

# Seminarkatalog Angewandte Statistik

## Allgemeine Hinweise

Der Seminarkatalog bietet Kurzbeschreibungen der Themen des Seminars. Die Pflichtliteratur zu jedem Seminarthema besteht aus einem Kapitel aus einem deutschsprachigen Lehrbuch und einem englischsprachigen Artikel aus einer internationalen Fachzeitschrift. Darüber hinaus kann es weitere Literaturhinweise geben.

In der Seminararbeit soll neben der theoretischen Ausarbeitung auch eine empirische Anwendung des Seminarthemas durchgeführt werden. Dazu empfehlen wir die Nutzung der Programmiersprache R.

## Makroökonomische Zeitreihenanalyse und Prognose

### Makroökonomische Prognosen in einer *Big Data* Umgebung

Zur Prognose makroökonomischer Zeitreihen steht häufig eine große Anzahl potenzieller Regressoren zur Verfügung. Die Aufnahme aller Variablen in ein einzelnes Modell ist aufgrund der Fülle an Variablen und der in der Regel limitierten Anzahl an vorliegenden Beobachtungen nicht sinnvoll, da die Parameterschätzung instabil wird oder im Extremfall nicht identifiziert ist. Einen Lösungsansatz, mit hochdimensionalen Datensätzen umzugehen, bieten Faktormodelle. Hierbei werden die Variablen zu einer Handvoll Faktoren zusammengefasst. Diese Faktoren können dann beispielsweise in einem Prognosemodell verwendet werden. Das Deutsche Institut für Wirtschaftsforschung in Berlin verwendet einen solchen Ansatz, um ein monatliches Konjunkturbarometer zu berechnen. Die Seminararbeit soll das statistische Faktor Modell sowie seine Schätzung über die Hauptkomponentenmethode vorstellen und zur Konjunkturprognose verwenden.

Lehrbücher:

- Fahrmeir et al. (1996) (Kap. 11.1–11.3)
- Härdle and Simar (2015) (Kap. 11–12)

Artikel:

- Stock and Watson (2002)

Sonstiges:

- Konjunkturbarometer des DIW Berlin:  
[https://www.diw.de/de/diw\\_01.c.623308.de/seiten\\_redakteure/konjunkturprognosen.html](https://www.diw.de/de/diw_01.c.623308.de/seiten_redakteure/konjunkturprognosen.html)

## **Gemeinsam genauer? Wie lassen sich Prognosen kombinieren?**

Viele empirische Studien zeigen, dass die Kombination von Forecasts sogar genauere Ergebnisse liefern kann als der beste individuelle Forecast. Zudem scheinen einfache Kombinationsmethoden erfolgreicher zu sein als ausgefeilte Methoden zur Wahl der Kombinationsgewichte. Die Seminararbeit soll in die Methoden zur Prognosekombination einführen und empirisch überprüfen, ob ein kombinierter *Forecast aus dem Survey of Professional Forecasters* tatsächlich bessere Ergebnisse liefert als die einzelnen Prognosen.

Lehrbücher:

- Döhrn (2014) (Kap. 9)
- Timmermann (2006)

Artikel:

- Kenny et al. (2010)
- Stock and Watson (2004)

## **Umgang mit Daten gemischter Frequenzen**

In der Praxis liegen viele Zeitreihen in gemischten Frequenzen vor. Das BIP wird quartalsweise erhoben, während Arbeitslosenzahlen monatlich ausgewiesen werden. Finanzmarktdaten stehen sogar minütlich zur Verfügung. Die Nutzung von Daten unterschiedlicher Frequenzen in einem Modell erfordert besondere Ansätze, wie zum Beispiel die *Mixed-Data Sampling* (MiDaS) Methode. Diese Methoden sollen vorgestellt und zur Prognose des BIP auf Basis von monatlichen Daten angewandt werden.

Lehrbücher:

- Wohlrabe (2009) (Kap. 2)

Artikel:

- Foroni and Marcellino (2014)

## **Der *Leading Indicator Index* zur Konjunkturprognose**

Das Conference Board veröffentlicht monatlich einen *Leading Indicator Index*, der als frühzeitiger Konjunkturindikator einen Ausblick auf die gesamtwirtschaftliche Entwicklung in den USA geben soll. Dieser setzt sich aus verschiedenen Variablen zusammen. Die Güte dieses Indikators soll überprüft und es soll analysiert werden, wie ein ähnlicher Indikator für Deutschland aussehen könnte.

Lehrbücher:

- Döhrn (2014) (Kap. 5 und 12)

Sonstiges:

- Der *Leading Indicator Index* des *Conference Boards*:  
<https://www.conference-board.org/data/bciarchive.cfm?cid=1>

### **Multivariate Zeitreihenanalyse mit Vektorautoregressiven Modellen**

Vektorautoregressive (VAR) Modelle ermöglichen es, die Dynamiken mehrerer univariater Zeitreihen simultan zu modellieren und Wirkungszusammenhänge einer Variablen zu verzögerten eigenen Werten und zu verzögerten Werten der anderen Variablen des Modells zu untersuchen. Somit können beispielsweise „Ansteckungs-“ und „*Spillover*-Effekte“ identifiziert und häufig bessere Vorhersagen als mit univariaten autoregressiven Modellen getroffen werden.

