Quantilsregression

Martin Gubisch

Universität Konstanz

13. April 2011

Warum ein Berufspraktikum?

Erfahrungen aus der „Realität“

- Berufliche Erfahrungen
- Berufliche Perspektiven
- Qualifikation
- „Hardsklills“ & „Softskills“
- Neugier auf Arbeitsweise, Methoden ... außerhalb der reinen Wissenschaft

Die SAF Simulation, Analysis and Forecasting AG

- Schweizerische Aktiengesellschaft, gegründet 1996
- Hauptsitz in Tägerwilen/Schweiz; Niederlassungen in Dallas/USA und Bratislava/Slowakei.
- Softwarepakete zur automatischen Disposition und Bestandsoptimierung in Handelsfilialen
- Bestandsminimierung nicht von Produzentenseite her, sondern auf Grundlage der Kundennachfrage
- Kunden sind u.A. Lebensmittelgroßhändler, Drogerien, Discounter, Warenhäuser und Baumärkte
- Enge Anpassung der Produkte an individuelle Kundenwünsche
- Anforderungen: Prognosegenauigkeit, hoher Automatisierungsgrad & kurze Datenverarbeitungszeiten

Einige Unterschiede zum akademischen Arbeiten

- Viel Kommunikation mit Nicht-Experten und als Nicht-Experte
- Integration in und Kooperation mit anderen Teams
- strengere Fristen, verbindliche Vorausplanung
- Weniger penibles, ergebnisorientierteres Arbeiten
- Dokumentationen, die man auch nach Jahren noch versteht
- Differenzieren zwischen Vielfältigkeit und Benutzerfreundlichkeit der Programme
- Berücksichtigung „exotischer“ Kundenwünsche

Aufgabenfeld und Tätigkeiten

- Einarbeitung in verschiedene Verfahren zur Erstellung von Prognosen
- Präsentation des theoretisch-mathematischen Hintergrunds
- Entwicklung und Programmierung effizienter Algorithmen
- Hierbei: Einarbeitung in eine Programmiersprache
- Erstellen einer ausführlichen Dokumentation
- Test und Vergleich verschiedener Prognoseverfahren in einer Fallstudie
- Aufbereitung und Transformation großer Datenmengen
- Analyse und Präsentation der Resultate

Prognoseverfahren

Auf Grundlage von Abverkäufen $y_1, \ldots, y_n$ der vergangenen n Wochen soll eine Prognose für die Anzahl $\hat{y}$ der Verkäufe in der folgenden Woche erstellt werden.

Möglichkeit I:

$$\hat{y} = \frac{1}{n} \sum_{i=1}^{n} y_i,$$

d.h. die Prognose $\hat{y}$ ist das **Arithmetische Mittel** der Historie $y = (y_1, \ldots, y_n)$.

„Nachteil": Ausreißer, d.h. vereinzelte sehr große oder sehr kleine Werte, haben einen verhältnismäßig großen Einfluss auf das Mittel.

Prognoseverfahren

Zugehöriges Optimierungsproblem:

$$\hat{y} = \underset{y \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - y)^2.$$

Möglichkeit II:

$$\hat{y} = \begin{cases} \operatorname{sort}(y)_{\frac{n+1}{2}} & \text{falls } n \text{ ungerade} \\ \operatorname{sort}(y)_{\frac{n}{2}} & \text{falls } n \text{ gerade} \end{cases},$$

d.h. man nimmt den Wert der mittleren Beobachtung, den sog. **Median**.

„Nachteil“: Ein relativ großer Anteil an historischen Werten kann bei der Prognose ignoriert werden.

Prognoseverfahren

Zugehöriges Optimierungsproblem:

$$\hat{y} = \underset{y \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} |y_i - y|.$$

Bei Abverkaufsprognosen hat man in der Regel nicht als Ziel, dass sich Über- und Unterschätzungen die Waage halten, sondern man möchte möglichst selten unterschätzen (damit die Kunden möglichst selten vor leeren Regalen stehen), ohne die Nachfrage unnötig hoch zu überschätzen (damit im Winter die Verkaufsregale nicht rappelvoll mit Rasenmähern sind oder der Kunde die Erdbeeren von vorletzter Woche immer noch bestaunen kann).

