

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

Amanda Fernández-Fontelo¹, Pedro Puig¹, Montserrat Guillén², David Moriña²

Abstract

Health insurance is one of the branches of insurance with the greatest penetration in the Spanish market; the same happens in many developed countries. Its accident rate has suffered the impact of the Covid-19 pandemic in 2020 and 2021, especially regarding consultations and medical events that could be postponed. Mobility restrictions led to a decline in the use of insurance by policyholders and a transformation of the interaction between patients and health workers with greater use of telephone consultation. This work aims to study how to determine if (i) due to the effect of postponing visits or (ii) due to the consequences of having suffered the virus (persistent Covid or side effects), there will be an excess of accidents and, where appropriate when it will occur. This work describes the methodological proposal that is aimed to respond to the research questions that will be addressed in this work.

Keywords

Introduction

In the last months, there has been a big global concern around the 2019 novel coronavirus (SARS-CoV-2) infection, prompting the World Health Organization (WHO) to declare a public health emergency in early 2020. The pandemic induced by this virus has significantly impacted many aspects of human activity. In addition to the direct consequences concerning deaths caused by the Covid-19 infection and the saturation of health systems in many countries (including Spain and neighbouring countries), a decrease in the use of health services has been detected in both the public health system and services associated with private health insurances in 2020.

With more than 12 million insured, health insurance is one of the insurance branches with the highest penetration in the Spanish market. More than 25% of the population has this type of coverage, exceeding 35% in some areas (UNESPA (2020)). In 2020 and 2021, the Covid-19 pandemic impacted the companies' claim rates, especially regarding medical consultations and actions that were apparently of low priority and thus were postponed. The mobility restrictions led to a decline in the use of insurance services, as well as a transformation of the interaction between patients and health workers, with a larger use of telephone consultations. The concern is whether there will be an excess of claims in 2022 and later years, either due to postponing visits or the pandemic consequences (e.g., because of persistent Covid symptoms or secondary effects). In the public health system, there is already evidence of an increased frequency of use of health services. However, it is not straightforward to determine if the highest frequency of claims that will be observed will be equal to or greater than the infra-loss rate that was observed during the pandemic period. For this purpose, we present here a new method for misreporting (i.e., under-reporting or over-reporting) estimation, which

builds on Fernández-Fontelo, A., Cabaña, A., Puig, P. and Moriña, D. (2016); Fernández-Fontelo, A., Cabaña, A., Joe, H., Puig, P. and Moriña, D. (2019). We use this method to determine: (i) whether the rebound effect occurs uniformly for all health insurance coverages or only for certain health insurance coverages, (ii) whether the rebound effect occurs homogeneously or varies depending on the insured characteristics, and (iii) to estimate the time point at which the initial benefit level is recovered. Many analyzes will be conducted on the effects of health spending at the public system level, but the implications for private health insurance will also be of interest. Above all, it is expected that to monitor the effects of the pandemic in the coming years, these forms of analysis –such as the one presented in the current work– will be used as population groups with different sociodemographic characteristics or impacts on the use of health services cannot be directly compared.

This work aims to estimate and evaluate the pandemic's impact on health insurance, estimating the degree of under-reporting of health insurance usage, mainly in 2020, as well as the degree of over-reporting of health insurance usage nowadays as a result of the pandemic consequences. In addition, we want to develop a system that monitors claims to detect changes in the dynamics of medical insurance use in particular and any other branch in general. Although the main focus of the current research is on the development of a new method for misreporting estimation, which has also been empirically tested using simulated data, this method can

¹ Departament de Matemàtiques, Universitat Autònoma de Barcelona

² Department of Econometrics, Statistics and Applied Economics, Riskcenter-IREA, Universitat de Barcelona

Corresponding author:

Amanda Fernández-Fontelo

Email: amanda@mat.uab.cat

also be applied to aggregated (anonymized) data from the health portfolio. We believe that the results and conclusions derived in the current research can be extended to other branches, as well as used to assess potential inequalities between countries or regions. In 2020, it was estimated that the total benefits provided by health insurance had totalled 6,300 million euros, of which 6,200 million correspond to the provision of medical services. In 2019, it was estimated that the total benefits provided by this type of insurance had totalled 6,600 million euros, of which 6,500 million correspond to the provision of medical services.

