Data-driven health insurance optimal contract specification: comparing insurance policies in a risk-averse context

**Introduction**

Insurance aims to provide financial support to atypically high losses financed by agents that are part of the insurance pool. According to Chhibber et al. (1997), insurance implies market distortions such as information asymmetries that might cause losses in efficiency and harm the desirable risk-spreading that benefits society by improving welfare through covering high agent losses (Cutler and Zeckhauser, 1998). The industry of health insurance is of particular concern to policymakers because of the risk of catastrophic health expenditures, that is, emergency expenditures in health that can significantly impact the economy of households. Catastrophic expenses typically make a family poor if it does not have enough economic resources to face significant losses.

The US health insurance industry is characterized by expensive spending, almost three times the OECD average, that are twice the OECD average in volume, showing evidence of the high cost of healthcare services (OECD, 2021). In this context, the expenditure triggered by an illness or an accident could be catastrophic for the healthcare consumer if the government does not subsidize healthcare. In such a costly market, the main challenge is to balance the premium magnitude charged by the underwriter with the likelihood of the event occurrence and the respective financial recovery due to the policyholder (Torraca and Fanzeres, 2021).

In health insurance, the insurers or underwriters face information asymmetries regarding the health status of the insured or policyholders. For instance, if policyholders are smokers, they are more likely to trigger charges due to respiratory diseases. If underwriters ignore this fact, they will design a contract that may not cover enough of the losses. Cutler and Zeckhauser (1998) showed that if underwriters create two arrangements: [1] moderate, which is cheap and provides low coverage, and [2] generous, which is expensive and offers high coverage, then high-risk individuals may prefer the most affordable contract, even if it does not cover all the medical costs when an illness or an accident is triggered. In all the cases above, the desirable risk-spreading is affected. This phenomenon is called adverse selection.

In the case of health insurance, insurers need more information regarding the risk drivers of policyholders' to design contracts. We define insurance supply as the coverage that insurers are willing to offer for a given premium risk. Along the same line, supply is a function of marginal costs that seem to be affected by information asymmetries. We argue that the supply curve can shift to the right if information asymmetries are reduced. Thus, the main objective of this paper is to investigate the optimal health insurance contract specification under uncertain demand and policyholders segmentation.

The proposed methodology uses a holistic approach to evaluate the optimal parameter specification. Following Torraca and Fanzeres (2021), a risk-averse optimization problem is formulated to specify the optimal magnitude of the contract's parameters that maximize a risk-adjusted value of the companies' final wealth. A functional form built on quantile-based Conditional Value-at-Risk is adapted to quantify the insurers' absolute wealth value (Fanzeres, Street, and Barroso, 2014; Street, 2010). This formulation uses the standard safety-loading-based premium pricing framework (Bernard et al., 2015; Cummins and Mahul, 2004; Sun, Weng, and Zhang, 2017). However, a simulation approach named Sample Average Approximation (SAA) must be used to approximate the optimization problem, as endogenous variables depend on stochastic parameters such as health charges.

The uncertainty regarding health charges was characterized in the literature by probabilistic distributions and then populated with Monte Carlo simulations (Arenas et al., 2017). The authors acknowledge that uncertainty must be adequately represented if not directly representing actual outcomes, so a dataset from US health insurance charges is used to perform empirical simulations to solve the stochastic problem above. This approach overcomes the limitations of distributional assumptions regarding uncertainty, but it comes with increasing computational complexity, as is common in empirical simulations. We use the variance-reduction Quasi-Monte Carlo Latin Hyper Cube Sampling technique to tackle this limitation.

This paper proposes solving the optimization problem considering two big scenarios: [1] one global insurance contract for the entire population and [2] different contracts for subpopulations derived from the market segmentation produced by insurers' features using clustering methodology. These experiments may shed light on the optimal contract specification under different settings. In both the single and multiple contracts approaches, sensibility analysis on the probability of illness or accidents, level of companies' risk-aversion, and risk-loading parameter specification was conducted to investigate pricing behavior under different specifications. The insights drawn from these experiments would contribute to academics improving their understanding of health insurance risk management and policymakers to have an experimental framework to formulate and test policies to improve social welfare.

The remaining of this paper is divided into five sections. Section 2 describes the leading work on methodologies to investigate the response of health insurance companies to uncertainty. Section 3 main ideas of the theoretical framework and drawbacks of the mathematical modeling approach. Section 4 details the data collection methods and processing pipeline, the experimental setting is also defined in this section. In Section 5, the main results are highlighted and discussed under the scope of economic theory. Finally, in the conclusions section, the paper is closed with recommendations for improvement in risk management based on the use of optimization under uncertainty techniques.