Im Rahmen der Arbeit sollen VAR-Modelle vorgestellt und deren Verwendung anhand eines Beispiels wie der Prognose des Wachstums des Bruttoinlandprodukts illustriert werden.

Lehrbücher:

- Neusser (2011) (Kap. 13 und 14)
- Tsay (2013) (Kap. 2)
- Canova (2011) (Kap. 4 und 10)

Artikel:

### **Zustandsraummodelle und Kalman-Filter**

Zeitreihenmodelle können im Allgemeinen auch in einem sogenannten Zustandsraum dargestellt werden. Ein Zustandsraummodell wird durch zwei Gleichungen beschrieben, die Zustands- und die Messgleichung. Sämtliche Beziehungen der beobachtbaren Daten und der latenten, nicht direkt beobachtbaren Prozesse werden in Form von Vektoren und Matrizen modelliert. Der Kalman-Filter ist ein Algorithmus zur rekursiven Schätzung und Prognose des latenten Prozesses.

Im Rahmen der Arbeit sollen Zustandsraummodelle vorgestellt werden und ein illustrativer latenter Prozess mithilfe des Kalman-Filters in einem linearen Gaußschen Modell geschätzt werden. Anwendungsbeispiele in der Ökonomie sind die natürliche Zinsrate oder der Konjunkturzyklus.

Lehrbücher:

- Schlittgen (2015) (Kap. 10)
- Hamilton (1994) (Kap. 13)

- Durbin and Koopman (2012) (Kap. 3-7)

Artikel:

- Holston et al. (2017)

### **Kointegration**

In der Realität entwickeln sich die meisten ökonomischen Zeitreihen wegen einer Trendkomponente keineswegs immer stationär. Das Problem bei der Schätzung mit solchen Zeitreihen ist, dass es sehr schnell zu Scheinregressionen und damit zu falschen Schlussfolgerungen kommen kann. Eine Möglichkeit zur Lösung dieser Problematik ist es, mittels einer Trendbereinigung stationäre Zeitreihen zu erhalten. Der Nachteil dieser Methode ist der dabei auftretende Informationsverlust. Das Kointegrationsverfahren bietet sich bei der Analyse von nicht stationären Zeitreihen an, da mit dieser Methode das oben beschriebene Problem vermieden werden kann. Die vorhandenen statistischen Gleichgewichtsbeziehungen zwischen nicht stationären Variablen werden korrekt erfasst und analysiert.

In der Seminararbeit soll die Kointegrationsanalyse vorgestellt und anhand von Beispielen aus der Praxis in der Arbeit erläutert werden. Klassische Anwendungsbeispiele sind der Zusammenhang zwischen kurz- und mittelfristigen Zinssätzen oder dem Öl- und US-Dollarpreis.

Lehrbücher:

- Assenmacher (2010) (Kap. 15)
- Kirchgässner and Jürgen (2006) (Kap. 6)
- Wollschläger (2017)

Artikel:

- Engle and Granger (1987)

## **Finanzmarktökonometrie und empirische Kapitalmarktforschung**

### **Modellierung von Volatilitäten auf Finanzmärkten**

Hinter dem Begriff der Volatilität verbirgt sich die Schwankung eines Finanztitels um seinen langfristigen Mittelwert und somit das Risiko, welches mit dieser Anlage einhergeht. Die Messung der Volatilität liegt daher im Zentrum der Bewertung von verschiedenen Anlageformen, wie z.B. Aktienkursindizes, und wird in eigenen Volatilitätsindizes, wie z.B. dem *CBOE Volatility Index* (VIX), abgebildet.

Lehrbücher:

- Shiryayev (1999) (Kap. 4.3.a)
- Franke et al. (2012) (Kap. 6.3.4)
- Alexander (2008) (Kap. II.3)
- Brooks (2019) (Kap. 9.1-9.7)

### **Langfristige Änderungen in Finanzmarktrisiken**

Stilisierte Fakten, wie die leptokurtische Verteilung und die Clustereigenschaft der Renditen von Finanzzeitreihen können mit klassischen Modellen der Zeitreihenanalyse nicht erklärt werden. Abhilfe können hier Modelle der sogenannten *Autoregressive Conditional Heteroskedastic* (ARCH) Modellklasse schaffen. Anwendungsbeispiele sind z.B. Aktienkursindizes wie der DAX.

Lehrbücher:

- Bollerslev et al. (1994)
- Cryer and Chan (2008) (Kap. 12)
- Schröder (2012) (Kap. 6)
- Martin et al. (2012) (Kap. 20)

Artikel:

- Engle (1982)
- Bollerslev (1986)

### **Portfoliotheorie und Tobin-Separation**

Die Portfolio Theorie nach Markowitz gibt Antworten auf wesentlichen Fragen der Portfoliostrukturierung und gilt auch heute als gebräuchlichstes Instrument zur Bestimmung der Anlagekategorien. Sie unterstellt eine Risikoaversion der Investoren und analysiert Anlagesituationen vor dem Hintergrund ihrer erwarteten Renditen (Erwartungswert) und ihres Risikos (gemessen anhand der Standardabweichung). Die Grundidee der Portfoliotheorie ist die Erzielung einer maximalen Rendite bei geringerem Risiko. Die Tobin-Separation ist ein Konzept aus der Kapitalmarkttheorie. Sie befasst sich mit der Bestimmung des optimalen Portfolios, bei der Wertpapiere in zwei Gruppen unterteilt werden: risikofreie und riskante Wertepapiere. Dabei spielt die Risikoeinstellung des Investors keine Rolle.

