

Möglichkeiten einer
automatischen Modellselektion
von saisonalen Modellen
zur Prognose von Verbrauchsdaten

Lothar Nottekämper

Geschäftsführer
Senior-Consultant

- ◆ Wir sind
 - ◆ 8 Mitarbeiter mit unterschiedlichen Schwerpunkten

- ◆ Sitz
 - ◆ Marburg an der Lahn
 - ◆ Büro in Bad Vilbel

- ◆ Schwerpunkte

- ◆ Beratung beim Aufbau von
 - ◆ Statistischen Informationssystemen
 - ◆ Scoring-Anwendungen
 - ◆ Datamining

- ◆ Anwendungsentwicklung
 - ◆ Internet
 - ◆ Intranet
 - ◆ B-2-B

- ◆ Kunden aus dem Finanzdienstleistungsbereich



CoBaCash ist eine Software der
Commerzbank AG mit dem Ziel die
Versorgung und Entsorgung von
Geldausgabeautomaten und Kassen mit
Bargeld sicherzustellen

Partner



Implementierung
Basiskomponenten
Front-End
Prognose

Wechsel

Agenda

- ◆ Voruntersuchung
- ◆ Identifikation von Modellen
- ◆ Halbautomatische Strategie
- ◆ Implementierung
- ◆ Fazit

Agenda

- ◆ Voruntersuchung
- ◆ Identifikation von Modellen
- ◆ Halbautomatische Strategie
- ◆ Implementierung
- ◆ Fazit

Anforderungen

- ◆ Vielzahl von Zeitreihen müssen verarbeitet werden (ca. 2000)
- ◆ Eine individuelle Betrachtung und Beurteilung der einzelnen Zeitreihe ist nicht möglich
- ◆ Ein täglicher Update ist notwendig und erhöht die Genauigkeit
- ◆ Verarbeitungsgeschwindigkeit muss ausreichen
- ◆ Prognosen werden während der Nacht erneuert

- ◆ Entwicklungsdauer war kurz

Vorteile statistischer Methoden

- ◆ Der analytische Aufwand zur Ermittlung geeigneter Modelle hält sich in Grenzen
- ◆ Eine automatische Vorauswahl von Modellen ist anhand einfacher Kriterien durchführbar
- ◆ Eine automatische Schätzung der Modellparameter ist möglich
- ◆ Eine schnelle Verarbeitung von ca. 2000 Reihen im Batch-Betrieb kann implementiert werden
- ◆ Es stehen Verfahren für eine weitgehende automatische Modellauswahl zur Verfügung