Konkret: $\hat{y}$ **soll ein möglichst kleiner Wert sein, so dass mindestens** $\tau\%$ **der historischen Werte kleiner oder gleich** $\hat{y}$ **sind für ein** $\tau \approx 95$**.**

Prognoseverfahren

Modifikation für Möglichkeit I:

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} y_i + \sigma \cdot \tau\%,$$

wobei $\sigma = \left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2\right)^{\frac{1}{2}}$ die Standardabweichung bezeichnet.

Modifikation für Möglichkeit II:

$$\hat{y} = \begin{cases} \mathrm{sort}(y)_{\lfloor n\cdot\tau\% \rfloor + 1} & \text{falls } n\cdot\tau\% \notin \mathbb{N} \\ \mathrm{sort}(y)_{n\cdot\tau\%} & \text{falls } n\cdot\tau\% \in \mathbb{N} \end{cases} .$$

Prognoseverfahren

Zugehöriges Optimierungsproblem:

$$\hat{y} = \underset{y \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \varrho_\tau(y_i - y),$$

wobei

$$\varrho_\tau(y) = \begin{cases} (\tau\% - 1) \cdot y & \text{falls } y < 0 \\ \tau\% \cdot y & \text{falls } y \geq 0 \end{cases}.$$

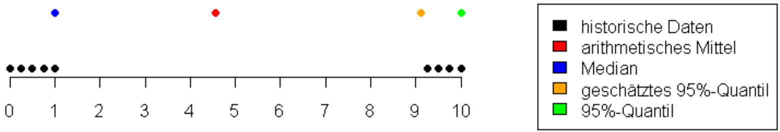
$\hat{y}$ heißt das $\tau\%$-**Quantil**.

Vorteil: Im Gegensatz zu (I) werden Quantile nicht geschätzt, sondern exakt bestimmt.

Prognoseverfahren

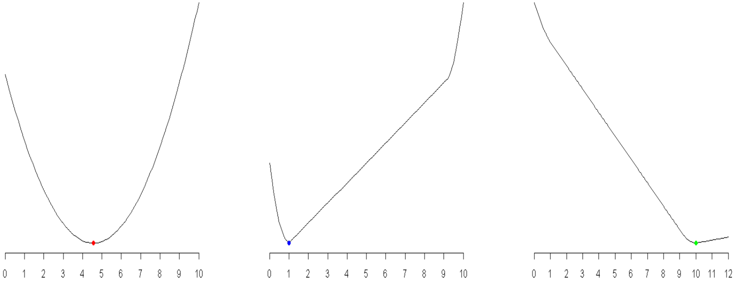
Beispiel: Badezimmerfliesen, 10er-Packungen.

Fünf Kunden wollen nur einzelne Fliesen ersetzen, vier Kunden im ganzen Bad neue Fliesen legen.



Prognoseverfahren

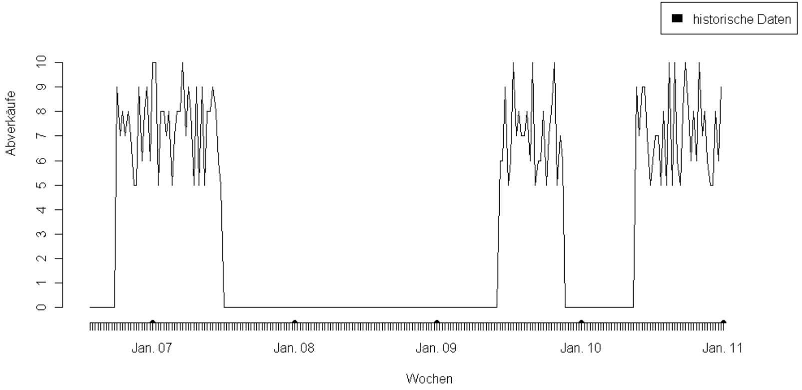
Zugehörige Optimierungsprobleme:



Bislang ist nicht ersichtlich, wozu die korrespondierenden Optimierungsprobleme benötigt werden; bei allen Verfahren ist es viel einfacher, die Prognosen direkt über die Formeln zu bestimmen.

Prädiktoren

Beispiel: „Out-of-Stock“.



