



Anwendung von GLMs (Generalized Linear Models) in der Praxis: Stärken und Schwächen

Bernhard König
Bern, 15.11.2013

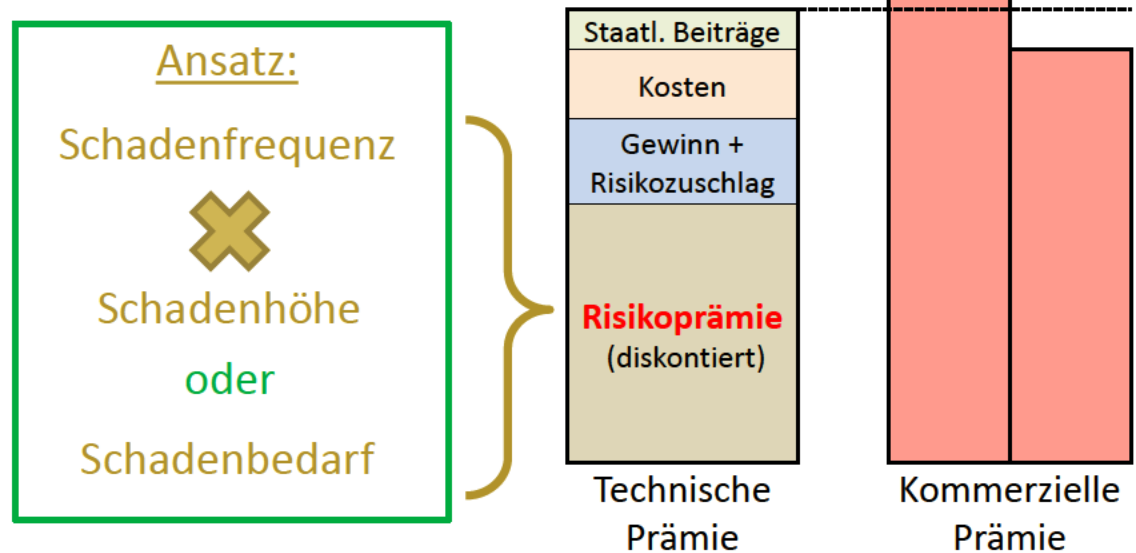


Pricing von Versicherungsprodukten

Die Schadenkosten in einem Portfolio unterscheiden sich für verschiedene Kundengruppen. Beim Pricing untersucht man dies und bestimmt den Preis für die verschiedenen Kundengruppen.



Man modelliert die Differenzierung der Risikoprämie.

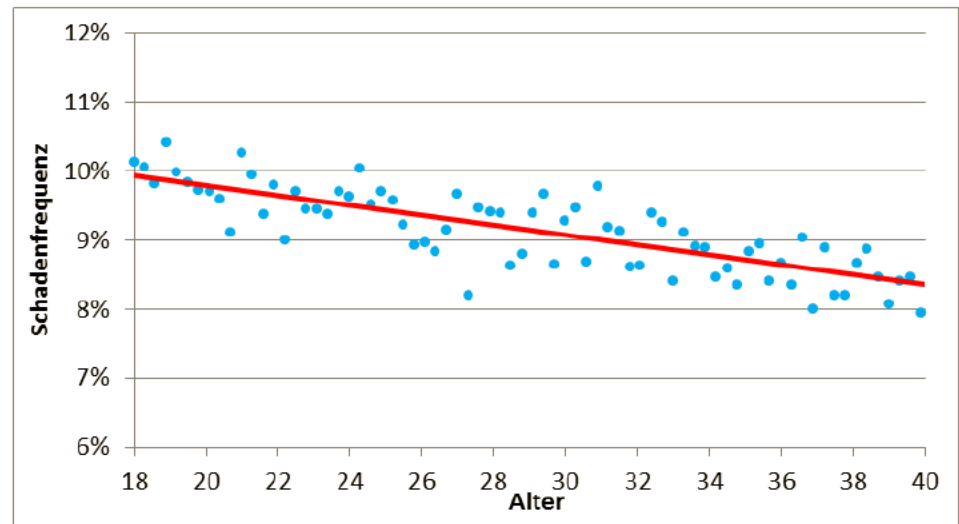


Dabei werden mithilfe von **historischen Daten** und durch die Berücksichtigung von Trends Abschätzungen über die **Zukunft** getroffen.