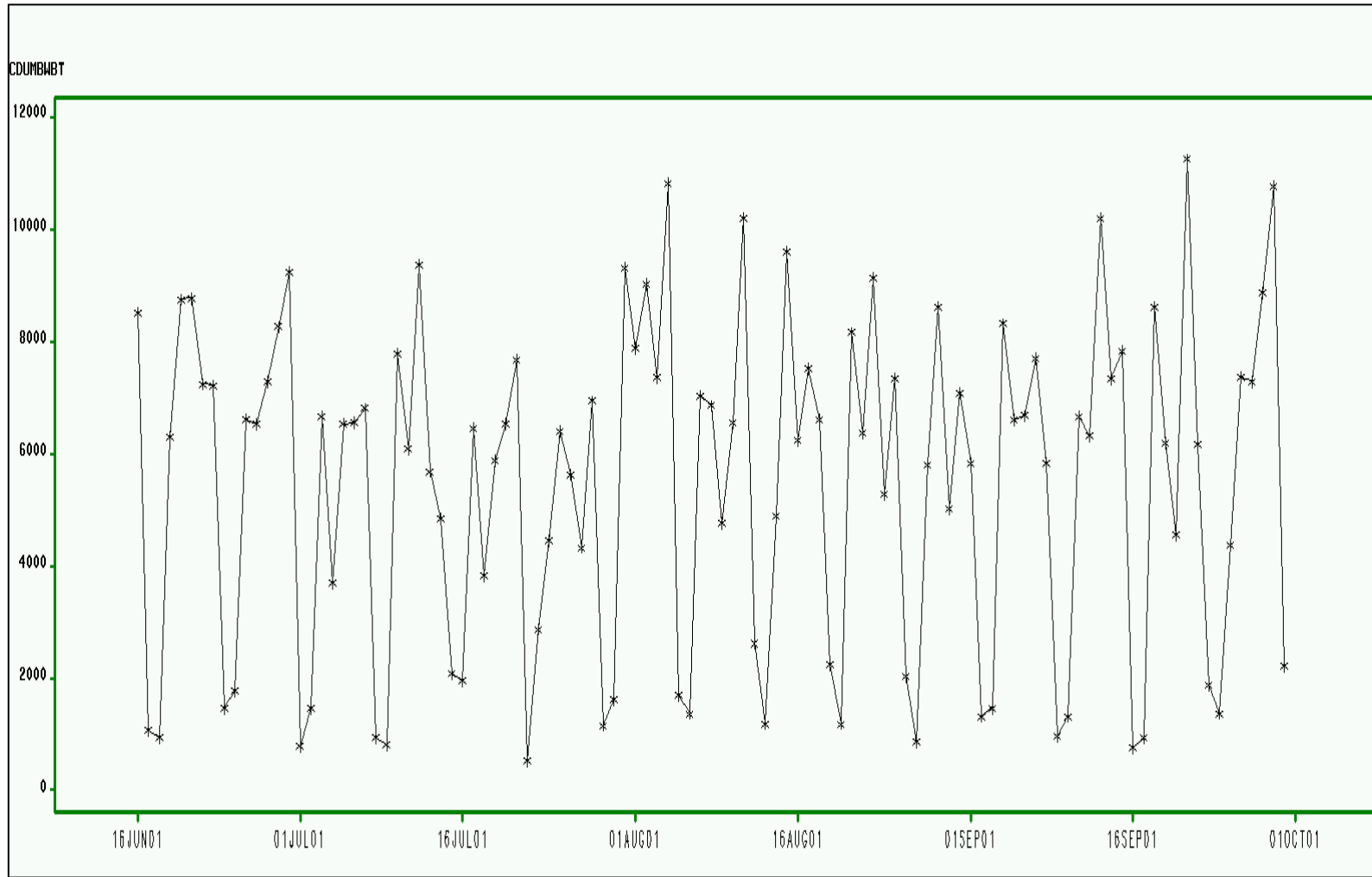
Voruntersuchung

- ◆ Ziel
 - ◆ Analyse der Zeitreihen
 - ◆ Vorauswahl statistischer Modelle

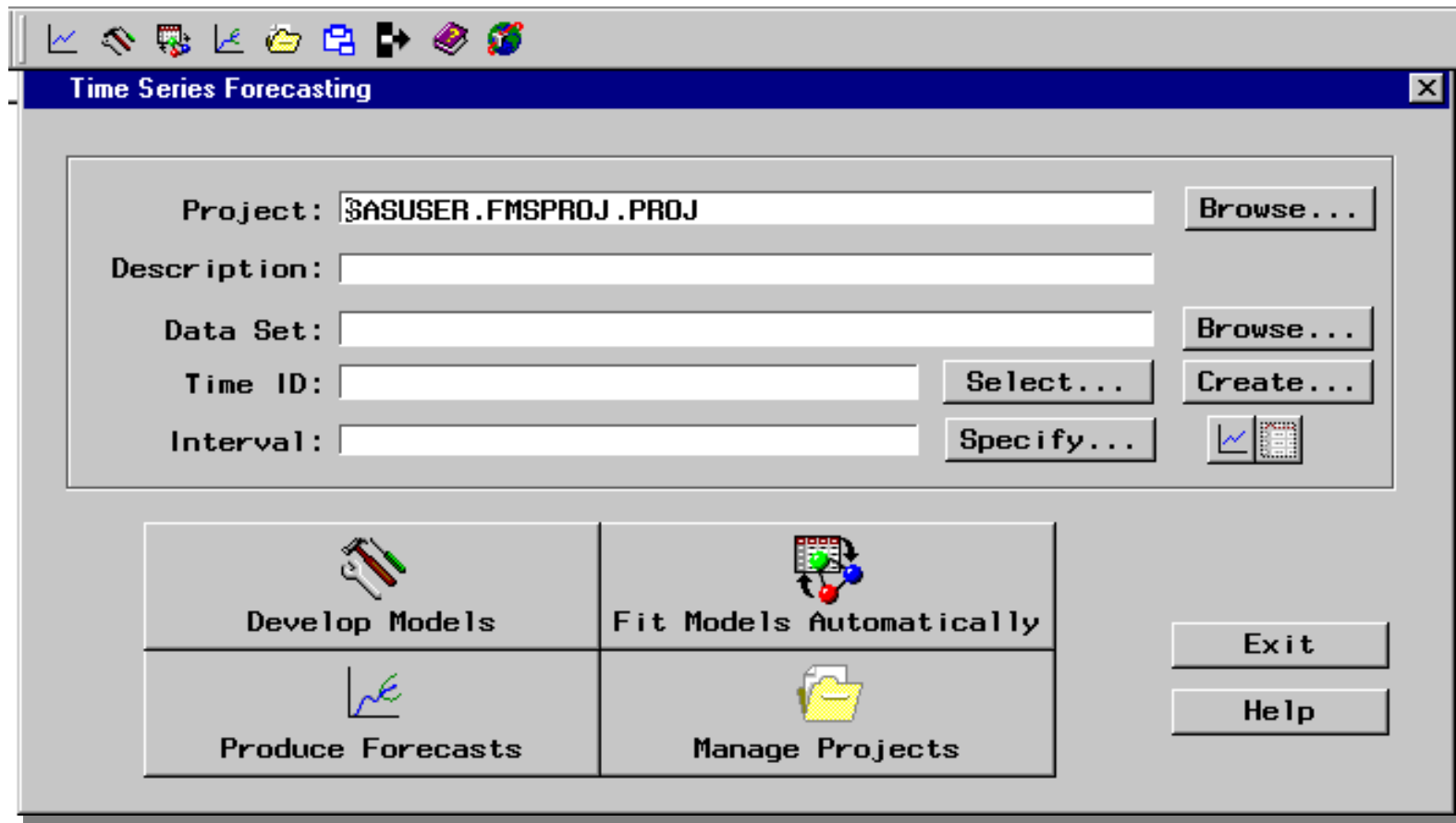
 - ◆ Stichprobe von Zeitreihe wurde gezogen
 - ◆ Intensive Untersuchung dieser Reihen