Prädiktoren

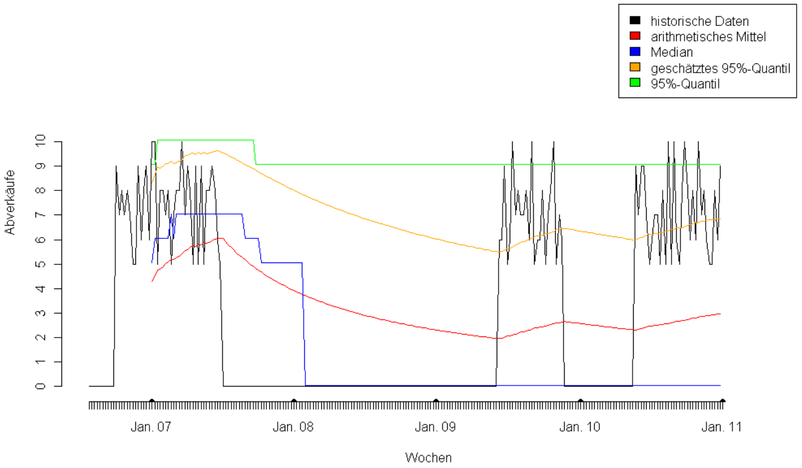
Problem: „Out-of-stock“-Perioden sollten nicht wie Null-Nachfragen interpretiert werden: Die Kunden würden kaufen, wenn Ware da wäre.

Bei den bisherigen Verfahren würden die Prognosen in den Perioden, in denen der Artikel inaktiv ist, langsam sinken (und dabei viel zu hoch sein) und in den aktiven Perioden langsam wieder steigen (und dabei viel zu niedrig sein).

Wenn zusätzlich zu den Abverkaufsdaten die inaktiven Perioden des Artikels übermittelt werden, können Nullabverkäufe und „Out-of-Stock“-Perioden bei der Prognose berücksichtigt werden.

Modifikation: **Prädiktoren** werden gesetzt.

Prädiktoren



Prädiktoren

Bisher wurde zur Schätzung von $\hat{y}$ das lineare Modell

$$y_i = \hat{y} + \varepsilon_i$$

verwendet. $\hat{y}$ wurde dabei so gewählt, dass die Fehler ε_i – quadratisch (I) oder absolut (II) – minimiert wurden.

Wir erweitern das Modell zu

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \varepsilon_i,$$

wobei

$$X_i = \begin{cases} 1 & \text{falls der Artikel in der } i\text{-ten Woche aktiv ist} \\ 0 & \text{falls der Artikel in der } i\text{-ten Woche inaktiv ist} \end{cases}$$

Prädiktoren

Die Koeffizienten $\hat{\beta}_0, \hat{\beta}_1$ werden dann wieder Fehler-minimierend bestimmt mittels

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - (\beta_0 + X_i\beta_1))^2$$

beziehungsweise

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \varrho_\tau(y_i - (\beta_0 + X_i\beta_1)).$$

Gibt $x \in \{0, 1\}$ an, ob der Artikel zum Zeitpunkt der Prognose aktiv oder inaktiv ist, dann lautet die Prognose

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x.$$

Prädiktoren

Bei Bedarf können mehrere Prädiktoren $(X_{i1}, \dots, X_{iN})_{i=1}^{n}$ gesetzt werden:

$$y_i = \hat{\beta}_0 + \sum_{j=1}^{N} \hat{\beta}_j X_{ij} + \varepsilon_i.$$

X wird **Design-Matrix** genannt.

Die Prognose lautet dann

$$\hat{y} = \hat{\beta}_0 + \sum_{j=1}^{N} \hat{\beta}_j x_j,$$

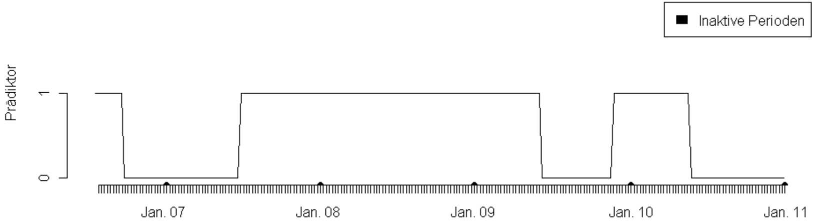
wobei x_j angibt, ob der j-te Prädiktor zum Zeitpunkt der Prognose aktiv ist oder nicht.

