

Fachbereich Wirtschaftswissenschaften
Goethe Universität Frankfurt

Selbstüberschätzung bei angehenden Finance- und Management-Studierenden

Eine quantitative Analyse der Determinanten von Overconfidence

Bachelor-Arbeit

Abgabetermin: 21. Januar 2016

Catharina Köhler

Prof. Dr. Horst Entorf
Applied Econometrics and
International Economic Policy

Zusammenfassung

Haben Studierende der Schwerpunkte *Finance & Accounting* und *Management* bereits zu Studienbeginn eine verzerrte Selbsteinschätzung bezüglich ihrer mathematischen Leistungsfähigkeit? Falls ja, welche Determinanten bestimmen diese Fehleinschätzungen maßgeblich?

Die individuelle mathematische Selbsteinschätzung von Erstsemester-Studierenden der Wirtschaftswissenschaften kongruiert häufig nicht mit deren tatsächlichen mathematischen Leistungen. Empirischen Studien zufolge ist neben der Fehleinschätzung der eigenen Fähigkeiten eine Diskrepanz zwischen Selbst- und Fremdbild sowohl bei Berufseinsteigern als auch bei Führungskräften zu beobachten.

Ziel dieser Bachelor-Arbeit ist die Einschätzung des Ausmaßes an Selbstüberschätzung (*Overconfidence*) bei Studienanfängern. Im Fokus liegt hierbei der Vergleich von *Finance- & Accounting*-Studierenden in Abgrenzung zu *Management*-Studierenden. Das Confidence-Level wird durch das Verhältnis zwischen Selbsteinschätzung und erreichtem Klausurergebnis aus drei Übungs- und einer finalen Abschlussklausur im Modul *Mathematik für Wirtschaftswissenschaftler* quantifiziert. Bei Kontrolle von soziodemographischen, die Schulleitung, die universitäre Leistung und deren Einschätzung betreffenden Variablen ergibt sich ein signifikanter Einfluss des Geschlechts und der Schwerpunktwahl auf das Ausmaß an Confidence.

Stichworte: Overestimation, Mathematical Confidence, Finance & Accounting, Schwerpunkt, OLS-Regression, Heckit-Methode.

Inhaltsverzeichnis

1	Einleitung	5
2	Aktueller Forschungsstand	6
3	Relevanz der Mathematik im Studium der Wirtschaftswissenschaften	8
4	Hypothesen und Methoden	9
5	Daten	10
5.1	Datenerhebung	10
5.2	Datenauswertung	11
5.3	Items und Skalen	12
5.4	Trimmung des Datensatzes	14
6	Deskriptive Statistiken	14
6.1	Das Kriterium $MCON_t$	16
6.1.1	Eingangsklausur Mathe-Vorkurs (T_1)	16
6.1.2	Ausgangsklausur Mathe-Vorkurs (T_2)	17
6.1.3	Probeklausur der OMAT-Vorlesung (T_3)	18
6.1.4	Klausur OMAT-Vorlesung (T_4)	19
6.1.5	Fazit	20
6.2	Mittelwertvergleich	21
6.2.1	Auswahl des Testverfahrens	21
7	Multivariate Analysen	22
7.1	Regressionsergebnisse	24
7.1.1	OLS-Regressionen T_2	24
7.1.2	Heckman-Zweistufen-Verfahren T_2	25
7.1.3	OLS-Regressionen T_4	26
7.1.4	Heckman-Zweistufen-Verfahren T_4	29
7.2	Annahmen der multiplen Regression	31
8	Konklusion	31
A	Zusatz: Deskriptive Statistiken	34
A.1	Relative Selbsteinschätzung und Relatives Ergebnis	34
A.2	Regressionen der Mathematical Confidence	34
B	Zusatz: Mittelwertvergleich	35
B.1	Der Shapiro-Wilk-Test	35
B.2	Der Levene-Test	36
B.3	Der Rangsummentest nach Wilcoxon	37

C Zusatz: Multivariate Analysen	38
C.1 Korrelationen der Prädiktoren mit dem Kriterium	38
C.2 Korrelationen der Prädiktoren	38

Abbildungsverzeichnis

1	Ablauf der Datenerhebung	10
2	MCON ₁ , Eingangsklausur des Mathe-Vorkurses	17
3	MCON ₂ , Ausgangsklausur des Mathe-Vorkurses	18
4	MCON ₃ , Probeklausur der OMAT-Vorlesung	19
5	MCON ₄ , Klausur der OMAT-Vorlesung	20

Tabellenverzeichnis

1	Übersicht relevanter Variablen und deren Eigenschaften	12
3	Deskriptive Statistiken der relevanten Variablen	15
4	MCON ₁ – MCON ₄ im Vergleich	20
5	Ergebnisse des Wilcoxon-Rangsummentest $MCON_1 - MCON_4$	22
6	OLS vs. Heckit, Kriterium: $\ln(MCON_2)$	27
7	OLS vs. Heckit, Kriterium: $\ln(MCON_4)$	30
8	Deskriptive Statistiken $RelSE_t$ und $RelRES_t$	34
9	Regressionen von $RelRES_t$ auf $RelSE_t$	35
10	Ergebnisse des Shapiro-Wilk-Tests $MCON1-MCON4$	36
11	Ergebnisse des Levene-Tests $MCON1-MCON4$	37
12	Korrelationen der Prädiktoren mit dem Kriterium	38
13	Korrelationen der Prädiktoren	40

"No problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence."