Lineare Regression

Bei der **linearen Regression** wird eine unbekannte Zielvariable Y (z.B. Schadenfrequenz) durch verschiedene erklärende Variablen X (z.B. Alter, Geschlecht, etc.) geschätzt. Dabei wird angenommen, dass $Y_i \sim N(\mu_i, \sigma^2)$ und $\mu_i = x_i \beta$ für Regressionskoeffizienten β .

$$Y = X \beta + \varepsilon$$



Dieses Modell ist unpassend für die Modellierung von **Versicherungsdaten**:

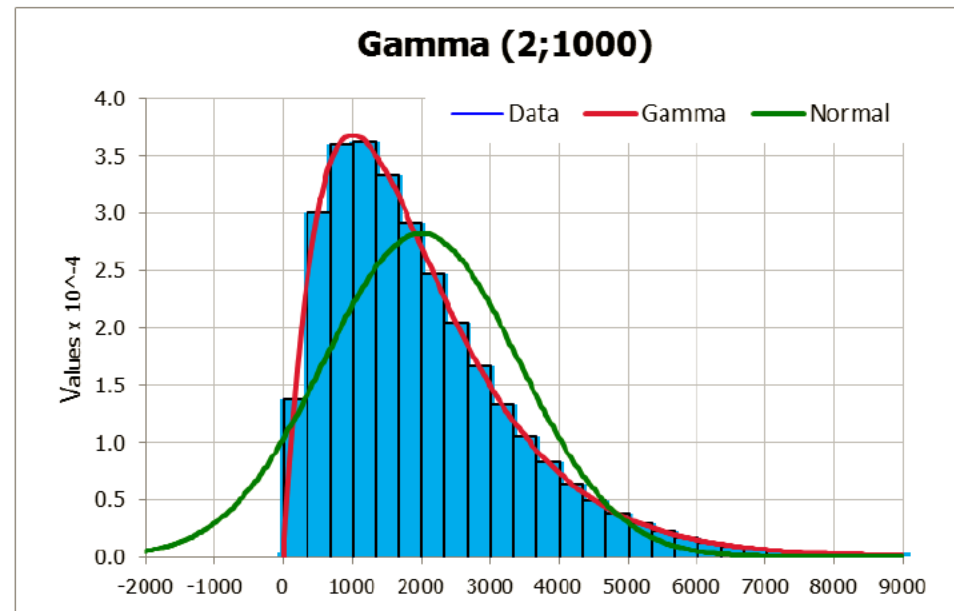
1. Die Annahme der Normalverteilung ist unrealistisch.
2. Der lineare Zusammenhang zwischen Regressionskoeffizienten und Mittelwert ist nicht wünschenswert.

Die Verallgemeinerung des linearen Modells

Ein GLM verallgemeinert die Annahmen des linearen Modells:

1. Die Zielvariable Y kann auch andere Verteilungen haben (Poisson, Gamma).
2. Die Daten werden durch eine **Link Funktion g** (z.B. Logarithmus) transformiert.

