Comparison of modeling approaches for healthcare beneficiary costs

Daniel Shapiro

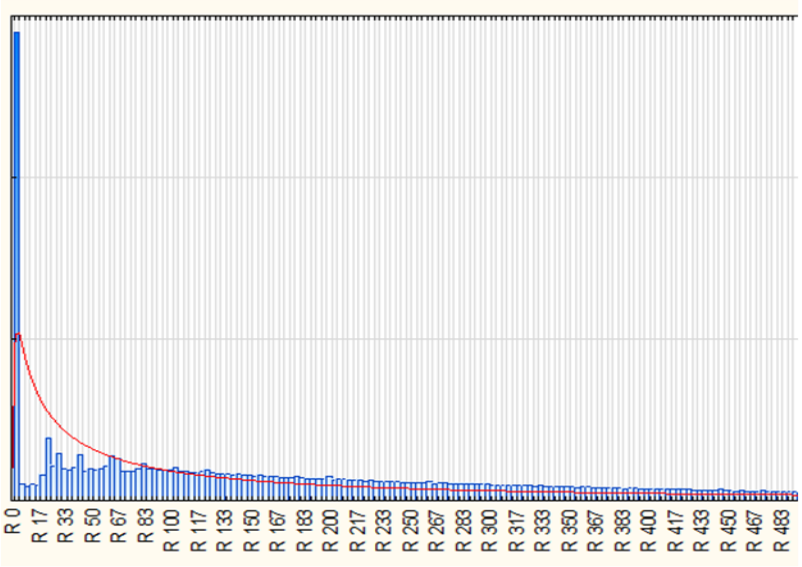
CareGauge
Health Systems Measurement

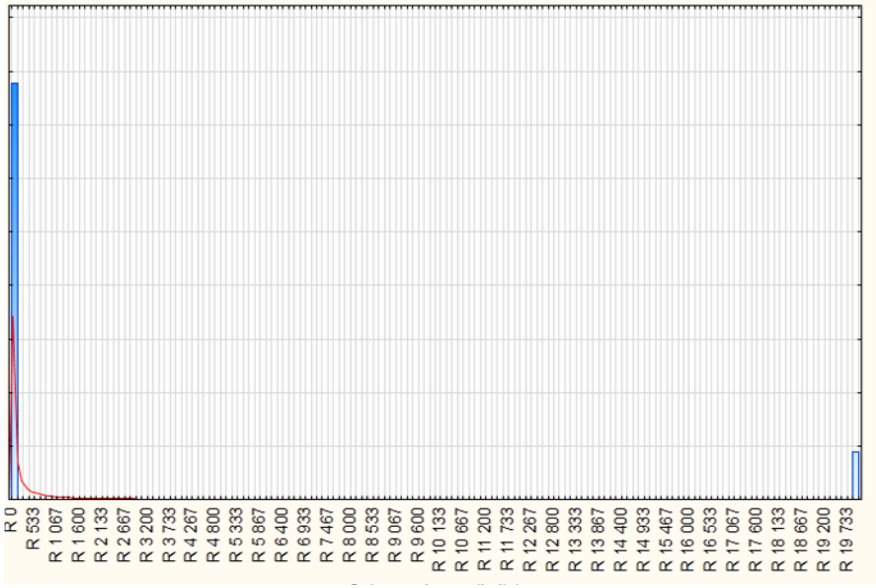
Agenda

- Background
- Data
- Models
- Methodology
- Preliminary results
- Conclusions and further discussion

Background

- Healthcare data exhibits certain distributional features





- Models built on traditional assumptions may not be accurate

Data

- Medical scheme population, claims and authorisations data considered
- Healthcare utilisation explanatory variables
  - Demographic
  - Conditions
  - Healthcare system
- Utilisation measured by total amount claimed

Models

- Log normal model
- Compound distribution (Tweedie model)
- Censored regression (Tobit model)
- Two-part model (Logit-Lognormal model)

Methodology

- Division into training, validation sample
- Models for IH, OH, episodes of care
- Model fitting
- Stratification of population
- In-sample model validation

Some preliminary results

- Models were ranked in order of lowest mean percentage error:

	Log normal	Tobit	Tweedie	Two-part
IH	4	3	2	1
OH	3	4	1	2
Overall	4	3	2	1
IH low risk	4	3	1	2
IH high risk	2	3	4	1
OH low risk	4	2	1	3
OH high risk	2	3	4	1
Chronic lives	3	2	4	1

Further insights

- Results provide support for principal-agent theory
- Analysis of IH and OH, and of frequency and amount components, do not require same sets of variables
- Amount component has fewer significant variables than count component

Conclusions and further discussion

- Different models appropriate for different uses
- Traditional models may give inappropriate results
- Two-part model performs well
- Greater dataset, further analysis and full results to come
- Model accuracy one of several considerations in modelling
- Modelling beyond medical scheme data

Thank you

CareGauge
Health Systems Measurement