(Plous, 1993, S. 217)

1 Einleitung

Das Phänomen Overconfidence ist ein zentrales Thema in der Forschung von Entscheidungen und Beurteilungen, nicht nur in der Psychologie, sondern auch in den Wirtschaftswissenschaften und insbesondere in den *Behavioral Economics*. Unsicherheiten sind ständiger Bestandteil unseres Alltags, seien es unerwartete Situationen, der Umgang in zwischenmenschlichen Beziehungen oder die Einschätzung der eigenen Person bezüglich bestimmter Fähigkeiten und Fertigkeiten. Um in derart unsicheren Situationen Handlungsalternativen abzuwägen und zu einer Entscheidung zu gelangen, unterscheidet die Verhaltensökonomik folgende zwei Konzepte:

Abhängig vom Grad der Rationalität des Entscheiders erfolgt die Entscheidungsfindung entweder auf Basis der rationalen Maximierung zugrundeliegender individueller Präferenzen mithilfe von Wahrscheinlichkeitsurteilen oder vereinfacht, und in der Empirie häufig beobachtet, mithilfe von Urteilsheuristiken [14]. Laut Gillenkirch sind Wahrscheinlichkeitsurteile quantifizierte Erwartungen von Handlungsalternativen unter Beachtung exogener Faktoren. Die Bildung von Wahrscheinlichkeitsurteilen erfolgt anhand Verarbeitung und Gewichtung verfügbarer Informationen. Hierbei gilt, je mehr Informationen vorhanden sind, desto fundierter das Wahrscheinlichkeitsurteil. Im Gegensatz hierzu stellen Urteilsheuristiken vereinfachte und zumeist unbewusste Denkmuster dar, die bei der Bildung von Urteilen in unsicheren, jedoch weniger gravierenden Situationen angewandt werden [15].

Unter der Voraussetzung psychischer Gesundheit, besitzt jede Person die meisten Informationen über sich selbst, da jede private Information idealerweise bekannt ist. Die Selbsteinschätzung auf Basis von Wahrscheinlichkeitsurteilen sollte auf Basis perfekter Information kein allzu großes Problem darstellen. Dennoch existiert das Phänomen der Fehleinschätzung eigener Fähigkeiten auch in wirtschaftlichen Berufen und ferner über verschiedene Karrierestufen hinweg: „Neben Berufsanfängern haben hierarchiehohe Manager sicherlich die größten Wahrnehmungsschwierigkeiten. Ihnen werden meist nur Erfolgsmeldungen weitergeleitet - läuft hingegen etwas schief, traut sich niemand, was zu sagen“, so Rüdiger Hossiep [13, WirtschaftsWoche]. Dieses Schicksal ereilt laut Hossiep (2009) einen nicht kritikfähigen Manager bereits nach durchschnittlich drei bis vier Jahren. In einem weiteren Interview stellte dieser fest, dass „Selbst- und Fremdwahrnehmung von Führungskräften [...] immer weiter auseinander [klaffen].“ [19]. Dies führt zu Unzufriedenheit der Mitarbeiter und hemmt Weiterentwicklungsmöglichkeiten der Führungskräfte. Die Folge ist eine Stagnation der Produktivität. Die Basis für eine realistische Selbstreflexion der Führungskraft sieht Hossiep (2009) insbesondere in einem guten Vertrauensverhältnis zwischen Führungskraft und Mitarbeitern. Durch konstruktive Kritik könnten Führungskräfte „ihr eigenes Führungshandeln sensitiver wahrnehmen [...] und als Führungskraft [...] wachsen.“

Besteht dieser Hang zur Fehleinschätzung eigener Fähigkeiten bereits zu Studienbeginn oder bildet sich dieser erst im Laufe des Studiums oder der Karrierelaufbahn heraus? Existieren in diesem Zusammenhang maßgebliche Faktoren, die eine Selbstüberschätzung begünstigen?

Diese Arbeit untersucht die Fähigkeit der Leistungseinschätzung von Studienanfängern im ersten Semester des Bachelor-Studiengangs *Wirtschaftswissenschaften* an der Goethe-Universität Frankfurt mit dem Ziel herauszufinden, ob sich die Studierenden überschätzen und ob dies bereits zu Studienbeginn der Fall ist. Basierend auf einer Studie und Datenerhebung aus dem Sommersemester 2015 wird die Hypothese analysiert, dass die Studierenden des voraussichtlich gewählten Schwerpunktes *Finance- & Accounting* in höherem Maß zu Overconfidence neigen als Studierende der voraussichtlichen Spezialisierung *Management*. Aufgrund der hohen Relevanz der Mathematik¹ wird das Ausmaß

¹Beispiele zu mathematischen Anwendungen in der Wirtschaft und wirtschaftlichen Forschung sind insbesondere Optimierungsprobleme, die statistische und ökonometrische Analyse von beispielsweise Finanzdaten, die Modellentwicklung und Hypothesenbildung.

der Selbstüberschätzung in der quantitativen Einführungsveranstaltung *Mathematik für Wirtschaftswissenschaftler (OMAT)* untersucht. Zu diesem Zweck wird in 6.2 die Variable *Mathematical Confidence (MCON)* vorgestellt, anhand der die Einflussfaktoren auf das Confidence-Level der Studierenden identifiziert werden sollen. Der verwendete Datensatz ermöglicht es, die Confidence-Level der Studierenden der beiden Schwerpunkte in vier Messzeitpunkten über das erste Semester zu verfolgen und vergleichen.