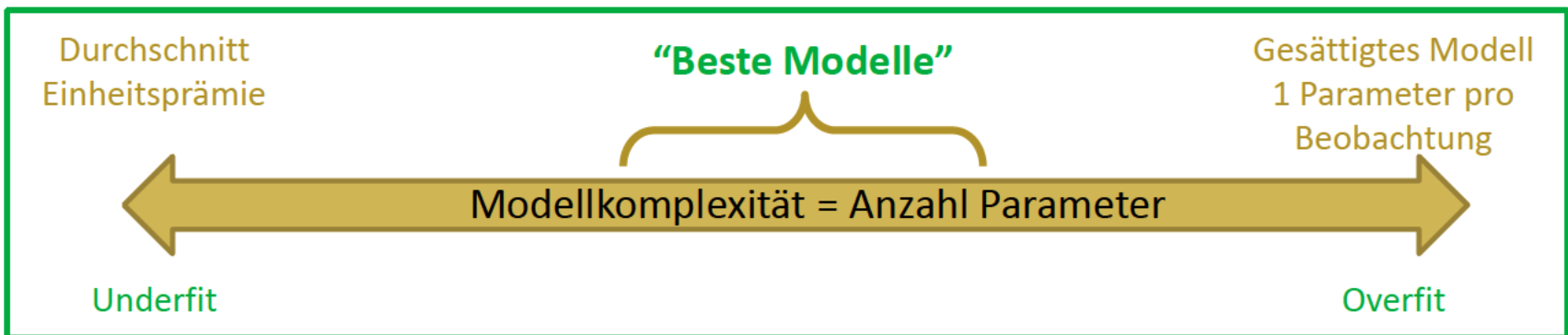
$$Y = g^{-1}(X \beta) + \varepsilon$$



$Y = g^{-1}$ (Linearkombination von Faktoren) + Fehler

GLMs in der Praxis

1. Ansatz:
 - 2 Modelle: Schadenfrequenz und Schadendurchschnitt
 - 1 Modell: Schadenbedarf
2. Modellierung:
 1. Wahl der relevanten Faktoren (Erklärende Variablen)
 2. Wahl der Relativitäten



Stärken und Schwächen der GLMs



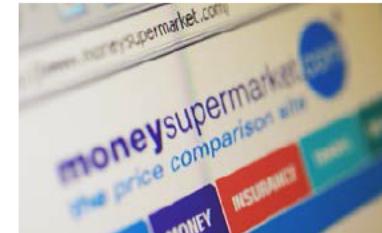
- **Modell**
 - Verteilungsfunktionen
 - Multiplikativer Tarif
 - Wird Korrelationen zwischen den Faktoren gerecht
 - Einbezug von Interaktionen
 - Einbezug von Randbedingungen
- **Struktur**
 - Interpretation & Erklärbarkeit
- **Verbreitet**
 - Bekannte Theorie (->statistische Tests)
 - Kommerzielle Software



- **Modellannahmen**
 - Folgt die Realität einer Wahrscheinlichkeitsverteilung bzw. der Struktur des Modells?
 - Unabhängigkeit von Frequenz & Durchschnitt (Personenschäden, Diebstahl, Hagel)
- **Modellierung**
 - Einfache GLMs modellieren alle Werte einer Variable einzeln. Auch Variablen wie Alter werden als kategoriell interpretiert, obwohl eine natürliche Abhängigkeit zwischen benachbarten Werten besteht.
 - Korrelierte Erklärende: Assoziation <-> Kausalität
- **Interaktionen**
 - Wie findet man (neue) Interaktionen?

Probleme in der Praxis

- **Risikodifferenzierung:** Wie weit will man differenzieren?
 - Regulator & Gesetzgebung
 - Versicherungsgedanke: Ausgleich im Kollektiv
 - Preisdruck auf dem Markt: Konkurrenz, Vergleichswebsites, Kunden
 - In der Regel gilt: Je höher der **Preisdruck** und die **Preissensitivität** desto wahrscheinlicher ist es, dass die Modelle überparametrisiert werden (Overfit).
 - Stark abhängig vom Land
- Das Umfeld, insbesondere **IT und Vertrieb**, hat einen grossen Einfluss auf die Möglichkeiten der Pricing Aktuare



Alle Marktteilnehmer machen «**Best Practice**», dennoch beobachtet man bei Preisvergleichen sehr grosse Verwerfungen auf dem Markt



Fazit und persönliche Meinung

- Mit GLMs lässt sich eine **gute Basis** für einen Tarif mit multiplikativer Struktur herleiten.
- Heute nimmt man oft **Polynome** zur Hand um einem Tarif, insbesondere bei kleinen Datenmengen, eine sinnvolle Struktur zu geben (Beispiel Alterskurve).
- Alternative Modelle wie **Machine Learning** Ansätze können helfen den GLM Tarif durch **nicht-lineare** Abhängigkeiten zu verfeinern.