Im Rahmen der Seminararbeit sollen die Portfoliotheorie und die Tobin Separation ausführlich dargestellt und anhand von einfachen praktischen Beispielen erklärt werden.

Lehrbücher:

- Spremann (2014) (Kap. 7-8)

Artikel:

- Markowitz (1952)

### **Das Capital Asset Pricing Model**

Das *Capital Asset Pricing Model* (CAPM) ist eine Erweiterung der klassischen Portfoliotheorie nach Markowitz. Hier wird die erwartete Rendite eines Wertpapiers als eine lineare Funktion des Portfoliorisikos dargestellt. Heutzutage wird das Modell zur Aktien- und Unternehmensbewertung sowie bei der Vermögensstrukturierung eingesetzt.

Im Rahmen der Seminararbeit soll das CAPM inklusive seiner Herleitungen dargestellt und anhand von einfachen Beispielen aus der Praxis erklärt werden.

Lehrbücher:

- Spremann (2014) (Kap. 10)

Artikel:

- Sharpe (1964)

### **Prognose populärer Risikomaße**

*Value-at-Risk* (VaR) und *Expected Shortfall* (ES) sind zwei wichtige Risikomaße an den Finanzmärkten, die auf den Regelungen von Basel basieren. Beide Maßzahlen werden intensiv von Finanzinstituten und Investoren weltweit genutzt. VaR und ES geben den größten potenziellen Verlust ausgehend von einem bestimmten Niveau an. Eine Prognose des VaR und ES basiert auf historische Daten und kann Finanzinstituten helfen, ihre Investmentportfolios zu verbessern. Volatilitätsmodelle, wie z.B. das *Generalized Autoregressive Conditional Heteroskedastic* (GARCH) Modell, spielen dabei eine wichtige Rolle.

Im Rahm der Arbeit sollen VaR und ES anhand von empirischen Aktienmarktdaten prognostiziert werden. Mit einem Backtesting Ansatz lässt sich überprüfen, ob die Prognosen nah an den realisierten Werten liegen. Einen großen Einfluss auf die Prognoseperformance hat das verwendete Modell. Man erhält akkuratere Prognosen durch eine verbesserte Volatilitätsprognose.

Lehrbücher:

- Franke et al. (2012) (Kap. 15)

- Christoffersen (2011) (Kap. 2)

Artikel:

- Engle and Manganelli (2001)

### **Das Stochastische Volatilitätsmodell**

Neben dem bekannten *Generalized Autoregressive Conditional Heteroskedastic* (GARCH) Modell, das im Bereich der Finanzmarktökonomie zur Volatilitätsschätzung verwendet wird, bietet das Stochastische Volatilitätsmodell (SV Modell) einen anderen Ansatz, der eine größere Flexibilität ermöglicht. In dieser Arbeit sollen die analytischen und empirischen Aspekte des SV Modells erläutert und eine Anwendung des Modells auf Aktienmarktdaten durchgeführt werden.

Lehrbücher:

- Franke et al. (2012) (Kap. 12)
- Christoffersen (2011) (Kap. 4)
- Tsay (2010) (Kap. 3)

Artikel:

- Harvey et al. (1994)

## **Statistische Modelle und Parameterschätzung**

### **Lineare Regression und Variablenauswahl**

Die lineare Regression ist eines der wichtigsten Instrumente zur Analyse von linearen Beziehungen zwischen einer zu erklärenden Variablen und potenziell vielen erklärenden Variablen. Eine entscheidende Frage ist hierbei, welche Variablen zur Erklärung im Modell mit einbezogen werden sollen und welche irrelevant sind. Zudem ergibt sich bei hochdimensionalen Datensätzen oft die Situation, dass mehr Variablen als Beobachtungen zur Verfügung stehen. Damit wäre es nicht möglich eine lineare Regression durchzuführen und entsprechende Vorselektion der Variablen ist notwendig. Verschiedene Methoden zur Variablenselektion (*forward stepwise, backward stepwise, forward stagewise, least angle regression, ...*) sollen präsentiert und in einer Anwendung miteinander verglichen werden. Diese Methoden finden in vielen Bereichen Anwendung. Beispielsweise können sie genutzt werden, um aus der Fülle ökonomischer Kennziffern, diejenigen zu filtern, die gemeinsam eine aussagekräftige Konjunkturprognose bieten.

