Embedded predictive analysis of misrepresentation risk in insurance ratemaking

Michelle Xia

Division of Statistics

Northern Illinois University

August 6, 2016
2016 CARE Conference

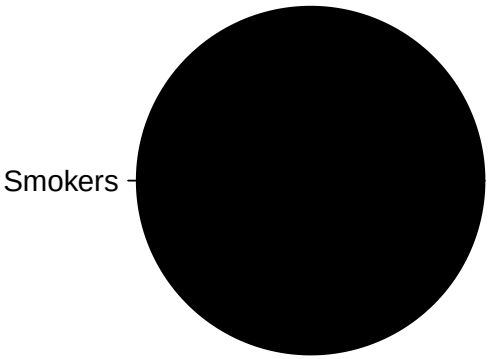
Funded by the Casualty Actuarial Society, joint work with Paul Gustafson (UBC)

Motivation

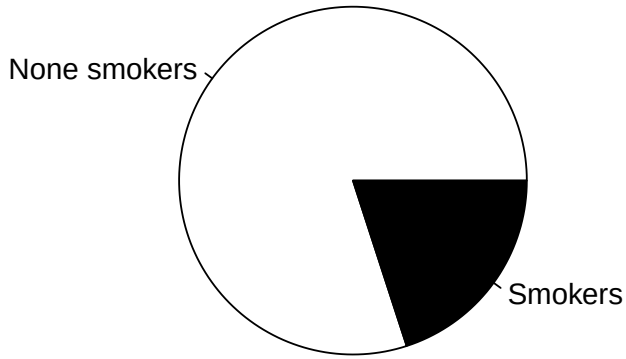
- **Misrepresentation** (see, e.g., Winsor [1995]) is a type of insurance **fraud** when the applicant choose to give a false statement on a trait (e.g., smoking status and age in health insurance) that may affect the insurance premium.
- In practice, insurance companies usually do not verify information provided by the applicant.
- Due to the financial incentive, misrepresentation happens frequently.

Misrepresentation is **unidirectional** and usually **unobserved**.

Misrepresentation on smoking status



(a) Report smoking



(b) Report nonsmoking

Figure: Here, we usually do **not observe** the true status, hence **cannot directly learn** the percentage of misrepresentation.

Ratemaking and misrepresentation

- In insurance ratemaking, actuaries determine health insurance rates based on regression models between **historical losses** and **risk factors** such as age, location and smoking status.
- In a traditional ratemaking model, misrepresentation will result in an **underestimation** of the risk/association.
- Misrepresentation is usually **unobserved**, with a typical **selection bias** for the confirmed cases. Hence, standard models may be **nonidentifiable** (e.g., for misrepresentation percentage and risk effect).