Wichtig zu beachten ist, dass die in dieser Arbeit verwendeten Termini *Overconfidence* und *Underconfidence* im Folgenden synonym für die von Moore (2008) definierte *Overestimation* bzw. *Underestimation* verwendet werden (s. Abschnitt 2).

Diese Arbeit ist folgendermaßen gegliedert: Abschnitt 2 beschreibt verschiedene Konzepte der Psychologie und Behavioral Finance zur Messung des Confidence-Levels allgemein und insbesondere der Overconfidence. Zusätzlich werden bisherige Forschungsergebnisse zur Thematik der Selbstüberschätzung von Studierenden aufgezeigt. In Abschnitt 3 folgt eine Beschreibung des Aufbaus des Studiengangs *Wirtschaftswissenschaften*. In diesem Zusammenhang wird auch die Einbettung des Fachs Mathematik in das Curriculum verdeutlicht. Daraufhin wird die Kernhypothese dieser Arbeit formuliert und die zur Analyse verwendeten Methoden vorgestellt. Nach einer Beschreibung des zugrundeliegenden Datensatzes und insbesondere der relevanten Variablen wird in Abschnitt 6 die Mathematical Confidence für die Schwerpunkte *Finance & Accounting (F&A)* und *Management* in allen vier Messzeitpunkten $T_1 - T_4$ gegenübergestellt. Im Anschluss wird jeweils ein Mittelwertvergleich pro Messzeitpunkt durchgeführt, der vermeintliche Gruppenunterschiede zwischen den besagten Schwerpunkten identifiziert. In Abschnitt 7 werden schließlich mithilfe von OLS-Regressionen und dem Heckman-Zweistufen-Verfahren mögliche Einflussfaktoren auf die Mathematical Confidence untersucht. Die Auswertung schließt in Abschnitt 8 mit einer Zusammenfassung der Ergebnisse und möglichen Ausblicken.

2 Aktueller Forschungsstand

Eine Befragung von Personalchefs von 460 Unternehmen jeglicher Größe in Deutschland, der Schweiz und in Österreich ergab, dass die überwiegenden Ursachen für das frühzeitige Scheitern von Berufseinsteigern zu 94 Prozent auf Selbstüberschätzung und zu 89 Prozent auf mangelnde Fähigkeit der Selbstkritik zurückzuführen ist [20]. Der Studie der Managementberatung Kienbaum zufolge sind die Karrierechancen für junge, hochqualifizierte Absolventen in Zeiten des „War for talents“ besonders hoch: Durchschnittlich 70 Prozent der befragten Unternehmen gaben 2011 an, dass es insbesondere in den Sparten der Forschung und Entwicklung, der IT und in der Produktion an Fachkräften mangelt. Den befragten HR-Leitern zufolge gelten als Einstiegsvoraussetzungen neben einem guten Abschluss, insbesondere soziale Kompetenzen wie Eigenmotivation und die Fähigkeit zur Selbstreflexion. Auch Bescherer (2003, S. 12) akzentuiert, dass „aus der Wirtschaft [...] seit Jahren Klagen über die mangelhaften Qualifikationen der Schulabgängerinnen und -abgänger [kommen]“. Dies betrifft nicht nur Fachkompetenzen, sondern auch Schlüsselqualifikationen.

Aber nicht nur Schulabgänger und Berufseinsteiger sind von dem Phänomen der Fehleinschätzung eigener Fähigkeiten betroffen: Außerdem weist Hossiep darauf hin, dass insbesondere bei hierarchiehohe Managern das Selbst- und Fremdbild divergiert [13, WirtschaftsWoche]. Laut diesem liegt die Ursache für das konstant wachsende Ego der Karrieristen in der sehr einseitigen Feedbackkultur. Den Managern werden hauptsächlich Erfolgsmeldungen zugetragen; Misserfolge hingegen werden seltener kommuniziert. Laut Kanning ist allerdings die Fähigkeit zur objektiven Selbstreflexion eine essentielle Kompetenz, insbesondere von Managern, die Verantwortung gegenüber dem Unternehmen sowie deren Mitarbeiter tragen [13, WirtschaftsWoche].