Lehrbücher:

- Fahrmeir et al. (2009)
- Hastie et al. (2013)

Artikel:

- Bai and Ng (2008)

### **Statistische Schätzprinzipien: Die Maximum-Likelihood-Schätzung**

Die Maximum-Likelihood Methode ist ein Schätzverfahren, bei dem derjenige Wert für einen Parameter ausgewählt wird, welcher gemäß der Verteilung der realisierten Beobachtungen am plausibelsten erscheint. Die Maximum-Likelihood-Methode wird beispielsweise zur Schätzung von Probit- und Logit-Modellen verwendet. Diese Modelle finden Anwendung, wenn die zu erklärende Zielgröße eine binäre Variable ist, die nur zwei Werte annehmen kann. Beispielsweise lässt sich mit diesen die Kaufentscheidung (Ja/Nein) eines Kunden für ein bestimmtes Produkt modellieren. Im makroökonomischen Kontext lassen sich mit Probit- und Logit-Modellen beispielweise der Einfluss von ökonomischen Indikatoren auf das Eintreten einer Rezession abbilden.

Lehrbücher:

- Assenmacher (2010) (Kap. 10)
- Fahrmeir et al. (2009) (Kap. 4)
- Verbeek (2017) (Kap. 6.1 und 7.1)

Artikel:

- Estrella and Mishkin (1998)

### **Analyse von Zeitreihen mit linearen Modellen**

Das *Autoregressive Integrated Moving Average* (ARIMA) Modell ist der klassische Ansatz der modernen Zeitreihenanalyse. Es umfasst sowohl autoregressive wie auch Moving-Average-Prozesse. Darüber hinaus können nichtstationäre (integrierte) Zeitreihen durch Differenzbildung behandelt werden. Die Bestimmung der Modellordnung erfolgt mit der Box-Jenkins-Methode oder mit Hilfe von Informationskriterien.

Im Rahmen der Seminararbeit soll die Arbeitslosenquote in Deutschland auf Basis eines geeigneten ARIMA-Modells modelliert und prognostiziert werden.

Lehrbücher:

- Franke et al. (2012) (Kap. 11)



- Schlittgen (2015) (Kap. 3)
- Neusser (2016) (Kap. 2 und 5)
- Tsay (2014) (Kap. 2 und 3)

### **Markov-Chain-Monte-Carlo Verfahren**

*Markov-Chain-Monte-Carlo* (MCMC) Verfahren bilden eine Klasse von Algorithmen, die (autokorrelierte) Stichproben aus Wahrscheinlichkeitsverteilungen ziehen. Hierfür wird eine Markov-Kette konstruiert, welche die erwünschte Verteilung als ihre stationäre Verteilung aufweist.

In dieser Arbeit soll ein Random-Walk Metropolis Algorithmus für eine zwei-dimensionale Normal-Inverse Gamma Verteilung programmiert und auf empirische oder simulierte Daten angewendet werden. Unterschiedliche Startwerte und verschiedene Vorschlagsdichten mit unterschiedlichen Akzeptanzraten sollen erprobt und die Resultate im Hinblick auf Autokorrelation und Konvergenzverhalten beschrieben werden.

Lehrbücher:

- Robert and Casella (2013) (Kap. 7)
- Givens and Hoeting (2012) (Kap. 7 und 8)

Artikel:

- Hastings (1970)

### **Periodizität und Spektralanalyse**

Wiederkehrende Muster und insbesondere Zyklen von Zeitreihen können im Frequenzbereich untersucht werden. Hierzu werden erweiterte Methoden und Konzepte, wie z.B. Periodogramme und Fourierfrequenzen verwendet. Die reichhaltige Analyse bietet u.a. Tests auf periodisches Verhalten und Autokorrelation. Ein weiteres wichtiges Thema ist die empirische Schätzung der Spektraldichte.

Lehrbücher:

- Schlittgen (2015) (Kap. 6)
- Hassler (2007) (Kap. 4)
- Neusser (2016) (Kap. 6)
- Shumway and Stoffer (2017) (Kap. 4)

Artikel:

- Uebele and Ritschl (2009)

## Statistisches und maschinelles Lernen

### Boosting

*Boosting* gilt als eine der erfolgreichsten Ideen im Bereich des maschinellen Lernens der letzten 30 Jahre. Der *Boosting*-Algorithmus schätzt eine unbekannte Funktion schrittweise, indem in jeder Iteration die Variable mit dem höchsten Erklärungsgehalt dem Modell hinzugefügt wird. Dadurch werden viele einzelne schwache Modelle zu einem starken kombiniert. *Boosting* kann somit auch als Methode zur Variablenselektion verstanden werden. Die Seminararbeit soll einen Überblick über die Methodik geben und den Algorithmus im Rahmen eines Regressions- oder Klassifikationsproblems anwenden, wie zum Beispiel der Modellierung und Prognose von Immobilienpreisen auf Basis einer Vielzahl an Determinanten.