Prädiktoren

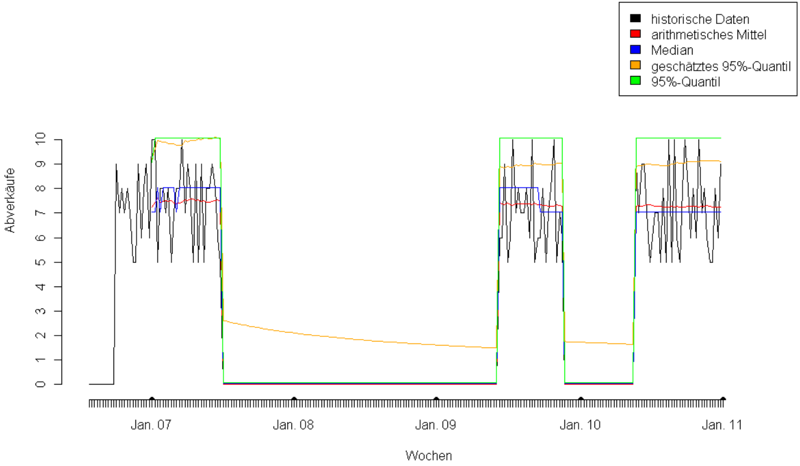
Ist der Prädiktor aktiv, so werden zur Prognose nur diejenigen historischen Abverkäufe in Betracht gezogen, bei denen der Prädiktor ebenfalls aktiv war. Entsprechendes gilt für die inaktiven Perioden.

Eindeutigkeitsprobleme können entstehen, wenn sich Prädiktoren überschneiden.

Neben Booleschen Prädiktoren sind auch metrische möglich, etwa um die Auswirkungen einer Preissenkungen einzukalkulieren.