- ◆ Mittel
 - ◆ Forecast-Menü des SAS/Systems
 - ◆ Interaktive Oberfläche

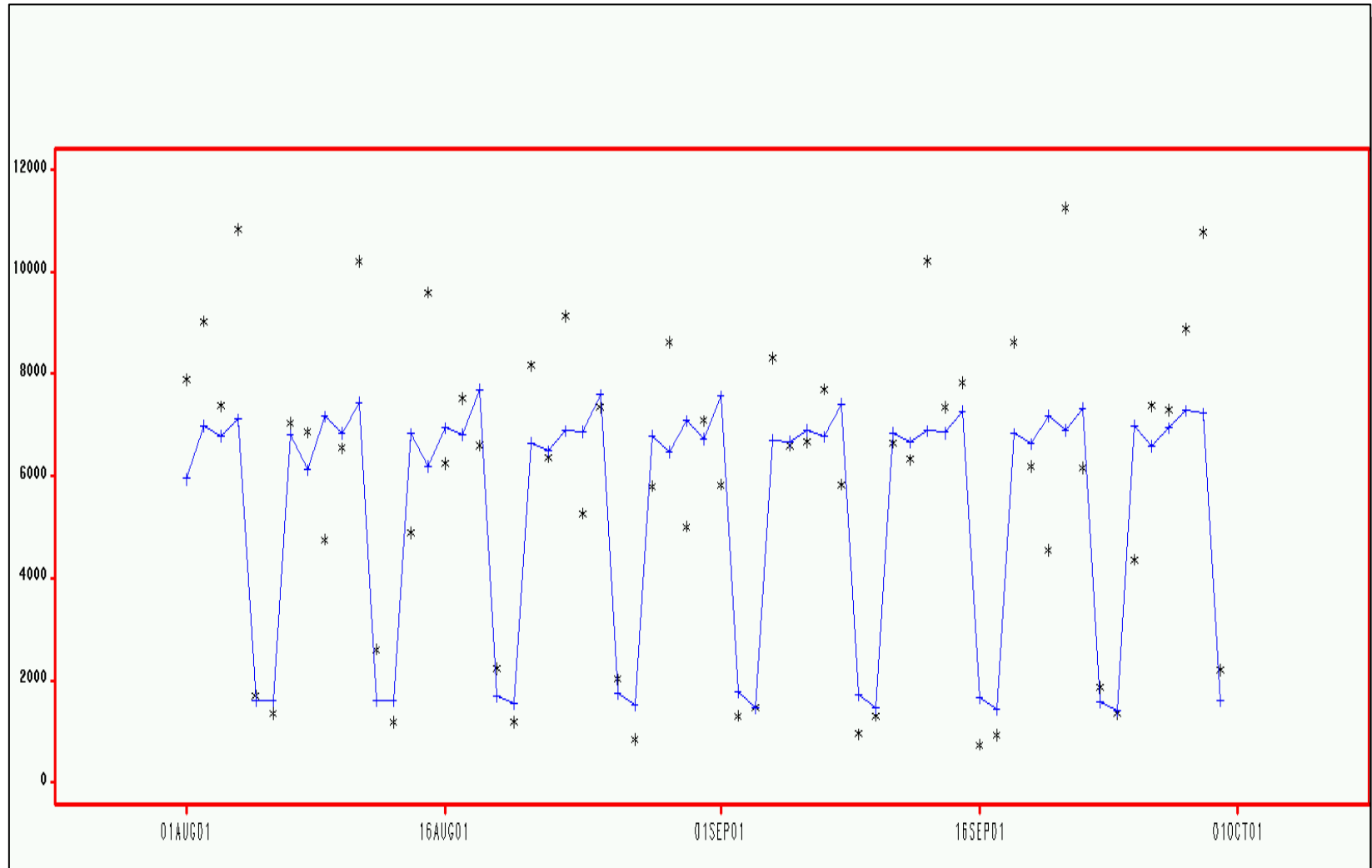
Zeitreihe



Time Series Forecasting Menü



Beispiel



Ergebnisse der Voruntersuchung

- ◆ Beantwortung grundlegender Fragen
 - ◆ Sind Voraussetzungen erfüllt?
 - ◆ Kann ein Trend in den Zeitreihen erkannt werden?
 - ◆ Weisen die Daten saisonale Effekte auf?
- ◆ Weitere Ergebnisse
 - ◆ Einfache Trends wurden nicht gefunden
 - ◆ Statistische Modelle mit saisonalen Parametern liefern die besten Werte bei den Anpassungsmaßen
 - ◆ Zeitreihen verhalten sich sehr unterschiedlich
 - ◆ Ausreißer müssen abgesichert werden

