Energiemodelle und Analysen
SS 2014

Übung 1

Regressionsanalyse

Beispiel 2

Beispiel 2:

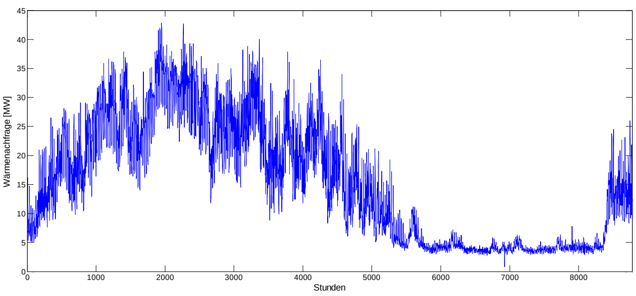
Modellierung der Wärmenachfrage eines Fernwärmenetzes mit linearer Regression - einfache Modellansätze

Gegeben ist die stündliche Nachfrage nach Wärme (Raumwärme und Warmwasser) in einem Fernwärmenetz (gemessene Leistungsmittelwerte der Einspeisung - stündlich) und die dazugehörige Umgebungstemperatur. Die Nachfrager sind hauptsächlich Haushalte, zum Teil aber auch Gewerbebetriebe. Vergleichen und interpretieren Sie unterschiedliche Modellansätze zur Abschätzung der Nachfrage in Abhängigkeit von der Temperatur und der Tageszeit:

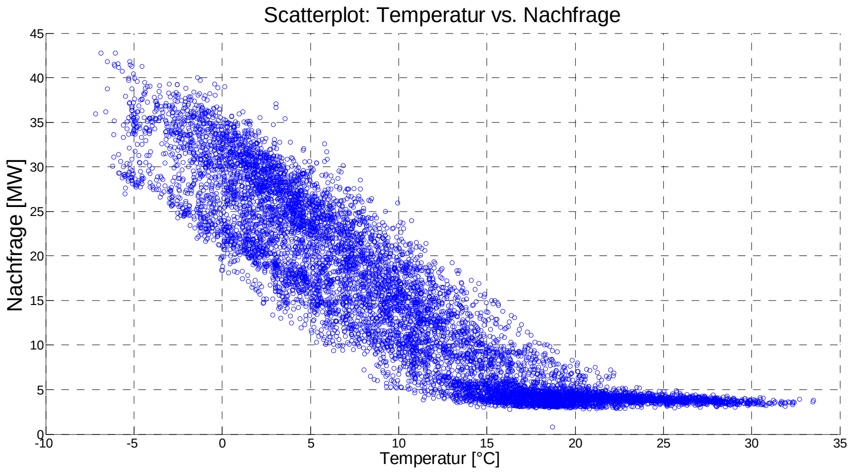
Beobachtungszeitraum: 1.10.2007 bis 1.10.2008