The current work is organized as follows. Section 2 presents the model, its properties, and a method for estimating the model's parameters. Section 3 shows the main results of a comprehensive simulation study to evaluate how the model behaves in practice in both the under-reporting and over-reporting scenarios.

The model

Let X_n be the following stationary INAR(1) process: $X_n = \alpha \circ X_{n-1} + Z_n$, where $\alpha \circ X_n$ is the so-called binomial thinning operator and $Z_n \sim \text{Poisson}(\lambda)$. Then, X_n follows a Poisson distribution with $\mathbb{E}(X_n) = \mu_X = \lambda/(1 - \alpha) = \sigma_X^2 = \mathbb{V}(X_n)$. Suppose that the processes X_{n-1} and Z_n are independent for all n . In addition, the auto-covariance and auto-correlation functions of this process are respectively $\gamma(k) = \text{Cov}(X_n, X_{n+k}) = \alpha^k \sigma_X^2$ and $\rho(k) = \text{Cor}(X_n, X_{n+k}) = \alpha^k$.

Let Y_n be an observed and potentially misreported (i.e., under-reported or over-reported) process such that:

$$Y_n = \begin{cases} X_n & 1 - \omega \\ \vartheta \diamond X_n & \omega, \end{cases} \quad (1)$$

where $\vartheta \diamond X_n$ is the so-called fattering-thinning operator defined as:

$$[\vartheta \diamond X_n | X_n = x_n] = \sum_{j=1}^{x_n} W_j, \quad (2)$$

and W_j are independent and identically distributed (i.i.d) random variables distributed following the probability mass function (pmf) below:

$$\mathbb{P}(W_j = k | \phi_1, \phi_2) = \begin{cases} 1 - \phi_1 - \phi_2 & \text{if } k = 0 \\ \phi_1 & \text{if } k = 1 \\ \phi_2 & \text{if } k = 2 \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

with $\vartheta = (\phi_1, \phi_2)$. Therefore, the observed process Y_n coincides with the true process X_n with probability $1 - \omega$; otherwise, the observed process Y_n is either less (under-reported) or more (over-reported) than the true process X_n with probability ω . The parameter $0 < \omega < 1$ is the frequency of the misreporting phenomenon, and parameters ϕ_1 and ϕ_2 indicate whether our process is under-reported or over-reported (i.e., if $\phi_1 + 2\phi_2 > 1$ the process is over-reported; if $\phi_1 + 2\phi_2 < 1$ the process is under-reported). Note that if $\phi_1 = 1$, the process is not misreported. In

addition, a less flexible version of the operator defined in expression (2) can be derived if $W_j \sim \text{Binomial}(2, \phi)$. Although the distribution in expression (3) is the easiest way to model over-reporting in count time series, other probability distributions with no compact support (e.g., the Poisson or Geometric distributions) can be taken into account here as well.

The operator defined in expression (2) follows a 2nd-order Hermite distribution with parameters $\mu_X \phi_1$ and $\mu_X \phi_2$, where $\mu_X = \lambda/(1 - \alpha)$ is the expectation of X_n and $\phi_j = \mathbb{P}(W = j)$ for $j = 1, 2$. We can see the above using the idea of probability generating functions (pgf) as follows: First, the marginal distribution of the observed process $X_n \sim \text{Poisson}(\mu_X)$ with pgf $G_X(s) = e^{\mu_X(s-1)}$. Second, the pgf of the random variable W in expression (3) is given by $G_W(s) = (1 - \phi_1 - \phi_2) + \phi_1 s + \phi_2 s^2$. Finally, the pgf of the operation defined in expression (2) is given by $G_{\vartheta \diamond X_n}(s) = G_X(G_W(s)) = \exp(\mu_X \phi_1(s-1) + \mu_X \phi_2(s^2-1))$, which is the pgf of a 2nd-order Hermite distribution with parameters $\mu_X \phi_1$ and $\mu_X \phi_2$. Hence, the expectation and variance of the random variable $\vartheta \diamond X_n$ are $\mathbb{E}(\vartheta \diamond X_n) = \mu_X(\phi_1 + 2\phi_2)$ and $\mathbb{V}(\vartheta \diamond X_n) = \mu_X(\phi_1 + 4\phi_2)$, respectively.