Agenda

- ◆ Voruntersuchung
- ◆ **Identifikation von Modellen**
- ◆ Halbautomatische Strategie
- ◆ Implementierung
- ◆ Fazit

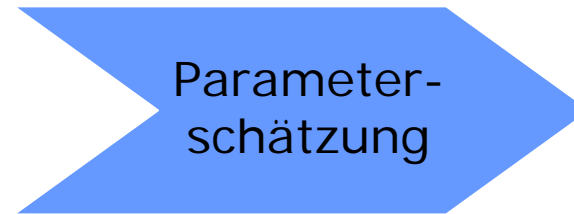
Detailuntersuchung einzelner Reihen

- ◆ Modellvarianten erkennen und festlegen
- ◆ Ausreißer-Analyse
- ◆ Klassifikation von Geräten nach Anpassungsmaßen

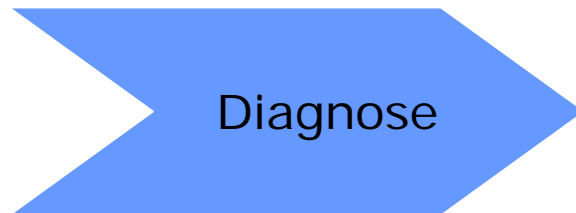
Identifikation von Statistischen Modellen



z.B. ARIMA (p,d,q)



$$X_t = \varepsilon_t - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}$$



Anpassung
Anzahl Parameter



Wochenzyklen

Spezifikation des geeigneten Modells

- ◆ Stationärer Prozess
- ◆ Autokorrelation (AR)
- ◆ Moving Average (MA)
- ◆ Gleitender Durchschnitt
- ◆ Exponentielle Glättung

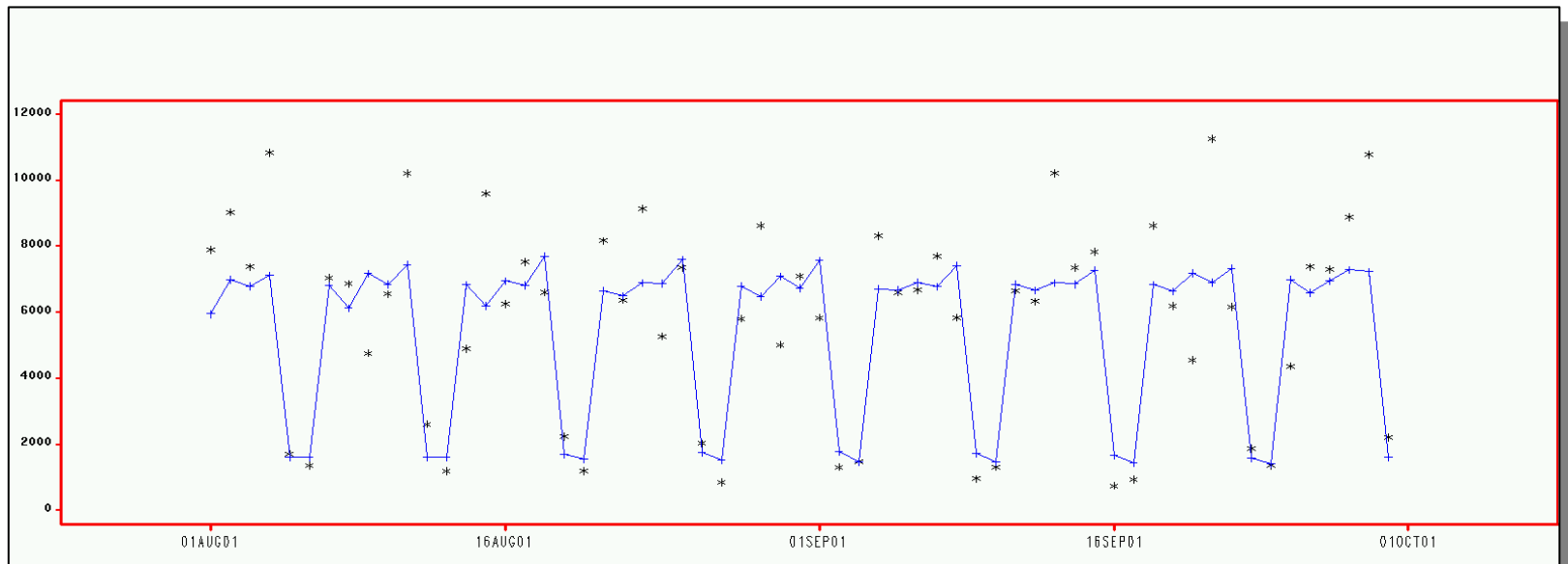
White-Noise-Test der Autokorrelationen

- ◆ White-Noise-Test der Autokorrelationen
 - ◆ Ist die Suche nach einem Modell überhaupt sinnvoll?
- ◆ Analyse steht am Anfang