Die Praxis verlangt also die Fähigkeit zur realistischen Selbstreflexion und Selbstkritik. Empirische Studien zeigen die Gründe für die Wichtigkeit einer akkuraten Selbsteinschätzung und verweisen gleichzeitig auf weitreichende Folgen von Overconfidence in der wirtschaftlichen Praxis: Im Fall von Overconfidence in Form des Better-than-average-Effects, der Überzeugung bessere Investitionsstrategien zu entwickeln als der durchschnittliche Fonds-Manager, ergibt

sich ein erhöhtes Handelsvolumen [16], häufig ohne den Markt zu übertreffen [34]. Laut Van Zant (2013) gehen sich selbst überschätzende CEOs zu häufig Firmenfusionen ein, was eine Schmälerung des Shareholder Values zur Folge hat. Zudem gehen diese durchschnittlich schneller das Risiko ein, hochriskante Produkte auf den Markt zu bringen. Zur Analyse der verschiedenen Ausprägungen und Facetten von Overconfidence ist eine Quantifizierung notwendig. In der empirischen Forschung existieren diesbezüglich unterschiedliche Ansätze. Eine in der *Behavioral Finance* bekannte Kennzahl zur Einschätzung des Ausmaßes von Overconfidence ist der *Brier Score (BS)* bzw. *Probability Score (P)* [5]. Um das Maß an Overconfidence von einzelnen Personen zu bestimmen, müssen folgende Voraussetzungen getroffen werden: Den Probanden werden jeweils N *Entscheidungsfragen*ⁱⁱ zur Beantwortung vorgelegt, zu denen anschließend jeweils ein *Wahrscheinlichkeitsurteil* $p_i \in [0, 5; 1]$ ⁱⁱⁱ abgegeben wird. Mit diesem wird eingeschätzt, wie sicher sich der Proband seiner Antwort zur Frage i ist. Dabei sei im folgenden $c_i \in \{0; 1\}$ der Indikator, ob Frage i richtig beantwortet ist. Der Brier Score berechnet sich wie folgt [5]:

$$BS = \frac{1}{N} \sum_{i=1}^N (p_i - c_i)^2$$

Zur Berechnung des Brier Scores gibt es eine mathematische Alternative, die diesen in drei Komponenten unterteilt [26]: *Uncertainty (UNC)*^{iv}, *Reliability (REL)*^v und *Resolution (RES)*^{vi}. Unabhängig von der Berechnung, kann der Brier Score als Kennzahl zur Messung von Overconfidence anhand der erhobenen Studie nicht genutzt werden, da weder Entscheidungsfragen mit ausschließlich Ja-Nein-Antworten gestellt werden, noch eine Einschätzung bezüglich der Sicherheit der Richtigkeit der eigenen Antwort erfragt wird.

Eine andere Möglichkeit zur Quantifizierung des Ausmaßes an Overconfidence ergibt sich bei Betrachtung der psychologischen Definition von Overconfidence. Die Psychologie unterscheidet zwei Kategorien von Overconfidence (Underconfidence): Zum einen die *absolute* Overconfidence (Underconfidence) und zum anderen die *relative* Overconfidence (Underconfidence). Die absolute Overconfidence betrifft die Einschätzung von Leistungen der Probanden auf individueller Ebene (within-subjects) und die relative Overconfidence betrifft den Vergleich von Einschätzungen von Leistung unter Probanden (between-subjects).

Die relative Overconfidence bezieht sich auf den Vergleich der eigenen Leistungen und Fähigkeiten mit den Leistungen und Fähigkeiten anderer. Moore und Healy (2008) bezeichnen dieses Phänomen als *Overplacement* während Svenson (1981) den Begriff *Better-than-average-Effect (BTA)* geprägt hat. Laut einer Forsa-Umfrage gaben beispielsweise 95 Prozent der befragten Manager an, sich für einen überdurchschnittlich geeigneten und allseits respektierten Vorgesetzten zu halten [13, WirtschaftsWoche]. Nach Svenson (1981) sei die Richtung des Better-than-average-Effects situationsabhängig: Ist die gegebene Situation für eine Person vertraut und die Fragestellung wird als (überraschend) leicht empfunden, so tendiert diese Person im Durchschnitt zur Überschätzung der eigenen Fähigkeiten relativ zu den anderen, bzw. die Fähigkeiten der anderen werden verhältnismäßig schlechter eingeschätzt [24]. Moore und Healy bezeichnen die übermäßig positive relative Selbsteinschätzung als *Overplacement*, Svenson hingegen als *Better-than-average (BTA)*. Ist die Situation dagegen ungewohnt oder die Fragestellung wird als schwierig wahrgenommen, so tendiert diese Person eher zur Unterschätzung der eigenen Fähigkeiten im Vergleich zu anderen, bzw. werden die Fähigkeiten anderer als verhältnismäßig besser bewertet [24]. Entsprechend definieren Moore und Healy die übermäßig konservative Selbsteinschätzung als *Underplacement* [25], während Svenson diese als *Worse-than-Average (WTA)* bezeichnet. Moore und Cain (2007) sehen die unterschiedliche, relative Selbsteinschätzung als eine Folge des Schwierigkeitsgrades der Aufgabe und als natürliche Konsequenz asymmetrischer Informationsverteilung: Die einschätzende Person verfügt per Definition über mehr Informationen zu sich selbst als zu anderen. Die Einschätzung

ⁱⁱEntscheidungsfragen sind Fragen, die nur mit Ja oder Nein zu beantworten sind.

ⁱⁱⁱWäre das Wahrscheinlichkeitsurteil $p_i < 0,5$, so würde sich der Proband nicht für die zugehörige Antwort entscheiden, sondern für die jeweils andere und wäre sich seiner Antwort zu $(1 - p_i)$ Prozent sicher.