Lehrbücher:

- Richter (2019) (Kap. 6.5)
- Hastie et al. (2013) (Kap. 10)

Artikel:

- Bai and Ng (2009)
- Buchen and Wohlrabe (2011)

### Regularisierungsmethoden

Im Big Data Kontext steht in der Regel eine große Anzahl an Variablen zur Verfügung. Eine unbeschränkte Aufnahme aller Variablen in ein zu schätzendes Modell (z.B. lineares Regressionsmodell) führt häufig zu einer Überanpassung des Modells an die Daten, sodass die Generalisierbarkeit des Modells auf neue, unbekannte Daten leidet. Regularisierungsmethoden (engl. *shrinkage methods*), wie z.B. der *Least Absolute Shrinkage and Selection Operator* (LASSO), bieten eine Möglichkeit ein Überanpassung an die Daten zu vermeiden. Die Anwendungsgebiete dieser Methoden sind äußerst vielfältig. Beispielsweise können sie zur Konjunkturprognose genutzt werden, um eine hohe Anzahl an Einflussfaktoren im Model zu erfassen. Ein weiteres Anwendungsbeispiel bietet die Modellierung und Prognose von Immobilienpreisen auf Basis einer Vielzahl an Determinanten.

Lehrbücher:

- Richter (2019) (Kap. 2)
- Hastie et al. (2013) (Kap. 3.4)

Artikel:

- Bai and Ng (2008)

## Random Forests

*Random Forests* bilden eine Klasse von nichtparametrischen Algorithmen. Durch ihre schnelle Berechenbarkeit und die darauf aufbauenden Erweiterungen sind die Bäume in der Praxis sehr beliebt.

Ziel dieser Seminararbeit ist eine Einführung in die Grundlagen der Random Forests im Kontext der Regression. Beispielhaft soll ein Modell implementiert und zur Schätzung von Immobilienpreisen verwendet werden.

Lehrbücher:

- Richter (2019) (Kap. 6)
- James et al. (2013) (Kap. 8.2)

Artikel:

- Liaw and Wiener (2001)

## Neuronale Netze

Neuronale Netze bilden eine flexible Funktionenklasse zur Approximation von stetigen Funktionen. Der Aufbau und die Funktionsauswertung eines neuronalen Netzes ist anschaulich der Signalübertragung von Nervenzellen nachempfunden. Wie Bäume erlauben neuronale Netzwerke sowohl die Behandlung von Regressions- als auch Klassifikationsproblemen.

Ziel der Seminararbeit ist die Einführung in die Grundlagen der neuronalen Netze. Beispielhaft soll ein Modell implementiert und zur Klassifikation der Kreditwürdigkeit von Kunden verwendet werden.

Lehrbücher:

- Richter (2019) (Kap. 7)
- Murphy (2012) (Kap. 28)

Artikel:

- Munkhdalai et al. (2018)

## Support Vector Machines

Eine *Support Vector Machine* unterteilt eine Menge von Objekten in Klassen, sodass um die Klassengrenzen herum ein möglichst breiter Bereich frei von Objekten bleibt. Sie dienen zur Erkennung von Mustern in großen Datensätzen.

Ziel der Arbeit ist es, eben diesen Klassifikator vorzustellen und zur Klassifikation der Kreditwürdigkeit von Kunden anzuwenden.

Lehrbücher:

- Richter (2019) (Kap. 4)
- James et al. (2013) (Kap. 9.1)
- Murphy (2012) (Kap. 14.5)

Artikel:

- Evgeniou and Pontil (2001)

### **Unüberwachtes Lernen: Der Erwartungs-Maximierungs-Algorithmus**

Clusteralgorithmen bilden Verfahren zur Erkundung von Mustern in großen Datensätzen (*Big Data*). Der Erwartungs-Maximierungs-Algorithmus (EM-Algorithmus) ist im Gegensatz zu den traditionellen Vorgehensweisen ein modellbasierter Ansatz, d.h. Objekte können mit einer gewissen Wahrscheinlichkeit zu einem Cluster zugeordnet werden. Diese Verfahren werden oft für Marketingzwecke eingesetzt, um gezielte Werbung für einzelne Kundengruppen zu schalten.

Ziel dieser Arbeit ist eine Einführung in die Funktionsweise des EM-Algorithmus. Beispielhaft soll ein Modell implementiert und zur Klassifikation von Kunden eines E-Shops verwendet werden.

Lehrbücher:

- Richter (2019)
- James et al. (2013)
- Murphy (2012) (Kap. 11.4)
- Fahrmeir et al. (1996) (Kap. 9)

Artikel:

- College and Dellaert (2002)

### **Statistische Test- und Vergleichsverfahren**

#### ***Resampling*-Verfahren**

In vielen Testsituationen ist die Verwendung kritischer Werte aus der Grenzverteilung (z.B. der Standardnormalverteilung) fragwürdig, weil nicht unendlich viele Beobachtungen zur Verfügung

stehen, sondern nur endlich viele und manchmal auch nur recht wenige. Die Grundidee des *Boots*traps, einem der wichtigsten *Resampling*-Verfahren, ist es, aus den bestehenden Daten viele neue Zufallsstichproben zu erzeugen und auf dieser Basis kritische Werte für die gegebene Statistik zu berechnen. Außerdem eignen sich diese Verfahren auch zur Bereinigung von Verzerrungen in Schätzern etc.