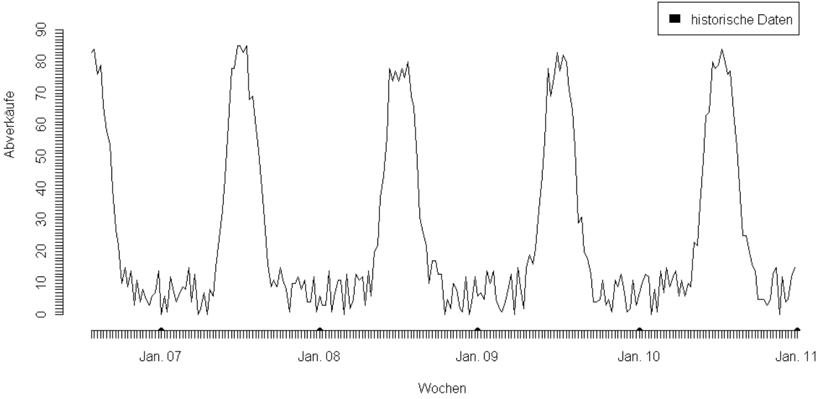


Prädiktoren

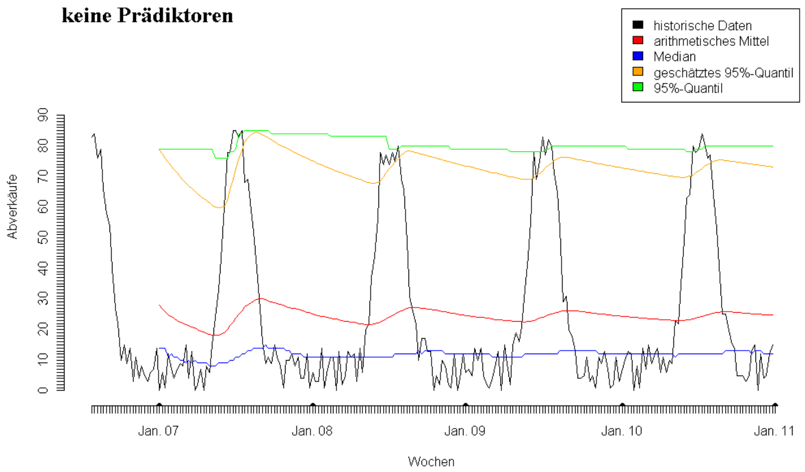


Saisoneffekte

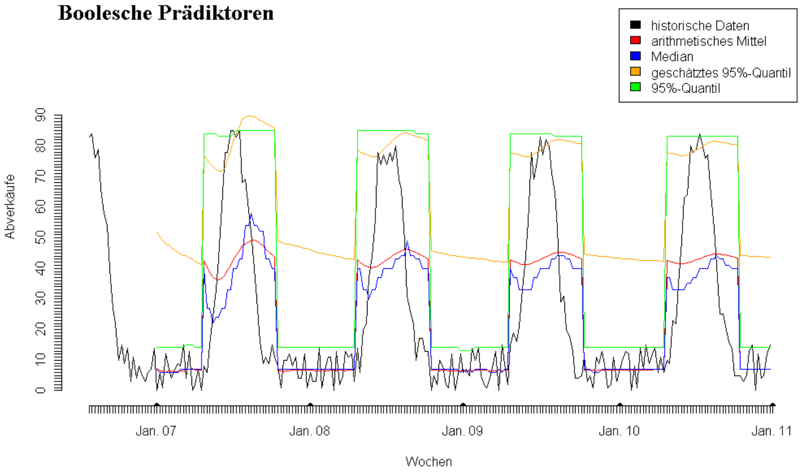
Beispiel: „Erdbeersaison“.



Prädiktoren



Prädiktoren



Lösen der Optimierungsprobleme

Für Methode I,

$$(\hat{\beta}_0, \ldots, \hat{\beta}_N) = \underset{(\beta_0,\ldots,\beta_N)\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - \left(\beta_0 + \sum_{j=1}^{N} X_{ij}\beta_j \right) \right)^2 ,$$

liefert Differenzieren der glatten, quadratischen Zielfunktion direkt ein lineares Gleichungssystem für $\hat{\beta}_0, \ldots, \hat{\beta}_N$.

Methode II,

$$(\hat{\beta}_0, \ldots, \hat{\beta}_N) = \underset{(\beta_0,\ldots,\beta_N)\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \varrho_\tau \left(y_i - \left(\beta_0 + \sum_{j=1}^{N} X_{ij}\beta_j \right) \right) ,$$

besitzt dagegen eine nur stückweise lineare und stetige, aber nicht differenzierbare Zielfunktion.

Lösen der Optimierungsprobleme

Lösungen des Problems korrespondieren mit Ecken der Zielfunktion und erfüllen daher

$$\beta_0 + X(I)\beta = y(I),$$

wobei $I \subseteq \{1, ..., n\}$, $\#I = N$ und $X(I) = (X_{i,j})_{i \in I}^{j \in \{1,...,N\}}$ und $y(I) = (y_i)_{i \in I}$.

Genauer: Entweder die Lösung ist eindeutig bestimmt, d.h. es gibt genau eine solche Indexmenge I und $X(I)$ ist invertierbar, oder die Lösungsmenge ist die konvexe Hülle (eine „Facette“ der Zielfunktion) gewisser Ecklösungen.

Das Optimierungsproblem kann in ein **lineares Programm** überführt und mit dem **Simplex-Algorithmus** gelöst werden.

Die punktweise Auswertung der $\binom{n}{N}$ möglichen Ecklösungen (so viele N-elementige Teilmengen von $\{1, ..., n\}$ gibt es nämlich) ist im Allgemeinen numerisch zu aufwendig.

Lösen der Optimierungsprobleme

Wir müssen das nichtlineare Optimierungsproblem ohne Nebenbedingungen auf die Form

$$\min_x c(x) = c^t x \quad \text{so dass} \quad \begin{cases} Px &= y \\ x &\geq 0 \end{cases}$$

transformieren.

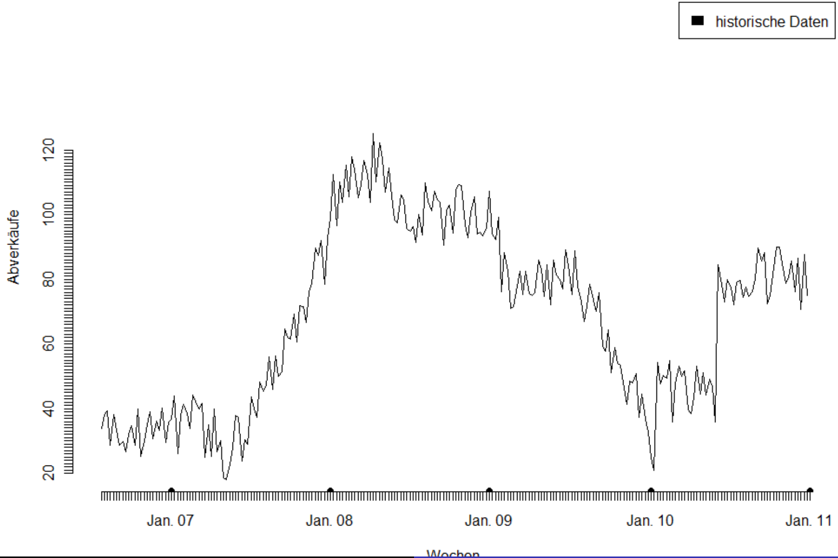
Wir teilen die Residuen $y_i - X_{i.}\beta$ hierfür in ihre positiven und negativen Anteile, u_i und v_i, auf und erhalten

$$\min_{u,v,\beta^+,\beta^-} \tau \sum_{i=1}^{n} u_i + (1-\tau) \sum_{i=1}^{n} v_i$$

