y_n can be misreported:

$$Y_n = \begin{cases} X_n & 1-\omega, \\ \vartheta \Diamond X_n & \omega, \end{cases}$$

where $\vartheta = (\varphi_1, \varphi_2)$ and $\vartheta \Diamond X_n$ is called the fattening-thinning operator: $[\vartheta \Diamond X_n | X_n = x_n] = \sum_{k=1}^{x_n} W_k$:

$$\mathbb{P}(W = k | \varphi_1, \varphi_2) = \begin{cases} 1 - \varphi_1 - \varphi_2 & j = 0, \\ \varphi_1 & j = 1, \\ \varphi_2 & j = 2, \\ 0 & \text{otherwise.} \end{cases}$$

Under-reporting or over-reporting ?

- To distinguish between under-reporting, no misreporting or over-reporting, the following can easily be computed once the model parameters are estimated:

under-reporting	$\varphi_1 + 2\varphi_2 < 1$
no misreporting	$\varphi_1 + 2\varphi_2 = 1$
over-reporting	$\varphi_1 + 2\varphi_2 > 1$

- Note that when $\varphi_2 = 0$, the model results in the model I, which only accounts for under-reporting.
- A less flexible version of the operator defined before can be derived if $W_k \sim \mathrm{Binomial}(2, \varphi)$.

Parameter estimation: MoM method

The marginal distribution of the observed (and stationary) process Y_n is essential to compute the moment-based estimates of the model:

$$Y_n = \begin{cases} \text{Poisson}(\mu_X) & 1-\omega, \\ \text{Hermite}\left(\mu_X\varphi_1, \mu_X\varphi_2\right) & \omega. \end{cases}$$

1. Fit the mixture above to obtain estimates of $\widehat{\omega}$, $\widehat{\mu}_X$, $\widehat{\varphi}_1$ and $\widehat{\varphi}_2$.
2. Use the expression of the ACF, $\rho_Y(k) = c(\alpha, \lambda, \omega, \varphi_1, \varphi_2)\alpha^k$, to estimate α.
3. Using $\widehat{\mu}_X$ (step 1) and $\widehat{\alpha}$ (step 2), λ can be easily estimated.

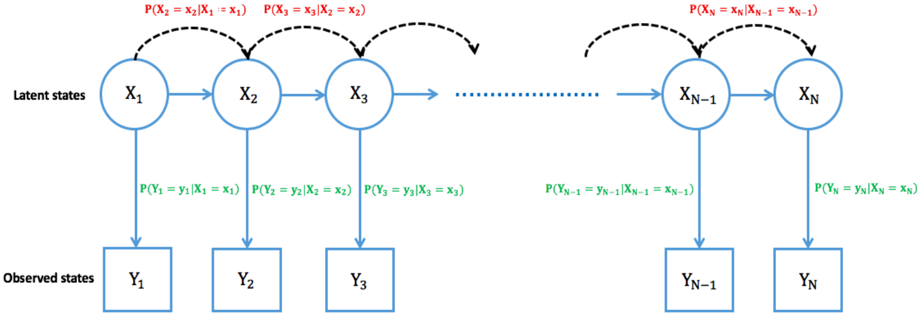
Parameter estimation: ML method

- Since the likelihood function (LF) of Y_n is not directly tractable (HMC with an infinite number of states), it is computed instead with the so-called **forward algorithm**.
- LF is then computed recursively using the forward probabilities given below:

$$\gamma_n(y_{1:n}, x_n) = \overbrace{\mathbb{P}(Y_n = y_n | X_n = x_n)}^{\text{emission probabilities}} \times \sum_{x_{n-1}=\frac{y_{n-1}}{2}}^{\infty} \underbrace{\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1})}_{\text{transition probabilities}} \gamma_{n-1}(y_{1:n-1}, x_{n-1}).$$

- Finally: $\mathbb{P}(Y_{1:N} = y_{1:N}) = \sum_{x_N=\frac{y_N}{2}}^{\infty} \gamma_N(y_{1:N}, x_N)$.

Visual model



Forward probabilities

- While the transition probabilities are computed using the cpmf of the Poisson(λ)-INAR(1), the emission probabilities are computed as below:

$$\mathbb{P}(Y_n = y_n | X_n = x_n) = \begin{cases} 0 & \text{if} \quad x_n < y_n/2, \\ (1-\omega) + \omega p_n & \text{if} \quad x_n = y_n, \\ \omega p_n & \text{if} \quad x_n > y_n, \\ \omega p_n & \text{if} \quad x_n < y_n,\, x_n \geq y_n/2. \end{cases}$$

- p_n can be computed using the following recursive relation:

$$p_n = \frac{1}{n(1-\varphi_1-\varphi_2)} \left[\varphi_1(x_n - (n-1))p_{n-1} + \varphi_2(2x_n - (n-2))p_{n-2} \right].$$

Preliminary results

Simulation study

- We perform a simulation study based on Monte Carlo providing different values for the parameters ω, φ_1 and φ_2.
- In particular, we generate an INAR(1)-Poisson process with $\alpha = 0.5$ and $\lambda = 3$, and different scenarios of over-reporting ($\omega = (0.3, 0.7)$, $\varphi_1 = (0.2, 0.3)$ and $\varphi_2 = (0.7, 0.5)$) and under-reporting ($\omega = (0.3, 0.7)$, $\varphi_1 = (0.2, 0.5)$ and $\varphi_2 = (0.2.0.1)$).
- We generate $M = 50$ repetitions of the processes X_n and Y_n above, all of length $n = 200$.