^{iv}Uncertainty misst die inhärente Unsicherheit bezüglich des Eintritts des Ereignisses, zum Beispiel die Unsicherheit, ob die Frage richtig beantwortet.

^vReliability (Reliabilität) misst inwiefern die a priori vorhergesagten Wahrscheinlichkeiten mit den realen Eintrittswahrscheinlichkeiten übereinstimmen.

^{vi}Resolution misst inwiefern die verschiedenen a priori vorhergesagten (konditionalen) Wahrscheinlichkeiten von deren Durchschnitt abweichen.

der Eigenleistung ist somit fundierter aber dennoch vorsichtig; die Einschätzungen der Leistungen anderer sind dabei noch vorsichtiger [24]. Laut Moore und Healy (2008) sind Selbsteinschätzungen somit regressiv; die Einschätzungen anderer hingegen in höherem Maße.

Eine weitere Facette der relativen Overconfidence (Underconfidence) im weiteren Sinne ist *Overprecision* [25]. Overprecision betrifft die systematische Überschätzung der eigenen Wahrscheinlichkeitsurteile bezüglich des Eintretens zukünftiger Ereignisse und in diesem Zuge, die systematische Unterschätzung des Konfidenzintervalls zum gegebenen Signifikanzniveau. Liegt Overprecision vor, so vertraut die jeweilige Person übermäßig in die Richtigkeit des eigenen Urteils. Sie geht also davon aus, dass die eigene Einschätzung einen geringen Fehler aufweist und setzt entsprechend die Intervallgrenzen um diese Punktschätzung zu eng an.

Betrifft die Einschätzung der eigenen Fähigkeiten lediglich die eigene Person, im Sinne eines Vorher-Nachher-Vergleichs oder einer absoluten Prognose ohne Vergleiche zu ziehen oder Wahrscheinlichkeiten anzugeben, so spricht man von *Overestimation*. Overestimation bezeichnen Moore und Healy (2008) als die absolute Überschätzung eigener Fähigkeiten im Vergleich zu der tatsächlich erreichten Leistung. Empirische Studien zeigen, dass insbesondere das Ausmaß der Overestimation sensitiv auf den Schwierigkeitsgrad der zu bewertenden Aufgabe reagiert [17]: Je schwieriger die gestellten Aufgaben erscheinen oder tatsächlich sind, desto schlechter schätzen sich Studierende selbst ein und umgekehrt. Dieses Phänomen wird als *Hard-Easy-Effect* bezeichnet und gilt über verschiedene Versuchsdesigns^{vii} hinweg als robust.

Moore und Healy (2008) zeigen, dass es wichtig ist, zwischen den verschiedenen Facetten der Overconfidence zu unterscheiden und die Daten vor diesem Hintergrund zu analysieren und interpretieren. Falls eine solche Unterscheidung in Anbetracht der Fragesituation zwar möglich ist, aber nicht erfolgt, so führt dies laut Moore und Healy (2008) zu Inkonsistenzen der Analyse.

Diese Arbeit betrachtet lediglich die Overestimation, also den durchschnittlichen Grad der Fehleinschätzung von *Finance- & Accounting-* und *Management-*Studierenden im Zeitablauf eines Semesters anhand dreier Übungsklausuren und einer Abschlussklausur. Eine andere vorgestellte Kennzahl zur Messung des Confidence-Level kann auf Basis des zugrundeliegenden Datensatzes nicht genutzt werden.

3 Relevanz der Mathematik im Studium der Wirtschaftswissenschaften

Der in Regelstudienzeit sechs Semester umfassende Studiengang *Bachelor of Science in Wirtschaftswissenschaften* der Goethe Universität Frankfurt besteht aus einem zweisemestrigen Orientierungsabschnitt und einem viersemestrigen Qualifizierungsabschnitt. Die Module des Orientierungsabschnitts dienen der Vermittlung von Grundlagen innerhalb von maximal drei Semestern. Nach Abschluss dieser Phase entscheiden sich die Studierenden entweder für den volkswirtschaftlichen oder betriebswirtschaftlichen Studienschwerpunkt. Im dritten und vierten Semester werden entsprechend entweder volkswirtschaftliche und ökonometrische oder betriebswirtschaftliche Inhalte vertieft. Nach Abschluss des vierten Semesters entscheiden sich die Studierenden des betriebswirtschaftlichen Studienschwerpunkts für eine weitergehende Spezialisierung entweder im Bereich *Finance & Accounting* oder *Management*.

Die Vergabe der Studienplätze erfolgt zu 80 Prozent nach einem universitätseigenen Qualifizierungsschlüssel und zu 20 Prozent nach längster Wartezeit. Der Qualifizierungsschlüssel ist eine Punktzahl, die sich zu 80 Prozent aus der Durchschnittsnote der Hochschulzugangsberechtigung und zu 20 Prozent aus der Durchschnittsnote der zwei letzten Schulnoten im Fach *Mathematik* zusammensetzt.