Lehrbücher:

- Wollschläger (2017) (Kap. 11)
- James et al. (2013) (Kap. 5)
- Horowitz (2001)
- Givens and Hoeting (2012) (Kap. 9)

Artikel:

- MacKinnon (2006)

### **Prognoseevaluation: Tests zum Vergleich der Genauigkeit und des Informationsgehalts konkurrierender Prognosen**

Genauere Prognosen sind wichtig, um gute Entscheidungen zu treffen. Und die Fähigkeit, die Güte zweier Prognosen von Analysten oder statistischen Modellen zu vergleichen, ist entscheidend, um zwischen guten und schlechten Prognosen unterscheiden zu können. Der Diebold and Mariano (1995) ist der am häufigsten verwendete Hypothesentest zur Überprüfung, ob ein statistisch signifikanter Unterschied bezüglich der Genauigkeit zweier konkurrierender Prognosemodelle besteht. Eine weitere verwandte Vergleichsmöglichkeit bieten Tests zum Vergleich des Informationsgehalts der konkurrierenden Prognosen (engl. *forecast encompassing tests*). Diese untersuchen, ob eine Prognose die andere umschließt in dem Sinne, dass eine Kombination beider Prognosen keine Verbesserung im Vergleich zu dem einzelnen umschließenden Modell liefert. Die schwächeren Prognosen bieten in diesem Fall keinen Informationsgehalt, der über denjenigen der konkurrierenden Prognosen hinausgeht. Die Seminararbeit soll die beiden Konzepte (Test auf gleiche Prognosegenauigkeit und Test zum Vergleich des Informationsgehalts) einführen und die Prognoseperformance zweier Modelle mit Hilfe dieser Methoden in einer empirischen Anwendung vergleichen.

Lehrbücher:

- Döhrn (2014) (Kap. 10; insb. 10.3)
- Clark and McCracken (2013)
- Clements and Harvey (2009)

Artikel:

- Diebold and Mariano (1995)
- Harvey et al. (1997)

### **Tests auf Stationarität**

Viele statistische Verfahren und Modelle setzen voraus, dass die zugrundeliegende Zeitreihe einem stationären oder zumindest einem schwachstationären Prozess folgt, da nur dann eine richtige Schlussfolgerung aus der empirischen Analyse gezogen werden kann. Zur Überprüfung der Stationaritätseigenschaft einer Zeitreihe werden sogenannte Einheitswurzel-Tests herangezogen. Zu den bekanntesten gehören: Dickey-Fuller (DF)-, Augmented-Dickey-Fuller-Test (ADF-Test), Kwiatkowski-Phillips-Schmidt-Shin (KPSS)-, sowie der Phillips-Perron (PP)-Test. Im Rahmen der Arbeit sollen die Einheitswurzeltests vorgestellt und angewendet werden. Anwendungsbeispiele sind das Bruttoinlandsprodukt und Inflationsrate in Deutschland.

Lehrbücher:

- Kirchgässner and Jürgen (2006) (Kap. 6)

Artikel:

- Dickey and Fuller (1979)
- Dickey and Fuller (1981)
- Kwiatkowski et al. (1992)
- Said and Dickey (1984)

### **Granger Kausalität**

Ein statistisch nachgewiesener Zusammenhang (Korrelation) zwischen zwei (oder auch mehreren) Variablen  $X$  und  $Y$  liefert keine Auskunft darüber, ob die Variable  $X$  die Variable  $Y$  verursacht oder umgekehrt die Entwicklung von  $X$  durch die Variable  $Y$  beeinflusst wird. Es kann auch vorkommen, dass die beiden Ereignisse  $X$  und  $Y$  durch eine dritte Variable  $Z$  verursacht werden oder der Zusammenhang nur zufällig oder sogar fehlerhaft besteht. Eine Aussage über die sogenannte Ursache-Wirkungs-Beziehung wird mit Hilfe des Kausalitätskonzepts getroffen. Kausalität kann sowohl empirisch als auch theoretisch begründet werden. Zum Zwecke der empirischen Untersuchung von Kausalitätsbeziehungen kommt meist das Granger-Kausalität-Verfahren zum Einsatz. Das Konzept der Granger-Kausalität kann beispielsweise zur Untersuchung des Einflusses des Immobilienvermögens auf den privaten Konsum angewendet werden. Ein anderes Beispiel ist die Analyse der

Auswirkungen von Ölpreisänderungen auf den internationalen Spotmärkten auf die deutschen Verbraucherpreise für Sprit. In der Seminararbeit sollen zunächst die theoretischen Überlegungen des Granger-Kausalität Konzepts vorgestellt werden. Im zweiten Schritt soll das Konzept anhand der statistischen Formeln ausführlich dargestellt und mit Hilfe eines einfachen Praxisbeispiels erklärt werden.