We already know that the latent process is distributed as $X_n \sim \text{Poisson}(\lambda/(1 - \alpha))$ and the fattening-thinning operator follows $\vartheta \diamond X_n \sim \text{Hermite}(\phi_1 \mu_X, \phi_2 \mu_X)$, where $\mu_X = \lambda/(1 - \alpha)$. In addition, we can see that the distribution of the observed process Y_n is the following mixture of a Poisson distribution and a 2nd-order Hermite distribution:

$$Y_n = \begin{cases} \text{Poisson}(\mu_X) & 1 - \omega, \\ \text{Hermite}(\mu_X \phi_1, \mu_X \phi_2) & \omega, \end{cases} \quad (4)$$

with expectation and variance defined as $\mathbb{E}(Y_n) = \mu_X(1 - \omega(1 - \phi_1 - 2\phi_2))$ and $\mathbb{V}(Y_n) = \mu_X(1 - \omega(1 - \phi_1 - 4\phi_2)) + \mu_X^2 \omega(1 - \omega)(1 - \phi_1 - 2\phi_2)^2$.

Finally, the auto-correlation function of the observed process Y_n is defined as:

$$\gamma(k) = \text{Cov}(Y_n, Y_{n+k}) = \mu_X \alpha^k (1 - \omega(1 - \phi_1 - 2\phi_2))^2, \quad (5)$$

which can be derived computing $\mathbb{E}(Y_n, Y_{n+k})$, $\mathbb{E}(Y_n)$ and $\mathbb{E}(Y_{n+k})$. In fact, we already know $\mathbb{E}(Y_n) = \mu_X(1 - \omega(1 - \phi_1 - 2\phi_2))$ (and similarly for $\mathbb{E}(Y_{n+k})$ given that Y_n is a stationary process). In addition, $\mathbb{E}(Y_n, Y_{n+k})$ can be computed as a function of $\mathbb{E}(X_n, X_{n+k}) = \text{Cov}(X_n, X_{n+k}) + \mathbb{E}(X_n)\mathbb{E}(X_{n+k}) = \alpha^k \sigma_X^2 + \mu_X^2$. The auto-correlation function can be thus computed as:

$$\begin{aligned} \rho(k) = \text{Cor}(Y_n, Y_{n+k}) &= \frac{\text{Cov}(Y_n, Y_{n+k})}{\sqrt{\mathbb{V}(Y_n)}\sqrt{\mathbb{V}(Y_{n+k})}} = \\ &= \frac{\alpha^k (1 - \omega(1 - \phi_1 - 2\phi_2))^2}{(1 - \omega(1 - \phi_1 - 4\phi_2)) + \mu_X \omega(1 - \omega)(1 - \phi_1 - 2\phi_2)^2} = \\ &= \alpha^k c(\alpha, \lambda, \omega, \phi_1, \phi_2). \end{aligned} \quad (6)$$

Note that $c(\alpha, \lambda, \omega, \phi_1, \phi_2)$ is constant with respect to k , and hence the ACF of the observed process Y_n is proportional (up to a multiplicative constant) to that of the latent process X_n (i.e., the ACF is geometrically decreasing at rate k given that $0 < \alpha < 1$). Note also that we always take by definition that $\rho(0) = 1$.