Stündliche Leistungsmittelwerte

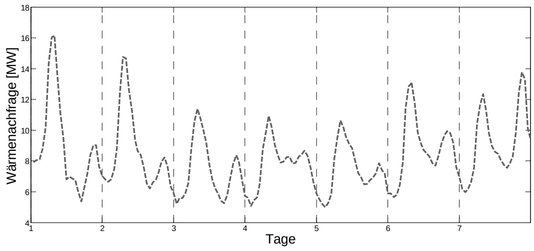
t ϵ {1….8774}



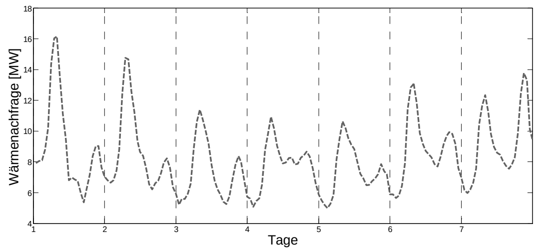
Vermutete Einflussfaktoren



Vermutete Einflussfaktoren

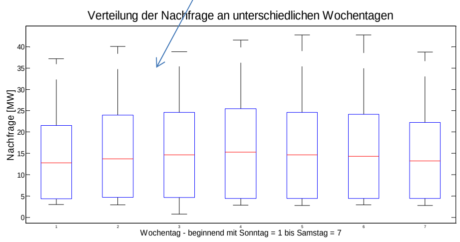


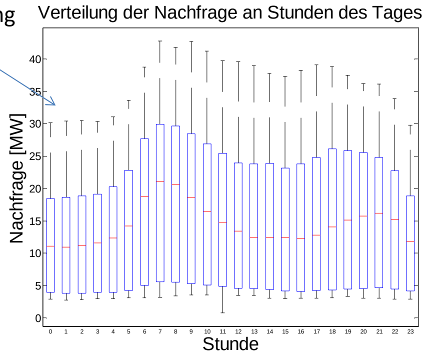
Vermutete Einflussfaktoren



eher geringer Einfluss des Wochentags?

Starker Tagesgang





Wie können die Einflussfaktoren in einem Modell abgebildet werden?

Wozu?

- Prognosen (falls Einflussfaktoren vorhersagbar)
- Analysen von Einflussfaktoren

2.1) Modell 1: Einfache Abschätzung über Temperatur

$$y_t = \beta_0 + \beta_1 \cdot T_t$$

In diesem Modellansatz hängt die Nachfrage nur von der Umgebungsemperatur ab. Die Koeffizienten β_i (i=0,1) ergeben sich jeweils aus der linearen Regression, wobei die Funktion fitlm(…) zu verwenden ist. (In der Matlab Ausgabe unter der Spalte „Estimate“ bzw. in der Variable *„Modellname“*.Coefficients.Estimate(*index*).) Der Koeffizient β_0 entspricht jeweils der Konstanten (=Intercept).

a) Geben Sie die Koeffizienten und die dazugehörige t-Statistik an. Wie interpretieren Sie die Ergebnisse?

b) Vergleichen Sie die modellierten Werte für die Beobachtungen von t=1680:1775 bzw. t=6480:6575. Erstellen Sie dazu eine Grafik. Wie interpretieren Sie die Abweichungen? Wodurch unterscheiden sich die Abweichungen in den beiden Beobachtungsperioden?

2.2) Modell 2: Versuch der zusätzlichen Modellierung des Tagesverlaufs

$$y_t = \beta_0 + \beta_1 \cdot T_t + \beta_2 \cdot h_t + \beta_3 \cdot h_t^{\,2} + \beta_4 \cdot h_t^{\,3}$$

Hier wird versucht, den typischen Tagesverlauf der Nachfrage (der nicht von der Temperatur abhängt) in das Modell zu integrieren. Die Variable h entspricht dabei der Spalte „Stunde“ im Datenblatt (bzw. data_heat.Stunde im Mat-File), weist also jeder Beobachtung, die dazugehörige Stunde zu. Die Stunden gehen hier zusätzlich zur Temperatur als Polynom 3.Grades in das Modell ein (Die Daten müssen also dementsprechend aufbereitet werden bevor die Regression durchgeführt wird). Die jeweiligen Koeffizienten werden wiederum über die lineare Regression mit der Funktion fitlm(…) geschätzt.

a) Geben Sie die Koeffizienten und die dazugehörige t-Statistik an. Wie interpretieren Sie die Ergebnisse?
b) Vergleichen Sie erneut die modellierten Werte für die Beobachtungen von t=1680:1775 bzw. t=6480:6575 und erstellen Sie eine Grafik. Was beobachten Sie? Wieso kann der gewählte Modellansatz den charakteristischen Tagesverlauf der Beobachtungen nicht wiedergeben?