Lehrbücher

- Assenmacher (2010) (Kap. 15)
- Kirchgässner and Jürgen (2006) (Kap. 6)
- Bortz and Schuster (2011) (Kap. 10)

Artikel:

- Granger (1969)

### **Varianzanalyse**

Die Varianzanalyse (engl. *Analysis of Variance*, kurz ANOVA) untersucht, ob die Varianz einer abhängigen metrischen Variablen durch die Variation einer unabhängigen nominalen Variablen erklärt werden kann. Bei der Durchführung der Varianzanalyse werden Mittelwertunterschiede zwischen mehreren Stichproben überprüft. Die ANOVA kann beispielsweise sowohl im Marketingbereich als auch zum Zwecke der Hypothesenüberprüfungen im Rahmen von Experimenten zur Anwendung kommen. Sie kann z.B. verwendet werden, um die Zahlungsbereitschaften von Kunden einer Supermarktkette zu untersuchen. Ein Untersuchungsgegenstand könnte lauten, ob die Kunden nach einer Degustation von sechs Schokoladensorten unterschiedlich hohe Zahlungsbereitschaften für die verschiedenen Sorten aufweisen. In der Seminararbeit soll das Prinzip der ANOVA vorgestellt werden und anhand eines Beispiels erklärt und angewandt werden.

Lehrbücher

- Bortz and Schuster (2011)
- Fahrmeir et al. (1996)
- Wollschläger (2017)

Artikel:

- Kaufmann and Schering (2014)
- Ganesan (2007)

## Literatur

- Alexander, C. (2008). Market Risk Analysis, Practical Financial Econometrics -. John Wiley & Sons, New York.
- Assenmacher, W. (2010). Einführung in die Ökonometrie. Walter de Gruyter.
- Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. Journal of Econometrics, 146(2):304–317.
- Bai, J. and Ng, S. (2009). Boosting diffusion indices. Journal of Applied Econometrics, 24(4):607–629.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3):307–327.
- Bollerslev, T., Engle, R., and Nelson, D. B. (1994). Arch models. In Engle, R. F. and McFadden, D., editors, Handbook of Econometrics, volume 4, chapter 49, pages 2959–3038. Elsevier, 1 edition.
- Bortz, J. and Schuster, C. (2011). Statistik für Human- und Sozialwissenschaftler - Limitierte Sonderausgabe. Springer-Verlag, Berlin Heidelberg New York.
- Brooks, C. (2019). Introductory Econometrics for Finance. Cambridge University Press, Cambridge.
- Buchen, T. and Wohlrabe, K. (2011). Forecasting with many predictors: Is boosting a viable alternative? Economics Letters, 113(1):16–18.
- Canova, F. (2011). Methods for Applied Macroeconomic Research -. Princeton University Press, Kassel.
- Christoffersen, P. (2011). Elements of financial risk management. Academic Press.
- Clark, T. and McCracken, M. (2013). Advances in forecast evaluation. volume 2, chapter Chapter 20, pages 1107–1201. Elsevier.
- Clements, M. and Harvey, D. (2009). Forecast combination and encompassing. pages 169–198.
- College, F. D. and Dellaert, F. (2002). The expectation maximization algorithm. Technical report, Georgia Institute of Technology.
- Cryer, J. D. and Chan, K.-S. (2008). Time Series Analysis - With Applications in R. Springer Science & Business Media, Berlin Heidelberg.



- Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with a unit root. JASA. Journal of the American Statistical Association, 74.
- Dickey, D. and Fuller, W. (1981). The likelihood ratio statistics for autoregressive time series with a unit root. Econometrica, 49:1057–72.
- Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3):253–63.
- Döhrn, R. (2014). Konjunkturdiagnose und -prognose: Eine anwendungsorientierte Einführung. Springer Berlin Heidelberg.
- Durbin, J. and Koopman, S. J. (2012). Time Series Analysis by State Space Methods - Second Edition. OUP Oxford, New York, London.
- Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica, 50(4):987–1007.
- Engle, R. and Manganelli, S. (2001). Value at risk models in finance. Working Paper Series 75, European Central Bank.
- Engle, R. F. and Granger, C. W. (1987). Cointegration and error correction: representation, estimation, and testing. Econometrica: journal of the Econometric Society, pages 251–276.
- Estrella, A. and Mishkin, F. (1998). Predicting u.s. recessions: Financial variables as leading indicators. The Review of Economics and Statistics, 80(1):45–61.
- Evgeniou, T. and Pontil, M. (2001). Support vector machines: Theory and applications. volume 2049, pages 249–257.
- Fahrmeir, L., Brachinger, W., Hamerle, A., and Tutz, G. (1996). Multivariate statistische Verfahren. Walter de Gruyter, Berlin.
- Fahrmeir, L., Kneib, T., and Lang, S. (2009). Regression - Modelle, Methoden und Anwendungen. Springer-Verlag, Berlin Heidelberg New York.
- Foroni, C. and Marcellino, M. (2014). A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates. International Journal of Forecasting, 30(3):554–568.
- Franke, J., Härdle, W. K., and Hafner, C. M. (2012). Einführung in die Statistik der Finanzmärkte. Springer-Verlag.
- Ganesan, V. (2007). An analysis of working capital management efficiency in telecommunication equipment industry. Rivier Academic Journal, 3(2).