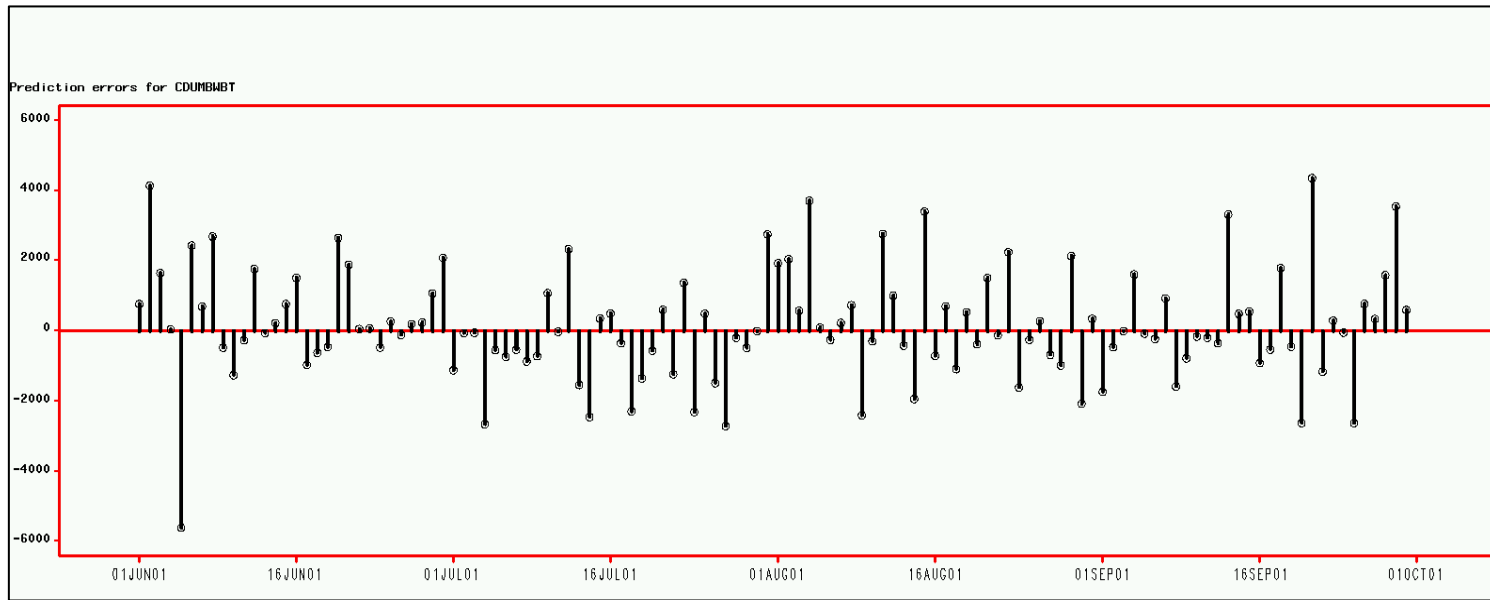
Stationäre Prozesse

- ◆ Prozesse bei denen sich der Mittelwert oder die Varianz mit der Zeit verändern
- ◆ Diagnose:
 - ◆ Visuelle Eigenschaften der Reihe
 - ◆ Statistischer Test
 - ◆ Autokorrelationen nehmen nur langsam ab
 - ◆ Formaler Test durch ein Regressionsmodell
- ◆ Abhilfe
 - ◆ Anwendung von Differenzfiltern
 - ◆ einfache
 - ◆ Saisonale
 - ◆ Transformation

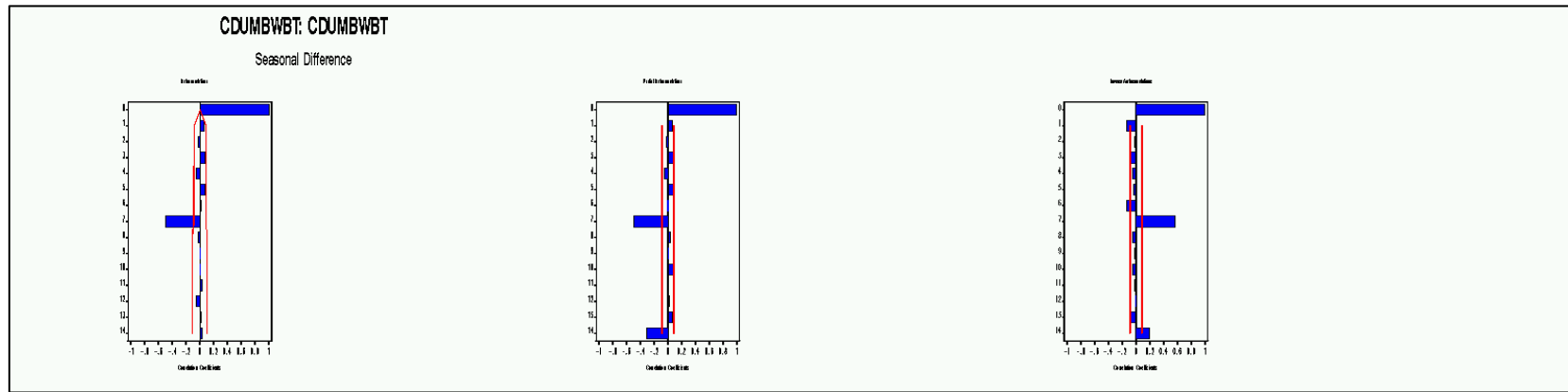
Kein Trend



Keine Varianzänderung



Autokorrelationsdiagramme



- ◆ Autokorrelationsfunktion (ACF)
- ◆ Partielle Autokorrelationsfunktion (PACF)
- ◆ Muster lassen auf ein bestimmtes Modell schließen