Entsprechend ist Mathematik eine essentielle Grundlage für das wirtschaftswissenschaftliche Studium. Auch allgemein gilt Mathematik als wichtiger Einflussfaktor der Studierfähigkeit von Abiturienten. Laut Bescherer (2003) beschreibt die Studierfähigkeit Kenntnisse, Fertigkeiten, Fähigkeiten, Haltungen und Einstellungen, die zur erfolgreichen Absolvierung eines bestimmten Studiengangs notwendig sind. In diesem Zusammenhang gilt der Konferenz

^{vii}Das Versuchsdesign wurde unter anderem hinsichtlich der Anordnung der Aufgaben, der Art der Bewertung und der monetären Anreize variiert.

mathematischer Fachbereiche zufolge die Schulmathematik als entscheidender Einflussfaktor der Studierfähigkeit von Abiturienten und insbesondere von Interessierten mathematikbasierter Studienfächer wie die Wirtschafts- oder Sozialwissenschaften [3]. Es ist laut Bescherer (2003) nicht nur entsprechendes mathematisches Fachwissen vorausgesetzt, sondern es wird auch die allgemeine Studierfähigkeit durch die Lehre der Mathematik gefördert, in erster Linie das Abstraktionsvermögen, die Konzentrationsfähigkeit, das logisch konsequente Denken und strukturierte Arbeitstechniken.

In dem beschriebenen Studiengang werden bereits im ersten Semester in den Veranstaltungen *Mathematik für Wirtschaftswissenschaftler* und *Statistik* quantitative Grundlagen vermittelt. Aufgrund des erfahrungsgemäß heterogenen, mathematischen Leistungsniveaus wurde im Sommersemester 2005 erstmals ein mathematischer Brückenkurs für Erstsemester-Studierende der Wirtschaftswissenschaften angeboten. Diese Einführung erfolgte im Rahmen des *Bologna Prozesses*, der eine europaweite Vereinheitlichung und Vergleichbarkeit der Hochschulabschlüsse durch die Einführung des Bachelor-Master-Systems verfolgt. Seither findet der mathematische Brückenkurs jedes Semester unmittelbar vor Semesterbeginn statt. Die Teilnahme an dem inzwischen achttägigen Kurs ist zwar fakultativ, wird seitens des Fachbereichs zur Wiederholung und Homogenisierung der Leistungsniveaus jedoch sehr empfohlen. Der durchschnittliche Anteil teilnehmender Studierender der letzten Semester zeigt zum einen eine positive Resonanz und zum anderen, dass sich Studierende ihrer mathematischen Defizite und der Wichtigkeit, diese abzubauen, bewusst sind.

4 Hypothesen und Methoden

Wie in Abschnitt 2 erläutert, zeigen Erfahrungen aus der Praxis, dass Selbstüberschätzung ein häufiges Problem darstellt. Nicht nur unter Absolventen, die gerade ihre Karriere beginnen, sondern auch auf höheren Karrierestufen ist das Phänomen nicht unbekannt. Hier wird der Frage nachgegangen, ob sich Studierende der Wirtschaftswissenschaften, insbesondere der Schwerpunkte *Finance & Accounting* und *Management* überschätzen. Dabei wird untersucht, ob eine vermeintliche Überschätzung bereits zu Studienbeginn vorherrscht oder sich vermutlich erst im Laufe des ersten Semesters herausbildet. Unter anderem zu diesem Zweck wurde eine empirische Studie unter Studienanfängern in der Mathematikveranstaltung *Mathematik für Wirtschaftswissenschaften* des Sommersemesters 2015 durchgeführt, deren Ablauf und Inhalt in Abschnitt 5 erläutert wird.

Im vorigen Abschnitt wird darauf eingegangen, welchen Stellenwert die Mathematik in dem wirtschaftlichen Studiengang hat, nicht nur in Bezug auf die Kenntnisse mathematischer Inhalte, sondern auch auf das miteinhergehende logische und strukturierte Denken. Aus diesem Grund bezieht sich die zentrale Forschungsfrage auf die Analyse der *Mathematical Confidence*, also die Einschätzung eigener Mathematikfähigkeiten im Vergleich zu der tatsächlichen Leistung. In diesem Zusammenhang werden insbesondere Determinanten identifiziert, die das Einschätzungsvermögen und damit das Confidence-Level der Studierenden beeinflussen und bestimmen.

Insgesamt werden die folgenden Hypothesen dazu in den nächsten Abschnitten deskriptiv und analytisch untersucht:

- Hypothese 1 Studierende der Wirtschaftswissenschaften überschätzen sich.
- Hypothese 2 *Finance- & Accounting*-Studierende überschätzen ihre mathematischen Fähigkeiten eher als *Management*-Studierende.

In 6.2 wird über Mittelwertvergleiche untersucht, ob sich Studierende der beiden Schwerpunkte *Finance & Accounting* und *Management* hinsichtlich ihres Ausmaßes an Selbstüber- bzw. unterschätzung im Zeitablauf des Semesters unterscheiden. Die Mittelwertvergleiche werden mithilfe des nicht-parametrischen Äquivalents zum Zweichstichproben-t-Tests, anhand des Rangsummentests nach Wilcoxon, durchgeführt. Im Anschluss untersuchen multiple OLS-Regressionen Determinanten dieser vermeintlichen Gruppenunterschiede, diesbezüglich insbesondere den Semesterbeginn mit dem Semesterende vergleichend.