2.3) Modell 3: Modellierung der Nachfrage getrennt für einzelne Stunden

$$y_t^j = \beta_0 + \beta_1 \cdot T_t^j$$

Der Modellansatz entspricht dem Ansatz aus Modell 1. Allerdings werden nicht mehr alle Beobachtungen in das Modell aufgenommen. Es wird jeweils eine Regression nur für jene Beobachtungen, die zur jeweiligen Stunde j aufgezeichnet wurden in das Modell aufgenommen. Führen Sie diesen Ansatz für die Stunde 7 und für die Stunde 23 durch. Bevor Sie die Regression durchführen, müssen Sie die Daten nach den jeweiligen Tagesstunden filtern (z.B. mit data=data(data.Stunde=x,:)). Im beigefügten Matlab Skript wird beispielhaft gezeigt, wie Sie aus dem Dataset Data_heat die Daten für eine bestimmte Stunde auswählen können.

a) Vergleichen Sie die beiden Konstanten (β_0) sowie die Koeffizienten des Temperatureinflusses (β_1) aus den Ergebnissen der Regression für Stunde 7 und Stunde 23. Wie interpretieren Sie diese und wie interpretieren Sie die Unterschiede zwischen den beiden Stunden?

b) Vergleichen Sie das jeweilige Bestimmtheitsmaß (R2) aus den beiden Modellen für Stunde 7 und 23 nach Modellansatz 3 mit dem Bestimmtheitsmaß aus Modell 1. Wie würden Sie die Qualität der beiden Modellansätze beurteilen? Woraus ergibt sich der Unterschied?

c) Vergleichen Sie die modellierten Werte für die Beobachtungen für Stunde 7 und für Stunde 23 mit den gemessenen Werten der jeweiligen Stunde über alle 366 Tage. Erstellen Sie dazu eine Grafik. Was beobachten Sie? Wieso schwanken zu einer bestimmten Zeit im Jahr die modellierten Werte um die relativ konstante gemessene Nachfrage? Vergleichen Sie dazu den modellierten Zusammenhang zwischen Temperatur und Nachfrage (Skizze) mit dem Scatterplot Temperatur vs. Nachfrage aus der Angabe. (Scatterplot siehe Matlab Skript oder Folien zur Übungsangabe).

2.4) Zusatzfrage: Verbesserungen des Modells

(Hier gibt es nur zusätzliche Punkte für die Übung aber ***keine Abzüge****, wenn Sie die Frage nicht beantworten. Maximale Punkteanzahl bleibt allerdings 12,5 Punkte auf die gesamte Übung.)*

a) Wie würden Sie vorgehen, um das Problem, das im letzten Punkt von Beispiel 2.3.c beobachtet wurde zu beheben und damit bessere Vorhersagen für diesen Zeitraum zu erhalten?
b) Welche Verbesserungsvorschläge für weitere Modellansätze fallen Ihnen ein? Gehen Sie dabei allerdings weiter davon aus, dass Ihnen nur die gegebenen Daten bzw. allgemein zugängliche Daten zur Verfügung stehen. Formulieren Sie wenn möglich einen verbesserten Modellansatz mathematisch.

Anhang:

Hier wird nur kurz die allgemeine Form eines Regressionsmodells angedeutet und Definitionen für Prüfgrößen des Modells beschrieben. Sie benötigen diese Angaben **nicht** für die Lösung der Übungsaufgabe!!! Sie dienen nur als Hintergrundinformation zu den Funktionen und Ausgabewerten der Regressionsergebnisse in Matlab bzw. anderer Software. Für Interessierte hier noch ein Link zu einem Skriptum des IHS Wien in dem die wichtigsten Begriffe und Herleitungen beschrieben werden:
http://homepage.univie.ac.at/robert.kunst/emwi.pdf

Zudem finden Sie in zahlreichen Online-Tutorials weitere Hintergrundinformationen.