Ratemaking data structure

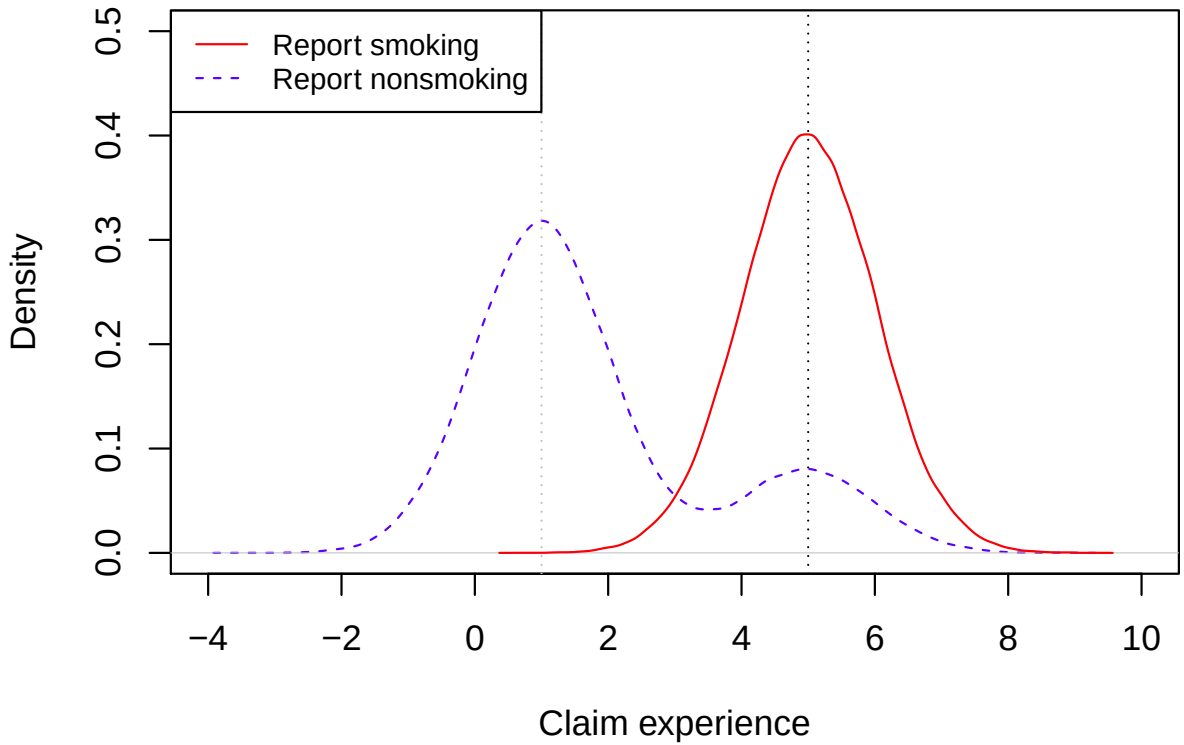


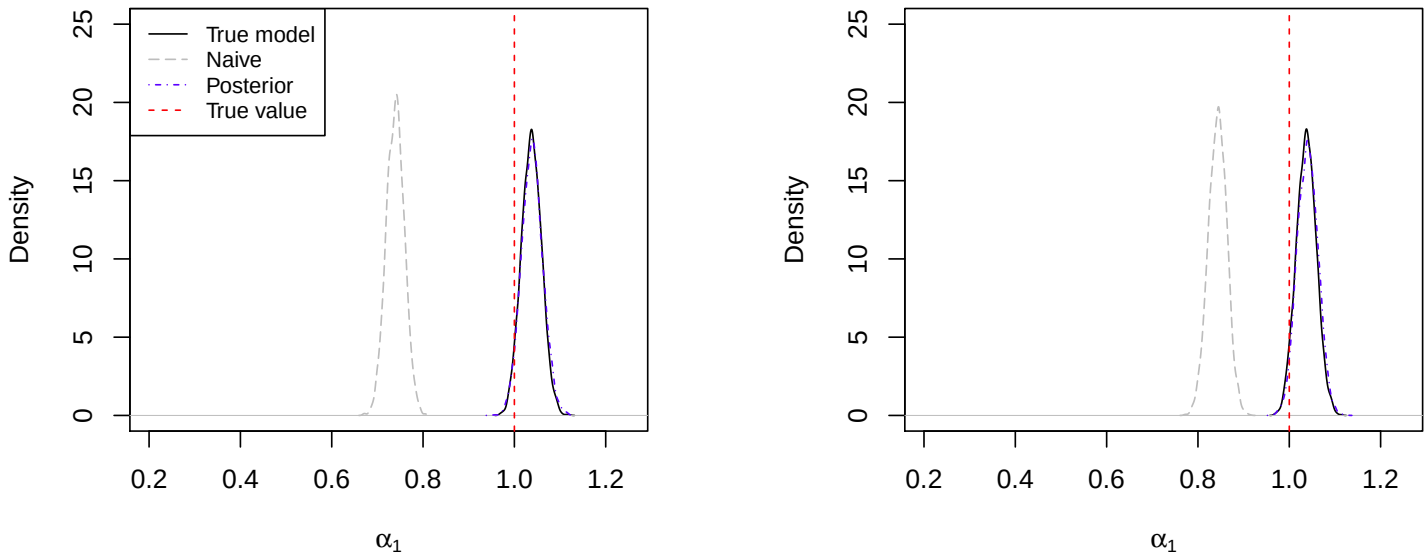
Figure: Loss experience by **reported** smoking status under **ratemaking** models, when comparing inviduals with same **other risk characteristics**.

