

Ökonometrisches Seminar für Masterstudenten

“Data mining”

Teilnahmevoraussetzungen

Mindestvoraussetzung: Bestandene Klausur Methoden der Ökonometrie. Sehr hilfreich ist darüber hinaus der Stoff aus Fortgeschrittene Ökonometrie. Je nach Themenwahl sind auch Kenntnisse aus der Zeitreihenanalyse relevant.

Themen

Die Seminararbeit kann etwa in Form einer eigenständigen Aufbereitung aktueller Literatur auf dem jeweiligen Gebiet, einer theoretischen Darstellung, einer empirischen Untersuchung oder einer Simulationsstudie erfolgen. Wir diskutieren moderne Verfahren des data mining. Denkbare Seminarthemen ergeben sich zum Beispiel aus der Bearbeitung von Werkzeugen des data mining, die nicht in der Vorlesung Fortgeschrittene Ökonometrie (Sommersemester 2015: Statistical Learning) behandelt wurden.

Mögliche Quellen, die als Ausgangspunkt zur selbständigen weiterführenden Literaturrecherche dienen können und sollen:

Bishop CM. 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning*. Springer, 2nd edn.

Kuhn M, Johnson K. 2013. *Applied Predictive Modeling*. New York, Heidelberg, Dordrecht, London: Springer.

Besonders honoriert werden sollen Anwendungen, die sich der Herausforderung stellen “unsaubere” Datensätze zu bearbeiten.

Ablauf des Seminars

1. Einführungsveranstaltung: 14.10.2015 von 10:00 - 12:00 Uhr. Hier werden unter anderem Hinweise zu wissenschaftlichem Arbeiten gegeben. Bitte melden Sie sich zu diesem Termin vorab an!
2. Verbindliche Anmeldung: Ab sofort bis zum 16.10.2015 mit drei priorisierten Themenwünschen (per email an Christoph Hanck), wobei auch eigene Themenvorschläge möglich und willkommen sind. Die Themenvergabe erfolgt am Tag darauf. Im (unseres Erachtens unwahrscheinlichen) Fall einer Überbuchung des Seminars richtet sich die Teilnahme an den Noten in Ökonometrie sowie nachrangig nach Eingang der Anmeldungen.
3. Evtl. Sprechstundentermine. Bis zum 10.11.2015 muss eine erste Gliederung des Themas abgegeben werden.
4. Abgabetermin der Seminararbeit: (spätestens) am 15.12.2015 am Lehrstuhl. Zusätzliche Abgabe aller relevanten Dateien (s.u.) in elektronischer Form per E-Mail/Dropboxlink o.ä. an christoph.hanck@vwl.uni-due.de.
5. Termin der Blockveranstaltung: Im Januar 2016. Die genauen Termine werden nach Abstimmung noch rechtzeitig bekannt gegeben.
6. Leistungsnachweis: Schriftliche Hausarbeit 40%, 10% Qualität der Replikationsdateien, Vortrag 25%, 15% Koreferat und aktive Teilnahme an der Diskussion 10%. Zum Bestehen des Seminars müssen alle Teile bestanden sein.

7. Es werden 6CP vergeben.

Anforderungen

1. Teilnahme an der Einführungsveranstaltung
2. Verfassen der Seminararbeit in deutscher oder englischer Sprache: ca. 15 Seiten netto, d.h. ohne Tabellen, Grafiken, Anhang etc.
3. Es sind zudem die Rohdaten sowie der Quellcode der statistischen Analyse einzureichen. Dieser ist so zu kommentieren, dass nachvollziehbar wird, welcher Teil der Programme für welche Ergebnisse verantwortlich ist ("reproducible research"). Er muss des Weiteren von Modifikationen wie Quellverzeichnissen abgesehen direkt ausführbar sein. Mit Werkzeugen wie RMarkdown werden (a) und (b) idealerweise in einem Dokument kombiniert.
4. Evtl. Besuch der Sprechstunde.
5. Vortrag: Im Rahmen einer Blockveranstaltung soll ein 30-minütiges Referat über den Inhalt der Arbeit gehalten werden. Daran schließt sich eine kritische Diskussion an, bei der eine aktive Teilnahme der Seminarteilnehmer vorausgesetzt wird. Für die Blockveranstaltung besteht Anwesenheitspflicht.
6. Thesenpapier: Für die mündliche Präsentation ist von allen Seminarteilnehmern ein Thesenpapier im Umfang von einer Seite vorzubereiten. Zweck des Thesenpapiers ist es, den Vortrag für die anderen Seminarteilnehmer zusammenzufassen sowie überschaubar und in seinem Aufbau nachvollziehbar zu machen. Das Thesenpapier ist von den Vortragenden selber zu vervielfältigen.
7. Koreferat: In einem 10-minütigen Kurzvortrag wird die Arbeit einer/s Kommilitonin zusammengefasst sowie kritisch, aber konstruktiv gewürdigt. Mögliche Aspekte: Welche alternativen Methoden hätten verwendet werden können? Welche weiteren Anwendungsmöglichkeiten bestehen für die verwendeten Methoden? Die Zuteilung der Koreferate kann gemeinsam abgestimmt werden, erfolgt aber durch den Lehrstuhl.
8. Formvorschriften: Bei der Anfertigung der Arbeit beachten Sie bitte die nachfolgenden Hinweise.