- **Over-reporting scenario**: $X_n = 0.5 \circ X_{n-1} + W_n(3)$ and:

$$Y_n = \begin{cases} X_n & 0.3, \\ (0.3, 0.5) \Diamond X_n & 0.7. \end{cases}$$

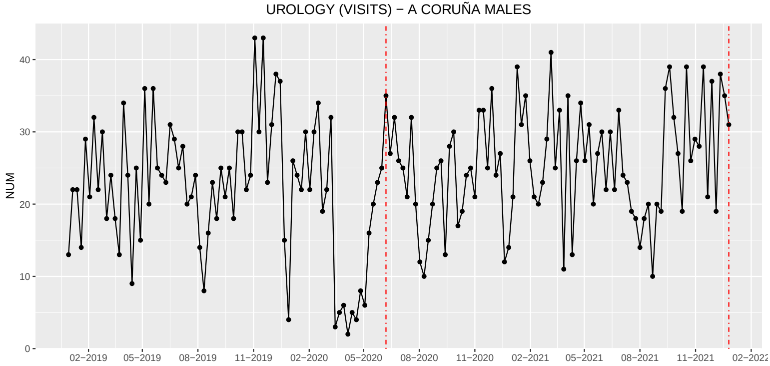
	α	λ	ω	φ_1	φ_2
true value	.5	3	.7	.3	.5
mean estimated value	.5172	3.0646	.6761	.2539	.4980
bias	.0172	.0646	-.0239	-.0461	-.0020
coverage (nominal: 0.90)	.9836	.9672	.9672	.6393	.9180

- **Under-reporting scenario**: $X_n = 0.5 \circ X_{n-1} + W_n(3)$ and:

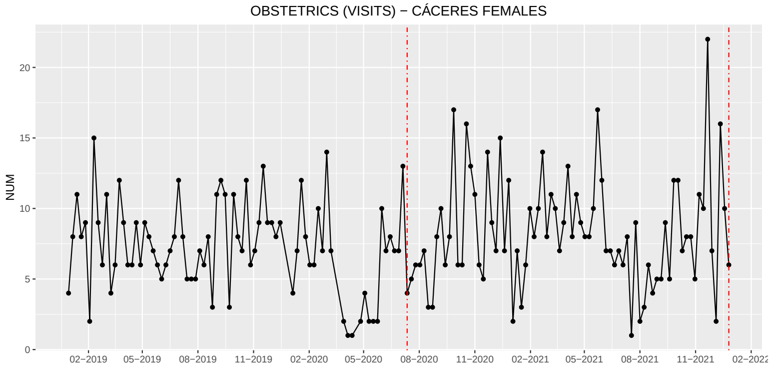
$$Y_n = \begin{cases} X_n & 0.3, \\ (0.2, 0.2)\Diamond X_n & 0.7. \end{cases}$$

	α	λ	ω	φ_1	φ_2
true value	.5	3	.7	.2	.2
mean estimated value	.4614	3.0228	.5953	.2010	.1892
bias	-.0386	.0228	-.1047	.0010	-.0108
coverage (nominal: 0.90)	.9121	.9890	.8901	.8571	.9121

Urology (Males) - A Coruña (06-20 to 12-21)



Obstetrics (Females) - Cáceres (07-20 to 12-21)



	ML point estimate	std. error
A Coruña (urology-males)		
α	.3256	.2029
λ	10.5644	3.1584
ω	.8809	.1020
φ_1	.1194	.0917
φ_2	.7988	.1467
Cáceres (obstetrics-females)		
α	.2232	.1576
λ	5.3185	1.2725
ω	.3897	.2446
φ_1	.1061	.0939
φ_2	.7105	.1596

- Note that in both cases $\hat{\varphi}_1 + 2\hat{\varphi}_2 > 1$, thus over-reporting.


Thank
You
Merci
Gracias
TapadhLeat
Köszönöm
Murakoze
aDank
BarakAllahuFiik
Grazzi
Grazie
Buznyg
WaadMahadsantahay
Takk
ThintKo
Mèsi
Chokrane
Kiitos
Blagodaram
Rahmat
Xièxie
Dziakuju
Mauruuru
Matondo
TerimaKasih
Aabhar
Barkal
GratiasAgimus
Mercé
Taiku
Tanemirt
Tenki
Ačiū
Dankie
Dakujem
Spacibo
Danke
Chnorakaloutioun
TänanVäga
Tak
Sagolun
Motashakkeram
Gràcies
Grandmercé
Hvala
Arigatô
Gratzias
GratiasAgo
Faleminderit
NajisTuke
KamSahHamnida
Gracie
KopKhunKha
Efharisto
CámOn
Nirringrazzjak
Bayarlalaa
Obrigado
Salamat
Saha
Multumesc
Dhanyavaad
Mamnun
aNiKie
KhobChaiDeu