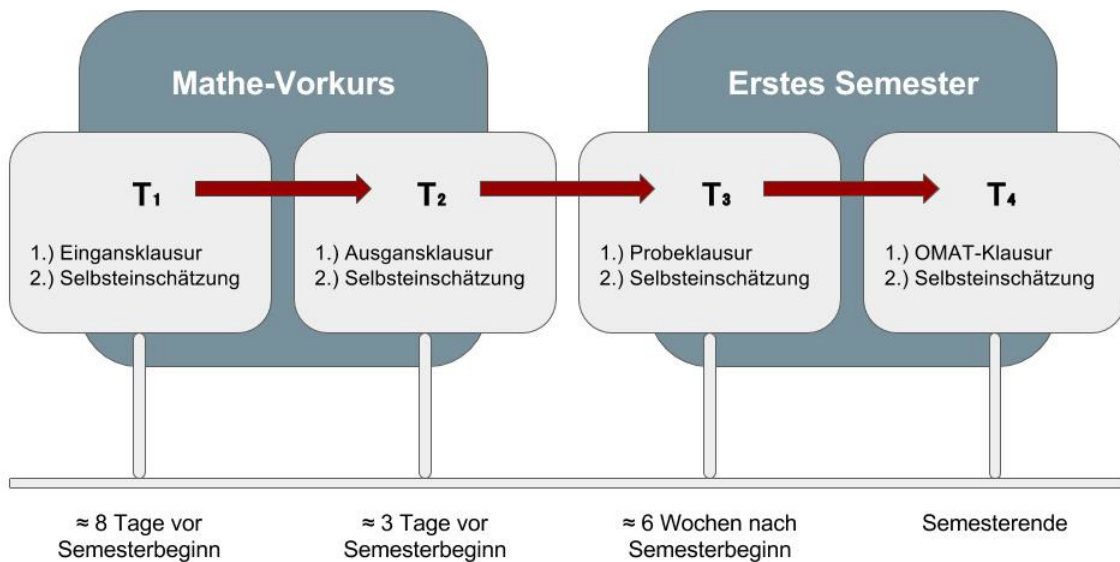
5 Daten

5.1 Datenerhebung

Bezüglich der mathematischen Studierfähigkeit und in diesem Kontext des Konzepts des Mathe-Vorkurses wurde im Sommersemester 2015 im Rahmen eines umfangreichen Forschungsprojekts eine Evaluation in dem mathematischen Brückenkurs vor Semesterbeginn und weiterführend in der quantitativen Einführungsveranstaltung OMAT im ersten Semester durchgeführt.^{viii}

Die Evaluation umfasst drei Fragebögen zur Erhebung von soziodemographischen Angaben der Studierenden. Diese enthalten auch Fragen zu Studienzielen, der Studienzufriedenheit, dem Studierverhalten, der Klausurvorbereitung, der Bildungsmobilität und der Studierfähigkeit. Zusätzlich zu den Fragebögen wurden die Studierenden gebeten, zu den drei Übungsklausuren und der Abschlussklausur OMAT eine Einschätzung in Form einer Punktzahl abzugeben, die die eigene erwartete Leistung bei bekannter Höchstpunktzahl widerspiegelt. Die Selbsteinschätzungen der drei Übungsklausuren wurden jeweils unmittelbar nach Bearbeitung der Klausur und entsprechend vor Bekanntgabe der jeweiligen tatsächlich erreichten Punktzahl abgegeben.

Abbildung 1: Ablauf der Datenerhebung



Die erste Übungsklausur, die *Eingangsklausur des Mathe-Vorkurses* (T₁), wurde vor Semesterbeginn an dem ersten Tag des Mathe-Vorkurses geschrieben und diente sowohl den Studienanfängern, als auch den Dozenten als Orientierung über den Status-Quo der Mathematik-Kenntnisse. Diese Klausur umfasste die für das Studium der Wirtschaftswissenschaften relevanten mathematischen Inhalte der gymnasialen Mittel- und Oberstufe, welche im folgenden Mathe-Vorkurs zur Homogenisierung der Leistungsniveaus wiederholt wurden. Die Selbsteinschätzung in der Eingangsklausur erfolgt nur bedingt auf rationaler Basis, da sie lediglich die in der Schul- und gegebenenfalls Ausbildungszeit gewonnenen Erfahrungen widerspiegelt [3, S. 130]. Auch zum Abschluss des Mathe-Vorkurses wurde eine Übungsklausur, die *Abschlussklausur des Mathe-Vorkurses* (T₂), gestellt. Diese sollte den Lernfortschritt im Zuge des Mathe-Vorkurses aufzeigen und gegebenenfalls auf Defizite hinweisen.

^{viii}Erste Ergebnisse der Studie wurden am „Tag der Lehre“ im Rahmen des Starken Start präsentiert [23]. Auf Basis des Datensatzes dieser Studie erfolgen die Auswertungen dieser Arbeit.