Allgemeine Hinweise

- Titelblatt
- Inhaltsverzeichnis
- ggf. Symbol- und Abkürzungsverzeichnis
- ggf. Abbildungsverzeichnis
- ggf. Tabellenverzeichnis
- Textteil (zentraler Bestandteil der Arbeit, s.u.)
- ggf. Anhang
- Literaturverzeichnis

Alle Seiten nach dem Titelblatt sind durchnummerieren. Die Vorseiten (Inhalts-, Symbol- und Abkürzungs-, Abbildungs-, und Tabellenverzeichnis) sind mit römischen Ziffern fortlaufend zu nummerieren. Im Anschluss daran ist die Bearbeitung durchnummerieren mit arabischen Ziffern zu nummerieren.

Titelblatt

Das Titelblatt der Seminararbeit enthält folgende Angaben:

- Thema der Arbeit
- Die Bezeichnung der Arbeit als Seminararbeit, laufendes Semester
- Angaben zum betreuenden Lehrstuhl
- Angaben zum Verfasser (Name, E-Mail-Adresse, Matrikelnummer, Anzahl der Fachsemester und Studiengang)
- Abgabedatum

Inhaltsverzeichnis

Das Inhaltsverzeichnis gibt einen Überblick über die Arbeit. In ihm werden die Verzeichnisse (nicht jedoch das Inhaltsverzeichnis selbst!) und die Gliederung des Textteils jeweils mit Seitenangaben aufgeführt. Die Gliederung verdeutlicht die logische Struktur der Bearbeitung. Dazu ist es erforderlich, dass die gewählten Überschriften aussagefähig sind und in einem konsistenten Zusammenhang stehen.

Abbildungen, Tabellen, Symbole

Abbildungen und Tabellen sind im Text an geeigneter Stelle einzufügen. Abbildungen und Tabellen sind fortlaufend zu nummerieren und mit einem Titel, der ober- oder unterhalb der entsprechenden Abbildung/Tabelle steht, zu versehen (z. B. "Abb. 1: Phillips-Kurve", "Tab. 1: Einheitswurzeltests") und werden mit Titel und Seitenangabe in ein entsprechendes Verzeichnis aufgenommen. Auf ein Verzeichnis kann verzichtet werden, wenn nur eine Abbildung oder Tabelle vorkommt. Verwendete mathematische Symbole werden in einem Symbolverzeichnis deklariert. Die gleiche ökonomische Größe sollte im Laufe des Textes nicht mit verschiedenen Symbolen bezeichnet werden bzw. dasselbe Symbol sollte nicht für verschiedene Bedeutungen verwendet werden.

Anhang

Ein Anhang ist nur dann notwendig, wenn Abbildungen und Tabellen entweder ihrer Art oder ihrem Umfang nach nicht im laufenden Text untergebracht werden können. In einen Anhang relegiert man z.B. die Ergebnisse statistischer Auswertungen, Wertetabellen usw., welche der Leser nur dann zu sehen braucht, wenn er die Thesen, die der Verfasser aus ihnen ableitet, bezweifelt. Üblicherweise werden im Anhang erst die Abbildungen und dann die Tabellen (oder umgekehrt) aufgenommen. Häufig verwendet man bei der Nummerierung der Abbildungen/Tabellen im Titel ein "A", zur Kenntlichmachung, dass sich die entsprechende Abbildung/Tabelle im Anhang befindet (also "Abb. A15: Arbeitslosenquote in OECD-Ländern" oder "Tab. A10: Autokorrelationstests"). Auch die im Anhang aufgeführten Abbildungen/Tabellen müssen in den vorangestellten Verzeichnissen angegeben werden. Wenn die Tabellen oder Abbildungen im Anhang jeweils höchstens eine Seite einnehmen, dann ist i.A. nicht einzusehen, warum sie einen Anhang bilden müssen.