- Givens, G. H. and Hoeting, J. A. (2012). Computational Statistics -. John Wiley & Sons, New York.
- Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3):424–38.
- Hamilton, J. D. (1994). Time Series Analysis -. Princeton University Press, Kassel.
- Harvey, A., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. The Review of Economic Studies, 61(2):247–264.
- Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. International Journal of Forecasting, 13(2):281–291.
- Hassler, U. (2007). Stochastische Integration und Zeitreihenmodellierung - Eine Einführung mit Anwendungen aus Finanzierung und Ökonometrie. Springer-Verlag, Berlin Heidelberg New York.
- Hastie, T., Tibshirani, R., and Friedman, J. (2013). The Elements of Statistical Learning - Data Mining, Inference, and Prediction. Springer Science & Business Media, Berlin Heidelberg.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. Biometrika, 57(1):97–109.
- Holston, K., Laubach, T., and Williams, J. (2017). Measuring the natural rate of interest: International trends and determinants. Journal of International Economics, 108(S1):S59–S75.
- Horowitz, J. L. (2001). The bootstrap. In Heckman, J. and Leamer, E., editors, Handbook of Econometrics, volume 5, chapter 52, pages 3159–3228. Elsevier, 1 edition.
- Härdle, W. K. and Simar, L. (2015). Applied Multivariate Statistical Analysis -. Springer, Berlin, Heidelberg.
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning - with Applications in R. Springer Science & Business Media, Berlin Heidelberg.
- Kaufmann, J. and Schering, A. (2014). Analysis of Variance ANOVA, pages 1–11. American Cancer Society.
- Kenny, G., Genre, V., Meyler, A., and Timmermann, A. (2010). Combining the forecasts in the ecb survey of professional forecasters: can anything beat the simple average? Working Paper Series 1277, European Central Bank.
- Kirchgässner, G. and Jürgen, W. (2006). Einführung in die moderne Zeitreihenanalyse. Vahlen.

- Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of Econometrics, 54(1-3):159–178.
- Liaw, A. and Wiener, M. (2001). Classification and regression by randomforest. Forest, 23.
- MacKinnon, J. (2006). Bootstrap methods in econometrics. Working Paper 1028, Economics Department, Queen’s University.
- Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7(1):77–91.
- Martin, V., Hurn, S., and Harris, D. (2012). Econometric Modelling with Time Series - Specification, Estimation and Testing. Cambridge University Press, Cambridge.
- Munkhdalai, L., Namsrai, O.-E., and Ryu, K. (2018). Credit scoring with deep learning.
- Murphy, K. P. (2012). Machine Learning - A Probabilistic Perspective. MIT Press, Cambridge.
- Neusser, K. (2011). Zeitreihenanalyse in den Wirtschaftswissenschaften -. Springer-Verlag, Berlin Heidelberg New York.
- Neusser, K. (2016). Time Series Econometrics. Springer, Berlin, Heidelberg.
- Richter, S. (2019). Statistisches und maschinelles Lernen - Gängige Verfahren im Überblick. Springer-Verlag, Berlin Heidelberg New York.
- Robert, C. and Casella, G. (2013). Monte Carlo Statistical Methods -. Springer Science & Business Media, Berlin Heidelberg.
- Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika, 71(3):599–607.
- Schlittgen, R. (2015). Angewandte Zeitreihenanalyse mit R. Walter de Gruyter GmbH & Co KG, Berlin.
- Schröder, M. (2012). Finanzmarkt-Ökonometrie - Basistechniken, fortgeschrittene Verfahren, Prognosemodelle. Schäffer-Poeschel, Stuttgart.
- Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance, 19(3):425–442.
- Shiryaev, A. (1999). Essentials Of Stochastic Finance: Facts, Models, Theory -. World Scientific, Singapur.

- Shumway, R. H. and Stoffer, D. S. (2017). Time Series Analysis and Its Applications - With R Examples. Springer, Berlin, Heidelberg.
- Spremann, K. (2014). Portfoliomanagement -. Walter de Gruyter GmbH & Co KG, Berlin.
- Stock, J. and Watson, M. (2004). Combination forecasts of output growth in a seven-country data set. Journal of Forecasting, 23:405–430.
- Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics, pages 147–162.
- Timmermann, A. (2006). Forecast combinations. In Elliott, G., Granger, C., and Timmermann, A., editors, Handbook of Economic Forecasting, volume 1, chapter 04, pages 135–196. Elsevier, 1 edition.
- Tsay, R. S. (2010). Analysis of financial time series. John Wiley & sons.
- Tsay, R. S. (2013). Multivariate Time Series Analysis - With R and Financial Applications. John Wiley & Sons, New York.
- Tsay, R. S. (2014). An Introduction to Analysis of Financial Data with R. John Wiley & Sons, New York.
- Uebele, M. and Ritschl, A. (2009). Stock markets and business cycle comovement in germany before world war i: Evidence from spectral analysis. Journal of Macroeconomics, 31(1):35–57.
- Verbeek, M. (2017). A guide to modern econometrics. Wiley, Chichester, 5. ed., reprint. with corr. edition.
- Wohlrabe, K. (2009). Forecasting with mixed-frequency time series models. Dissertation.
- Wollschläger, D. (2017). Grundlagen der Datenanalyse mit R - Eine anwendungsorientierte Einführung. Springer-Verlag.