Model inferences

Parameters of the model (i.e., $\alpha, \lambda, \omega, \phi_1$ and ϕ_2) are estimated using the likelihood function, but this is not directly tractable as described in [Fernández-Fontelo, A., Cabaña, A., Puig, P. and Moriña, D. \(2016\)](#); [Fernández-Fontelo, A., Cabaña, A., Joe, H., Puig, P. and Moriña, D. \(2019\)](#). According to these authors, we can compute the likelihood function here using the forward algorithm, which is based on the so-called transition and emission probabilities. In particular, the transition probabilities are given by the conditional pmf of the stationary INAR(1) process defined as:

$$\begin{aligned} \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) &= \\ &= \sum_{j=0}^{\min(x_n, x_{n-1})} \binom{x_{n-1}}{j} \alpha^j (1-\alpha)^{x_{n-1}-j} \mathbb{P}(Z_n = x_n - j), \end{aligned} \quad (7)$$

where $\mathbb{P}(Z_n = x_n - j)$ is computed in the following with the pmf of a $\text{Poisson}(\lambda)$ distribution. In addition, the emission probabilities here are defined as given:

$$\begin{aligned} \mathbb{P}(Y_n = y_n | X_n = x_n) &= \\ &= \begin{cases} 0 & \text{if } x_n < y_n/2 \\ (1-\omega) + \omega p_n & \text{if } x_n = y_n \\ \omega p_n & \text{if } x_n > y_n \\ \omega p_n & \text{if } x_n < y_n, x_n \geq y_n/2, \end{cases} \end{aligned} \quad (8)$$

and probabilities p_n are computed using the following recursive relation:

$$\begin{aligned} p_n &= \frac{1}{n(1-\phi_1-\phi_2)} [\phi_1(x_n - (n-1))p_{n-1} + \\ &+ \phi_2(2x_n - (n-2))p_{n-2}], \end{aligned} \quad (9)$$

which was computed following [Baena-Mirabete and Puig \(2020\)](#). Note that we consider under-reporting when $x_n > y_n$ and over-reporting when $x_n < y_n$, but $x_n \geq y_n/2$. Therefore, if there is over-reporting, our model assumes that what we see (i.e., y_n) is at most twice what happens (i.e., x_n).

Finally, with the transition and emission probabilities in expressions (7) and (8), the likelihood function of the observed process, Y_n , is recursively computed using the forward algorithm, which is essentially based on the forward probabilities. That is, the likelihood function is thus calculated as:

$$\mathbb{P}(Y_{1:N} = y_{1:n}) = \sum_{x_N = \frac{y_N}{2}}^{\infty} \gamma_N(y_{1:N}, x_N), \quad (10)$$

where the forward probabilities $\gamma_n(y_{1:N}, x_N)$ are defined as:

$$\begin{aligned} \gamma_n(y_{1:n}, x_n) &= \mathbb{P}(Y_n = y_n | X_n = x_n) \times \\ &\sum_{x_{n-1} = \frac{y_{n-1}}{2}}^{\infty} \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \gamma_{n-1}(y_{1:n-1}, x_{n-1}), \end{aligned} \quad (11)$$

considering $\mathbb{P}(X_1 = x_1) \sim \text{Poisson}(\lambda/(1-\alpha))$. Note that expression (11) is based on the transition probabilities and emission probabilities, that in our model here are defined as given in expressions (7) and (8), respectively. In practice, we use numerical optimization procedures, e.g., using the `nlm` function in R.

Results

Simulation study

Table 1. Performance of the proposed model on the under-reporting scenarios.

Scenario	Measure	α	λ	ω	ϕ_1	ϕ_2
1	True value	0.5	3	0.7	0.2	0.3
	Mean est. value	0.49	2.95	0.61	0.20	0.29
	Bias	-0.009	-0.05	-0.09	0.002	-0.006
	Coverage	99.6	99.6	92.4	88.8	94.0
2	True value	0.5	3	0.3	0.2	0.3
	Mean est. value	0.50	2.96	0.28	0.20	0.28
	Bias	0.0008	-0.04	-0.02	0.001	-0.02
	Coverage	99.6	99.8	88.8	85.6	92.0
3	True value	0.5	3	0.7	0.1	0.1
	Mean est. value	0.43	3.41	0.68	0.10	0.09
	Bias	-0.07	0.41	-0.01	0.0006	-0.006
	Coverage	96.2	96.8	90.8	93.0	97.2
4	True value	0.5	3	0.3	0.1	0.1
	Mean est. value	0.49	3.03	0.31	0.10	0.09
	Bias	-0.008	0.03	0.005	0.002	-0.006
	Coverage	97.0	97.0	95.6	93.2	95.4

Under-reporting scenarios

Table 2. Performance of the proposed model on the over-reporting scenarios.