Beispielcode Regressionsbeispiel_Strompreise.m

- Modell zur Abschätzung von Einflussfaktoren auf den stündlichen Strompreis
- Hier finden Sie mögliche Herangehensweisen und die wichtigsten Matlabbefehle
- Das File ist nur zur Unterstützung gedacht, ist aber nicht Teil der Übung!!!
- Auch als .pdf Report im zip-File

Beispiel 1

Ökonometrisches Nachfragemodell

1.1 Aufgabenstellung

Für zwei Länder ist vergleichsweise zu untersuchen, welche Parameter den STROMverbrauch im Sektor **Private Haushalte** beeinflussen. Dazu sind im Wesentlichen Preis- und Einkommenselastizitäten zu schätzen. Ausgangsdaten sind Zeitreihen dieser Länder für:

- Stromverbrauch Haushalte
- Strompreis
- Bruttoinlandsprodukt
- Index Real/nominal

Diese Daten werden zur Verfügung gestellt, siehe File attached.
Basierend auf den Ergebnissen der Regression sind **vier Szenarien** bis 2030 zu rechnen:

* Hochpreis + hohes GDP-Wachstum
* Hochpreis + niedriges GDP-Wachstum
* Moderater Preis + hohes GDP-Wachstum
* Moderater Preis + niedriges GDP-Wachstum

Annahmen:
Hochpreis: +4%/Jahr; + Moderater Preis +2%/Jahr; hohes GDP-Wachstum: +3%/Jahr; niedriges GDP-Wachstum + 1 %/Jahr;

Ausgangsmodell

Der einfachste Ansatz dazu lautet:

$$E_t = K \cdot p_t^{\alpha} \cdot Y_t^{\beta}$$

mit:

K	Konstante
E_t	Stromnachfrage im Jahr t
p_t	Haushaltsstrompreis im Jahr t
Y_t	Einkommen (z.B. Bruttoinlandsprodukt (BIP)
α	Preiselastizität
β	Einkommenselastizität

Durch logarithmieren können lineare Koeffizienten ermittelt werden!

$$\ln(E_t) = C + \alpha \cdot \ln(p_t) + \beta \cdot \ln(Y_t)$$

Vorgehensweise

- Überprüfen der Daten für die untersuchten Länder
- Umrechnen in reale Preise und reales Einkommen
- 3 unterschiedliche Modelle für jedes Land (z.B. Basismodell, Modell mit Trend, Lag, etc.)
- Dokumentation und Interpretation der Ergebnisse
- Erstellung der Szenarien basierend auf dem „brauchbarsten“ Modell.
- Verbesserungsvorschläge für das Modell

Hinweis zur Modellierung von Trend bzw. Lag

- Trend: der Trend entspricht einfach einer Laufvariablen und beginnt mit der ersten Beobachtung. Diese Variable wird einfach als neue unabhängige Variable in das Modell aufgenommen: z.B. $t \in \{1....23\}$ für einen Beobachtungszeitraum von 23 Jahren.

- Ein Lag kann in das Modell aufgenommen werden, in dem einfach die Beobachtung der abhängigen Variable der Vorperiode ($y_{(t-1)}$) als unabhängige Variable in das Modell aufgenommen wird. Damit fällt natürlich die erste Beobachtung aus dem Modell, da es keine Vorperiode dazu gibt.

$$y_t = \beta_0 + \beta_1 \cdot x_t + \beta_2 \cdot \overbrace{y_{t-1}}^{\text{Lag1}} + \beta_3 \cdot \overbrace{t}^{\text{Linearer Trend}}$$

- Manche Programme stellen diese Ansätze (Trend,Lag) automatisch zur Verfügung. In Matlab müssen Sie diese allerdings zuvor als neue unabhängige Variablen als neue Spalte des Modellinputs bereit stellen.

24.03.2014 15

Abgabe per mail als pdf bis **15.04.** an:

Die Abgabe ist in Form eines Protokolls abzugeben in dem die Ergebnisse und die **wichtigsten** Berechnungsschritte (nicht unbedingt alle) klar ersichtlich sind.

hartner@eeg.tuwien.ac.at
Tel: +43(0)-1-58801-370379

Fragen??

24.03.2014 16