Health insurance model and assumptions

- For health insurance, we specify a **regression** structure that characterizes the relationship between **medical losses** and **true** risk profiles such as age, location and smoking status.
- We assume there is a latent mechanism on the misrepresentation of **smoking** status, and we know the **direction** of error.
- In addition, we can specify an embedded predictive model that associate the chance of misrepresentation to the **age** variable.

In more complicated cases, the **risk factors can be selected or tested**, like in the case of regular regression analysis.

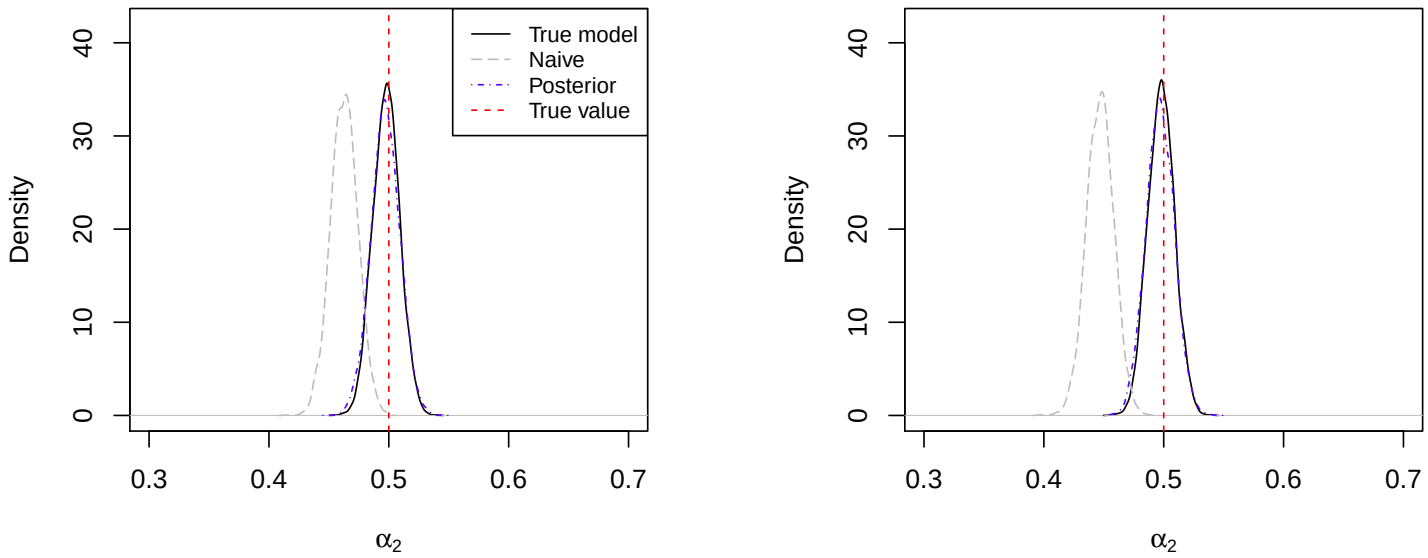
Smoking effect on loss frequency



(a) Low association (b) High association

Figure: Posterior distributions for **smoking effect** α_1 on Poisson losses ($n = 1000$).