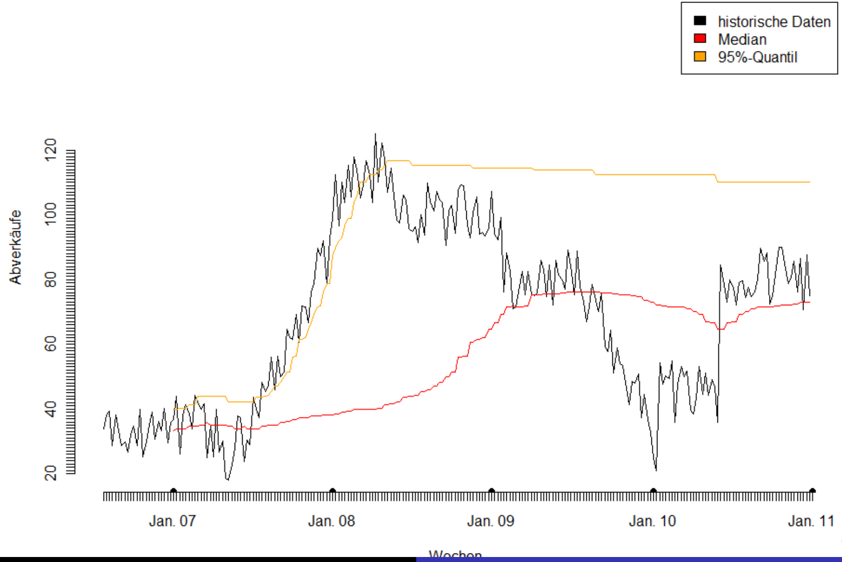
unter den Nebenbedingungen

$$\begin{cases} u - v + X\beta^+ - X\beta^- &= y \\ u, v, \beta^+, \beta^- &\geq 0 \end{cases}.$$

Gewichtung


historische Daten
Abverkäufe
20
40
60
80
100
120
Jan. 07
Jan. 08
Jan. 09
Jan. 10
Jan. 11
Wochen

Gewichtung


historische Daten
Median
95%-Quantil
Abverkäufe
20
40
60
80
100
120
Jan. 07
Jan. 08
Jan. 09
Jan. 10
Jan. 11
Wochen

Gewichtung

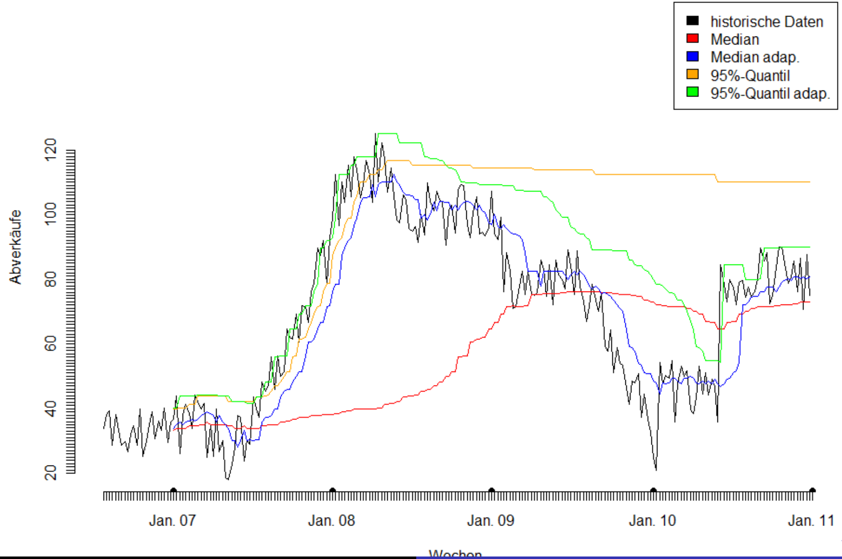
Wächst oder fällt die Nachfrage längerfristig, so haben die bisherigen Prognoseverfahren die unerfreuliche Angewohnheit, den Ereignissen „hinterherzuhinken“.