Scenario	Measure	α	λ	ω	ϕ_1	ϕ_2
1	True value	0.5	3	0.7	0.2	0.7
	Mean est. value	0.51	2.99	0.68	0.19	0.70
	Bias	0.011	-0.015	-0.019	-0.013	0.003
	Coverage	99.2	98.8	95.6	80.6	91.8
2	True value	0.5	3	0.3	0.2	0.7
	Mean est. value	0.51	2.92	0.34	0.18	0.70
	Bias	0.010	-0.076	0.042	-0.023	0.003
	Coverage	98.8	98.0	94.4	70.4	91.4
3	True value	0.5	3	0.7	0.3	0.5
	Mean est. value	0.511	2.99	0.66	0.27	0.52
	Bias	0.011	-0.013	-0.041	-0.030	0.018
	Coverage	100.0	99.6	96.6	78.0	92.8
4	True value	0.5	3	0.3	0.3	0.5
	Mean est. value	0.52	2.90	0.37	0.289	0.486
	Bias	0.021	-0.097	0.073	-0.011	-0.014
	Coverage	99.8	99.0	96.2	69.8	89.4

Over-reporting scenarios

Table 3. Estimates corresponding to the full lockdown period.

Parameter	Estimate (95% CI)
α	0.85 (0.56; 1.14)
λ	2.81 (-2.33; 7.95)
ω	0.80 (0.55; 1.06)
ϕ_1	0.025 (-0.015; 0.065)
ϕ_2	0.25 (0.13; 0.36)

Table 4. Estimates corresponding to the post-pandemic period.

Parameter	Estimate (95% CI)
α	0 (-0.31; 0.31)
λ	19.99 (13.92; 26.06)
ω	0.90 (0.74; 1.07)
ϕ_1	0.19 (-0.21; 0.58)
ϕ_2	0.61 (0.23; 0.99)

Application: Number of visits to obstetrics service in Tarragona province

Discussion

References

Acknowledgements

This research was funded by Fundación MAPFRE (Becas Ignacio H. de Larramendi 2021). A.F-F acknowledges Agencia Estatal de Investigación for the financial support IJC2020-045188I/AEI/10.13039/501100011033 and María Zambrano scholarship.

References

- Baena-Mirabete S and Puig P (2020) Computing probabilities of integer-valued random variables by recurrence relations. *Statistics & Probability Letters* 161: 108719. DOI:10.1016/j.spl.2020.108719. URL <https://www.sciencedirect.com/science/article/pii/S0167715220300225>.
- Fernández-Fontelo, A, Cabaña, A, Puig, P and Moríña, D (2016) Under-reported data analysis with INAR-hidden Markov chains. *Statistics in Medicine* 35(26): 4875–4890. DOI: 10.1002/sim.7026.
- Fernández-Fontelo, A, Cabaña, A, Joe, H, Puig, P and Moríña, D (2019) Untangling serially dependent underreported count data for gender-based violence. *Statistics in Medicine* 38(22): 4404–4422. DOI:10.1002/sim.8306.
- UNESPA (2020) Informe Estamos Seguros : 290–290URL [https://unespa-web.s3.amazonaws.com/main-files/uploads/2019/05/Informe+Estamos+Seguros+2018+\(Version+integrada\).pdf%0Ahttps://unespa-web.s3.amazonaws.com/main-files/uploads/2018/06/Informe-2017-Estamos-Seguros_V05.pdf%0Ahttp://www.unespa.es/que-hacemos/publ](https://unespa-web.s3.amazonaws.com/main-files/uploads/2019/05/Informe+Estamos+Seguros+2018+(Version+integrada).pdf%0Ahttps://unespa-web.s3.amazonaws.com/main-files/uploads/2018/06/Informe-2017-Estamos-Seguros_V05.pdf%0Ahttp://www.unespa.es/que-hacemos/publ).