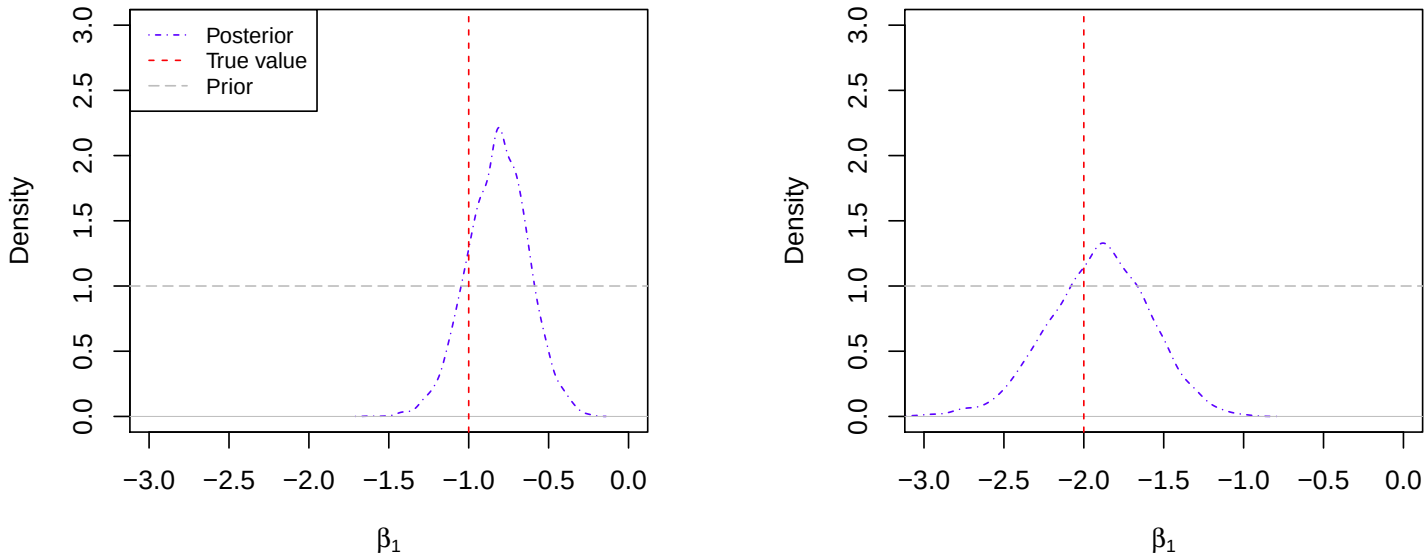
Age effect on loss frequency



(a) Low association (b) High association

Figure: Posterior distributions for **age effect** α_2 on Poisson losses ($n = 1000$).