Spezifikationsproblem

- ◆ Einfache Modelle mit einem Parameter sind leicht zu erkennen
- ◆ Gemischte Modelle höherer Ordnung
 - ◆ Systematische Überspezifikation
 - ◆ Verbessert das Hinzufügen eines Parameters die Anpassung?
 - ◆ Wie entwickeln sich die anderen Parameter?
 - ◆ Es gibt keine einfachen eindeutigen Tests!

Saisonale Modelle

- ◆ Saisonale und faktorielle Modelle sind schwierig zu diagnostizieren
 - ◆ Kataloge für die Mustererkennung abgleichen mit Erfahrungen aus der Literatur
 - ◆ Intuition

Ergebnisse der Detailanalyse

- ◆ Zeitreihen weisen ausgeprägte saisonale Komponenten auf
- ◆ Wochenzyklen dominieren
- ◆ Unterschiede zwischen den Automaten sind groß
- ◆ Extreme Ausreißer treten selten auf, aber sie treten auf

- ◆ Die Detailanalyse ist aufwendig und kann nicht für jede Zeitreihe durchgeführt werden
- ◆ Die interaktive Oberfläche liefert keine Möglichkeit für den Batch-Betrieb
- ◆ Keine Information über zeitliche Stabilität

- ◆ Notwendigkeit für ein Modul zur automatischen Modellselektion

Agenda

- ◆ Voruntersuchung
- ◆ Identifikation von Modellen
- ◆ **Halbautomatische Strategie**
- ◆ Implementierung
- ◆ Fazit

Vorgehen

- ◆ Aus einer Vielzahl geschätzter Modelle wird dasjenige ausgewählt, welches ein vorgegebenes Gütekriterium erfüllt

Kriterien für den Modellvergleich

- ◆ Kriterien
 - ◆ Anpassungsmaße
 - ◆ Abweichungsquadrate sollten möglichst klein sein
 - ◆ R-Square möglichst groß
 - ◆ Prinzip der Sparsamkeit
 - ◆ Anzahl der Parameter sollten möglichst klein bleiben
 - ◆ Sparsamkeit soll belohnt werden
 - ◆ Überspezifikation führt zu schlechteren Prognoseergebnissen
 - ◆ Akaike Informationskriterium (AIC)
 - ◆ u.a.

Anforderungen an das Modul

- ◆ Modellvielfalt abdecken
- ◆ Modellvarianten erfassen
- ◆ Kriterien der automatischen Selektion anwenden
- ◆ Parameterschätzung
- ◆ Anpassungsmaße speichern
- ◆ Historie aufbauen
- ◆ Batch-Betrieb
- ◆ Performance

- ◆ Prognose erstellen

Modellvielfalt

- ◆ Modellvielfalt sollte abgedeckt werden können
 - ◆ Datentransformation
 - ◆ Log-Transformation
 - ◆ Ausreißertest
 - ◆ Modelltyp
 - ◆ ARIMA
 - ◆ Exponentielle Glättung
 - ◆ ...
 - ◆ Modelle mit Differenzbildung
 - ◆ Ordnung der Autoregression Parameter
 - ◆ Ordnung der Moving-Average Parameter
 - ◆ Intercept