Äußere Form der Arbeit

Die in DIN A4-Format anzufertigende Arbeit ist in Normalschrift (z.B. Times New Roman 12 oder Arial 10) und mit 1,5-fachem Zeilenabstand zu schreiben. Auf jeder Seite ist links ein ca. 5 cm und rechts ein ca. 2,5 breiter Korrekturrand frei zu lassen. Fußnoten werden grundsätzlich auf den Seiten platziert, zu deren Text sie gehören. Innerhalb der Fußnoten ist einzeilig zu schreiben. Die Seiten der Arbeit sind fortlaufend zu nummerieren. Sie können in Word arbeiten, wir empfehlen aber \LaTeX zur Erstellung von vernünftig aussehenden Formeln.

Quellenangaben

Die Kennzeichnung wörtlicher oder sinngemäßer Zitate stellt einen der wichtigsten Aspekte wissenschaftlichen Arbeitens dar: Es hilft die eigene Leistung herauszustellen und verhindert Plagiate. Quellenangaben erfolgen in Fußnoten oder—wie in englischsprachigen Texten üblich—in Klammern direkt im Anschluss an das (wörtliche oder sinngemäße) Zitat mit Angabe der Seitenzahl. Wörtliche Zitate sind zusätzlich in Anführungszeichen zu setzen. Fußnoten werden durchgehend nummeriert. Fußnotentexte stehen auf derselben Seite wie die Fußnotenziffern.

Beispiele:

Mustermann (1996, S. 12)

Musterfrau et al. (2000, S. 10 ff)

(Mustermann, 2002, S. 12)

Bibliographische Angaben

Alle verwendeten und zitierten Literaturquellen sind unter Literatur in alphabetischer Reihenfolge aufzuführen. Dabei sind die Familiennamen der Autoren aufzuführen, Vornamen werden abgekürzt.

Monografien müssen mit Verfasser, Erscheinungsjahr, Titel, Auflage (außer 1. Auflage) und Erscheinungsort angegeben werden.

Zeitschriftenaufsätze erfordern folgende Angaben: Verfasser, Erscheinungsjahr, Titel des Beitrags, Namen und Jahrgang der Zeitschrift (+ evtl. Band oder Heft).

Beiträge in Sammelwerken erfordern folgende Angaben: Verfasser, Erscheinungsjahr, Titel des Beitrags, Namen des Herausgebers des Sammelwerkes ergänzt um "(Hrsg.)", Titel des Sammelwerkes, Erscheinungsort, Verlag und der Seitenzahl erfasst. Bei mehr als drei Herausgebern wird nur der erste genannt und die weiteren mit "et al." zusammengefasst.

Diskussionspapiere, Arbeitspapiere u. Ä. sollten folgende Angaben enthalten: Verfasser, Erscheinungsjahr, Titel des Beitrags, Name und Nummer der Reihe.

Beispiele:

Mustermann, A. und Musterfrau, B. (1996): Buchtitel. Erscheinungsort: Verlag.

Mustermann, A., B. Musterfrau, und C. Mustermann (1996): Titel des Aufsatzes. In: Titel der Zeitschrift, Jahrgang (Heft-Nr.), Seitenangaben.

Mustermann, A. (1996): Titel des Aufsatzes. In: Musterfrau, B. und Mustermann, C. (Hrsg.): Titel des Sammelbandes. Erscheinungsort: Verlag, Seitenangaben.

Sie können auch einen anderen anerkannten Zitierstil (etwa aus einer bestimmten Fachzeitschrift) wählen, so lange Sie ihn konsistent anwenden. Sie dürfen jedoch nicht mehrere Zitierstile in einem Quellenverzeichnis vermischen.

Inhalt und Aufbau der Arbeit

Trotz der Notwendigkeit der Einhaltung der oben genannten Vorgaben bzgl. der äußeren Form der Arbeit stellt der Inhalt natürlich den für die Beurteilung der Arbeit ausschlaggebenden Aspekt dar. Grundsätzlich kann die Seminararbeit in Form einer eigenständigen Aufbereitung der aktuellen (bzw. auf Bachelorniveau Lehrbuch-) Literatur auf dem jeweiligen Gebiet, einer empirischen Untersuchung oder einer Simulationsstudie erfolgen.

Die Arbeit könnte in etwa wie folgt aufgebaut sein. Es ist allerdings Ihre Aufgabe, darüber nachzudenken, ob gegeben Ihr Thema ein abweichender Aufbau sinnvoll ist.