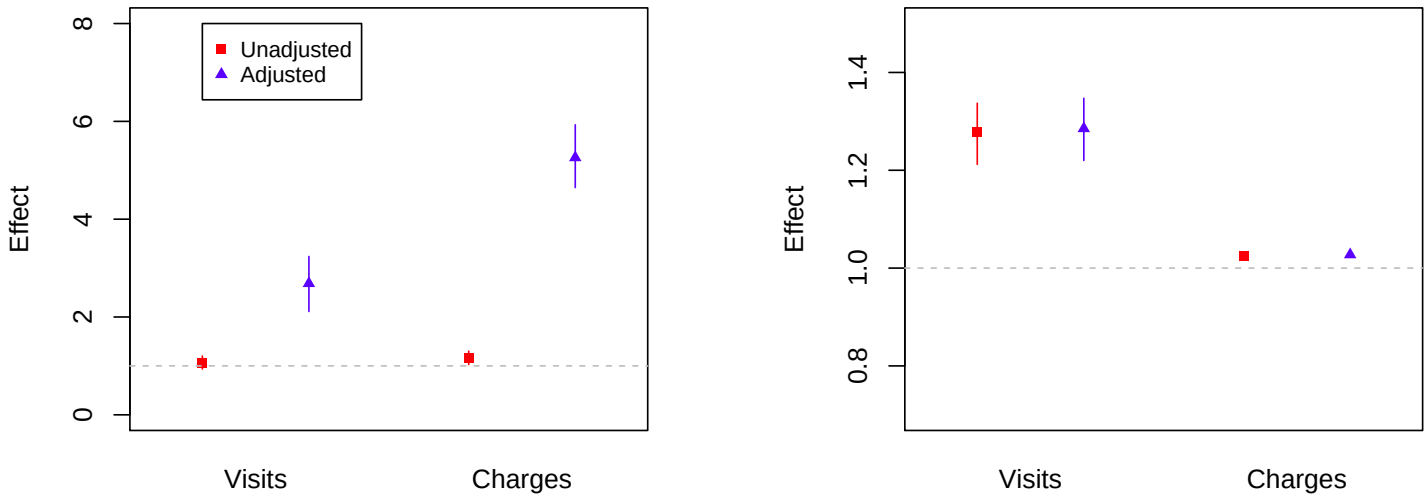
Age effect on misrepresentation risk



(a) Low association (b) High association

Figure: Posterior distributions for age effect β_1 on chance of misrepresentation ($n = 1000$).

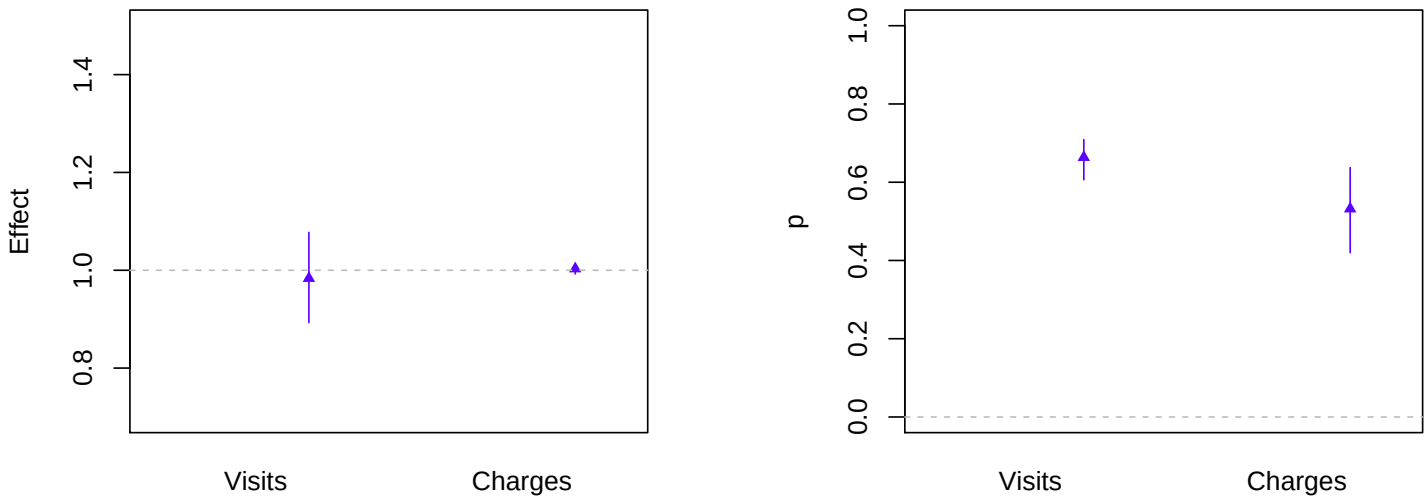
Healthcare expense risk factors



(a) Smoking (b) Age

Figure: Credible intervals for the effect of smoking and age, for the office-based visits and total medical charges.

Misrepresentation risk factor



(a) Age (b) p

Figure: Credible intervals for age effect on odds of misrepresentation, and the estimated misrepresentation probability p for individuals at the average age.

Application in fraud modeling

Fraud is **unidirectional** and usually **unobserved**.

- The methods may be constructed based on a regression model between published **revenues/expenses** and corporate fraudulent activities (e.g., accounting misconduct studied by Hahn et. al. [2016]), after adjusting for other corporate characteristics.
- Using **predictive models on misrepresentation risk**, we may be able to identify **risk factors** that are associated with corporate frauds, and hence have a **scoring system**.

More generally, even for fraud activities that are confirmed, there is a **selection bias**, since the characteristics of the unidentified frauds may be different. So there is a need for statistical models that adjust for the bias.