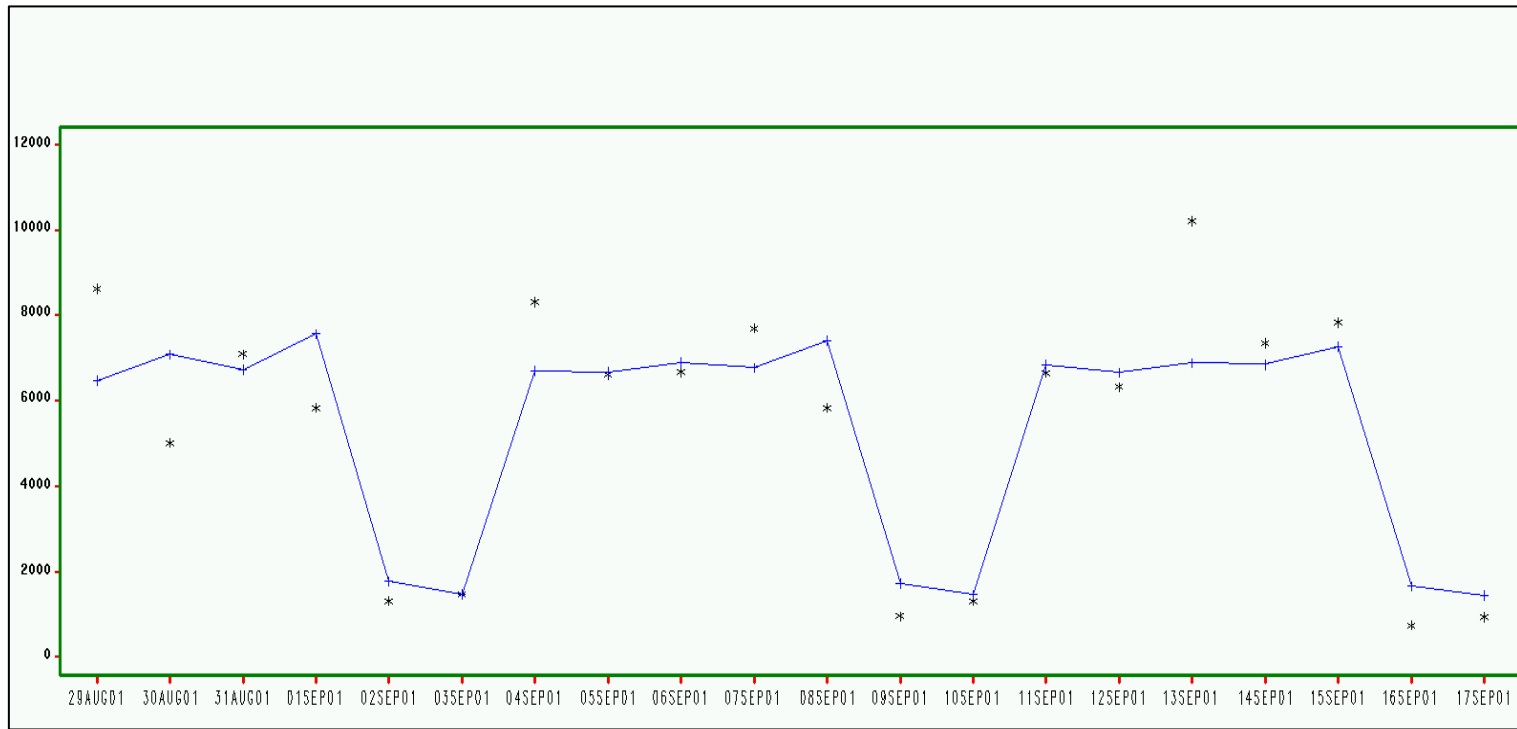
Modellerfassung

- ◆ Detailuntersuchung -> Hinweise auf relevante Modellparameter
- ◆ Notwendige Modellvielfalt wurde an Stichproben validiert und eingegrenzt
- ◆ Einfache Tabelle mit Feldern für die Merkmale des Modells
- ◆ So gestaltet, das ein Prozeduraufruf daraus generiert werden kann

Ablauf Erstellung einer Prognose

- ◆ Parameter für alle erfassten Modelle werden geschätzt
- ◆ Ein Modellvergleich nach dem eingestellten Vergleichskriterium wird durchgeführt
 - ◆ Das Beste gespeichert alle anderen verworfen
- ◆ Mit dem besten Modell wird eine Prognose erstellt
- ◆ Tabelle mit den Gütekriterien und allen Durchläufen wird gespeichert
- ◆ Die Modellvarianten mit den meisten Hits können in der Verarbeitung verbleiben