- Abstract: Kurzzusammenfassung der Arbeit
- Einleitung: Motivation, Zielsetzung
- Problemstellung: Welche Methoden werden verwendet/vorgelegt? Welche ökonomische Theorie wird empirisch untersucht?
- Literaturüberblick: Einbettung der Arbeit in die vorhandene Literatur
- Daten: exakte Definition, Quelle(n)
- Empirische Analyse/Simulationsstudie/Theoretische Zusammenfassung der aktuellsten Entwicklungen: Ergebnisse, Interpretation, etc.
- Zusammenfassung und Fazit
- Literaturverzeichnis

Bewertung der Arbeit

Hier geben wir einige Hinweise zur Bewertung der Arbeit. Wenn Sie den Eindruck haben, dass diese Angaben oft etwas vage sind, haben Sie wahrscheinlich recht: Es ist schwer präzise zu sagen, was ein gutes Papier ausmacht, aber man weiß, dass es eins ist, wenn man es liest.

Einige allgemeine Angaben können jedoch gemacht werden.

1. Relevanz und/oder Originalität des Themas. Hier besteht ein gewisser Tradeoff: Ein empirisches Papier, das den Zusammenhang von Inflation und Arbeitslosigkeit analysiert, ist sicher relevant, aber nicht sehr originell. Ihre Arbeit sollte mindestens eins der zwei sein.
2. Plausibilität. Sie haben keinerlei Verantwortung, in einer empirischen Arbeit etwa statistisch signifikante Ergebnisse zu produzieren, oder spektakuläre Ergebnisse in einer Simulationsstudie. Es sollte aber eine plausible Hypothese hinter Ihrer Arbeit stehen. Motivieren Sie diese Hypothese angemessen?
3. Tiefgang. Ist Ihre Arbeit "quick-and-dirty" oder untersuchen Sie das Thema in mehreren Facetten, von denen Sie einige vielleicht nur in einer Fußnote erwähnen? Gerade für das Bachelor-Seminar kann eine weitergehende Literaturrecherche hier hilfreich sein.
4. Korrektheit, selbstverständlich. Werden alle Methoden korrekt beschrieben und/oder angewandt? Modellieren Sie die Eigenschaften Ihrer Daten angemessen oder wenden Sie bestimmte Techniken nur an, weil Ihr Softwarepaket dafür einen Knopf bereithält?

5. Interpretation. Machen Sie klar, was Ihre Ergebnisse bedeuten—sowohl ökonomisch als auch ökonometrisch—oder drucken Sie nur Tabellen mit Rechneroutput? Das selbe gilt für theoretische Arbeiten. Schreiben Sie einfach Formeln auf oder zeigen Sie ihre Bedeutung auf?
6. Kompaktheit. Konzentrieren Sie sich auf die wichtigen Ergebnisse oder liefert das Papier viel nebensächlichen Output bspw. einer Regression, der nicht recht zum Thema gehört? Hat das Papier lange Absätze wo ein klarer Satz gereicht hätte? Haben Sie drei Grafiken mit bspw. jeweils einer Zeitreihe oder fassen Sie sie in einer Grafik zusammen, wenn sich das anbietet? Konzentrieren Sie sich auf das, was *Sie* gemacht haben oder besteht Ihr Aufsatz fast nur aus Zitaten?
7. Vollständigkeit. Hier ist eine gewisse Spannung zum vorigen Punkt: Haben Sie alle relevanten Techniken angewandt/betrachtet, die für Ihr Thema wichtig sind?
8. Sprache. Ist das Papier klar und korrekt geschrieben? Ist Ihr Stil professionell oder ähnelt er einem Post auf Facebook? Benutzen Sie ein Rechtschreibkorrekturprogramm, aber vertrauen Sie nicht nur auf ihn!
9. Struktur. Liest sich das Papier glatt? Machen Sie gleich klar, worum es in Ihrem Papier geht? Haben Sie die Formulierungen in Ihrem Papier mehrfach überarbeitet, oder ist die erste Version, die 15 Seiten lang ist, auch gleich die Endfassung (z.B. weil die Zeit bis zum Abgabetermin falsch geplant wurde)? Machen Sie klar, *warum* Sie bestimmte Schritte machen oder zählt das Papier einfach auf, was Sie im Laufe des Semester so alles für das Seminar gemacht haben? Konzentrieren Sie sich auf zentrale Ergebnisse oder ähnelt das Papier einem Tagebuch a la "...und dann lud ich die Daten nach eViews, dann betrachtete ich die Grafiken, fand jedoch nichts interessantes, dann...". Ich übertreibe, aber ein professionelles Papier hat nichts dergleichen.
10. Layout. Sieht das Papier aus wie ein professioneller Aufsatz einer Fachzeitschrift oder kopiert es nur Resultate aus (bspw.) eViews? Ist die Formatierung konsistent oder findet man drei verschiedene Arten, wie Tabellen, Regressionsergebnisse oder Grafiken dargestellt werden? Denken Sie gut darüber nach, wie etwa Grafiken gegeben die Aussage, die Sie treffen wollen, gestaltet werden oder nimmt das Papier immer die erste Option des Softwarepakets?