Führt man kein Verfahren ein, das solche „Trends“ automatisch erkennt und in die Prognose einbezieht, kann eine stärkere **Gewichtung** der jüngeren Abverkaufsdaten Abhilfe schaffen.

Nachteil: Eine zu starke Gewichtung führt dazu, dass viele Informationen der älteren Historie für die Prognose ignoriert werden. Die Prognosen können so erheblich schlechter werden.

Gut gesetzte Prädiktoren können diesen unerwünschten Effekt ausgleichen.

Gewichtung


historische Daten
Median
Median adap.
95%-Quantil
95%-Quantil adap.
Abverkäufe
20
40
60
80
100
120
Jan. 07
Jan. 08
Jan. 09
Jan. 10
Jan. 11
Wochen

Gewichtung

Zu einem Vektor $w = (w_1, \dots, w_n)$ von Gewichten lauten die zugehörigen Optimierungsprobleme

$$(\hat{\beta}_0, \dots, \hat{\beta}_N) = \underset{(\beta_0, \beta_1) \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \left(y_i - \left(\beta_0 + \sum_{j=1}^{N} X_{ij} \beta_j \right) \right)^2$$

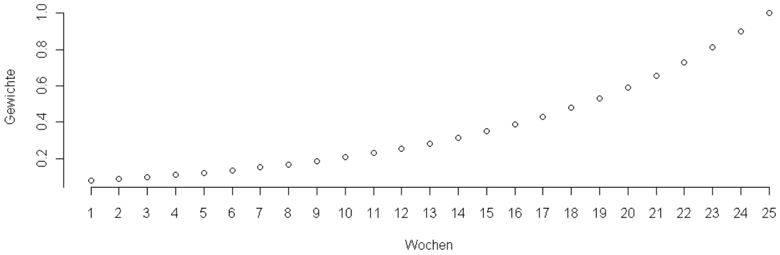
für Methode (I) beziehungsweise

$$(\hat{\beta}_0, \dots, \hat{\beta}_N) = \underset{(\beta_0, \beta_1) \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \varrho_\tau \left(y_i - \left(\beta_0 + \sum_{j=1}^{N} X_{ij} \beta_j \right) \right)$$

für Methode (II).

Gewichtung

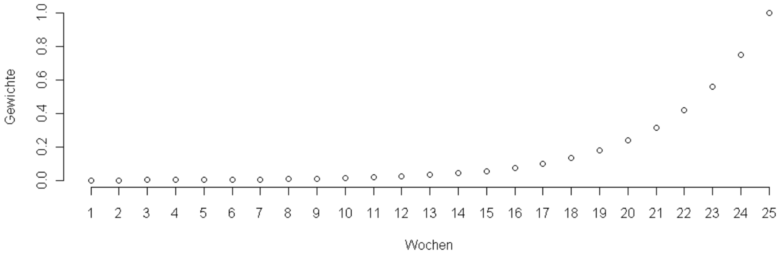
Verschiedene Gewichtungen



Jedes Gewicht beträgt 90% des vorherigen.

Gewichtung

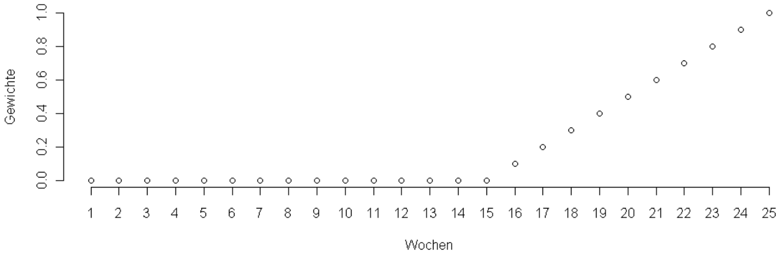
Verschiedene Gewichtungen



Jedes Gewicht beträgt 75% des vorherigen.

Gewichtung

Verschiedene Gewichtungen



Nur die ersten 10 historischen Werte werden berücksichtigt, und zwar zu 100%, 90%, ..., 10%.

Referenzen

Referenzen

- http://www.saf-ag.com
- Koenker, R.: *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press, 2005
- Koenker, R. & Hallock, K.: *Quantile Regression*, Journal of Economic Perspectives, Vol. 15, No 4, 2001, pp. 143-156

Vielen Dank für die Aufmerksamkeit.