Bedarfsprognose



Entscheidungs- und Zielkonflikt

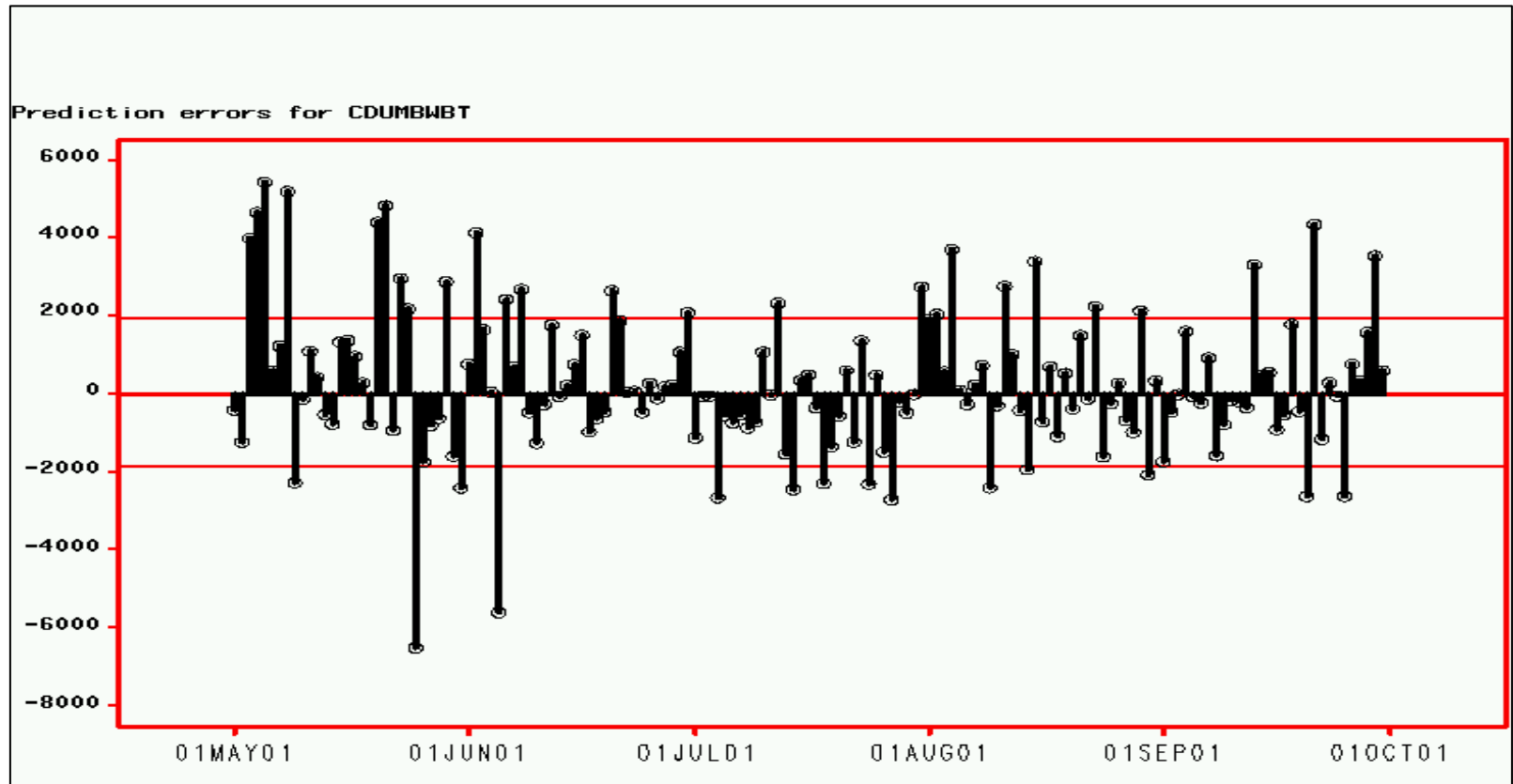
- ◆ 1. Ziel
 - ◆ Minimierung der Füllmengen
 - ◆ Reduzierung unnötiger Fahrten

- ◆ 2. Ziel
 - ◆ Vermeidung von Leerstand
 - ◆ Imageverlust

Entscheidungsproblem

- ◆ Eine zu geringe Bedarfsprognose führt zu einem erhöhten Leerstandsrisiko
- ◆ Einseitiges Problem
- ◆ Unterschätzungen sind gravierender und führen zum Leerstand
- ◆ Eine mögliche Lösung:
 - ◆ Bedarfsprognosen werden systematisch erhöht
 - ◆ Einseitiges Konfidenzintervall wird herangezogen

Entscheidungshilfe Konfidenzintervalle



Entscheidung

- ◆ Obere Grenze des Konfidenzintervalls
- ◆ Je größer das Konfidenzintervall
 - ◆ desto geringer wird das Leerstandsrisiko
 - ◆ Desto geringer aber auch der Einspareffekt

Agenda

- ◆ Voruntersuchung
- ◆ Identifikation von Modellen
- ◆ Halbautomatische Strategie
- ◆ **Implementierung**
- ◆ Fazit

Wechsel

Agenda

- ◆ Voruntersuchung
- ◆ Identifikation von Modellen
- ◆ Halbautomatische Strategie
- ◆ Implementierung
- ◆ Fazit

Fazit

- ◆ Informationen aus Verbrauchsdaten
 - ◆ Lassen sich durch statistische Zeitreihenanalysen zu Optimierungsaufgaben heranziehen
 - ◆ Verarbeitung im Batch erlaubt eine halbautomatische Selektion geeigneter Modelle
 - ◆ Die Anwendungsintegration sorgt für eine optimale Nutzung zur Entscheidungsunterstützung

Implementierung: Was ist zu tun?

Zeitreihen	<ul style="list-style-type: none">◆ Mindestens 12 Wochen◆ Schnittstellen zu liefernden Systemen herstellen◆ Daten bereinigen und integrieren◆ DB-Tabellen anlegen
Voruntersuchung	<ul style="list-style-type: none">◆ Erste Übersicht möglicher Modelle
Detailuntersuchung	<ul style="list-style-type: none">◆ Intensive Beurteilung der Modellvarianten
Implementierung	<ul style="list-style-type: none">◆ Erfassung der Modellvarianten
Kontrolle	<ul style="list-style-type: none">◆ Gespeicherte Kennwerte erlauben die Kontrolle der Modellauswahl während des laufenden Betriebs◆ Sicherheitsfaktoren vermeiden den Leerstand
Infrastruktur	<ul style="list-style-type: none">◆ Serveranschaffung◆ Installation SAS/System◆ SQL-DB◆ Scheduler