Die dritte Übungsklausur, die *Probeklausur OMAT* (T_3), wird in der sechsten Vorlesungswoche unmittelbar nach Abschluss des Kapitels „Lineare Algebra“ gestellt. In dieser wird das Wissen der in der Vorlesung besprochenen Inhalte überprüft; die Themen des Mathe-Vorkurses werden implizit vorausgesetzt. Die Probeklausur dient der eigenen Leistungsüberprüfung und bietet den Studierenden erstmals die Möglichkeit eine Klausur im großen Hörsaal mit durchschnittlich 600 weiteren Studierenden zu schreiben. Auch nach Bearbeitung der Probeklausur werden die Studierenden gebeten, eine Selbsteinschätzung abzugeben. Jede Übungsklausur wurde von studentischen Tutoren korrigiert, falls dies von den Studierenden erwünscht war.

Lediglich die Selbsteinschätzung der finalen *OMAT-Klausur* (T_4) wurde bereits in der letzten Vorlesungswoche abgegeben. Die Studierenden haben die Klausur zu diesem Zeitpunkt noch nicht bearbeitet, sodass sie deren Schwierigkeitsgrad nicht in ihre Selbsteinschätzung einbeziehen konnten. Entsprechend schätzten sich die Studierenden vor der OMAT-Klausur ausschließlich auf Grundlage gesammelter Erfahrungen aus den Übungsklausuren und auf Basis des zu dem Zeitpunkt aktuellen Lernstandes ein. Schließlich wurden nach Semesterende die Abschlussklausuren in den Einführungen *Mathematik für Wirtschaftswissenschaftler* (OMAT) und *Statistik* geschrieben.

Die Teilnahme an der Studie und die Bearbeitung von Fragebögen und der Übungsklausuren erfolgten auf freiwilliger Basis. In der Studie wurden ausschließlich Studierende der Wirtschaftswissenschaften befragt, die das Einführungsmodul OMAT noch nicht erfolgreich absolviert hatten. Die Auswahl der Population wurde bewusst getroffen: Zur Analyse, ob bereits zu Studienbeginn die Tendenz zur Selbstüberschätzung bei wirtschaftswissenschaftlichen Studierenden vorhanden ist, bezieht alleinig Studierende ein, die sich am Anfang ihres wirtschaftlichen Studiums befinden.

ix

Da die Befragungen und die Teilnahme an den Übungsklausuren auf freiwilliger Basis stattfanden, gab es pro Messzeitpunkt unterschiedliche Teilnehmerzahlen. Die verschiedenen Messzeitpunkte der Studie $T_1 - T_4$ werden im folgenden separat betrachtet und verglichen.

5.2 Datenauswertung

Die Fragebögen wurden mithilfe der Evaluations- und Umfragesoftware „EvaSys“ erstellt und online ausgewertet. Die Ergebnisse wurden automatisch in Excel-Dateien exportiert. Dieses Verfahren stellte sich als weitestgehend unproblematisch heraus, lediglich einzelne Antworten mussten aufgrund von Unleserlichkeiten seitens der Studierenden in EvaSys verifiziert werden.

Die drei Übungsklausuren und die Klausur in OMAT wurden zunächst von studentischen Hilfskräften und Tutoren korrigiert und anschließend über „EvaExam“, einer Software zur automatischen Prüfungsauswertung, ausgewertet. Die Ergebnisse wurden auch hier in Excel-Dateien exportiert.

In der Vorbereitung des Datensatzes für weitergehende Analysen werden zwei Arten von fehlenden Daten (Missing Data, Missings) unterschieden: Falls die Studierenden an der jeweiligen Klausur oder Befragung zu einem Messzeitpunkt nicht teilgenommen haben, beispielsweise aufgrund von Verweigerung oder Nichterreichbarkeit, werden die resultierend fehlenden Daten als *vollständiger Stichprobenausfall* bzw. *Antwortausfall* (*Unit Non-Response*, [11, S. 331]) bezeichnet. Fehlen hingegen nur einzelne Untersuchungsmerkmale bzw. wurden aus verschiedenen Gründen die Fragen nur teilweise beantwortet, beispielsweise aufgrund von fehlendem Wissen oder partieller Verweigerung, so werden diese fehlenden Daten *fehlende Werte* bzw. *partielle Antwortausfälle* (*Item Non-Response*, [11, S. 341]) genannt. Weder die Antwortausfälle noch die fehlenden Werte wurden in dem zugrundeliegenden Datensatz rekonstruiert, um eine Verzerrung zu vermeiden.

^{ix}Studierende, die sich im Zweit- oder Drittversuch des Moduls OMAT befinden, gelten hier annahmegemäß auch als Studienanfänger, da sich diese noch in der Orientierungsphase des Studiums befinden (s. Absatz 3).

5.3 Items und Skalen

Für die Untersuchung von Mathematical Confidence bei den Schwerpunkten *Finance & Accounting* und *Management* werden demographische Daten der Studierenden, die Abiturleistungen, Selbsteinschätzungen zu den Leistungen und die Leistungen selbst in dem Mathe-Vorkurs und der OMAT analysiert. Die übrigen Informationen, die aus der Studie hervorgehen, werden aufgrund der speziellen Forschungsfrage der Arbeit aus der Analyse ausgeschlossen.

In Tabelle 1 werden die betrachteten Items in einer Variablenübersicht hinsichtlich deren Skalierung und möglichen Ausprägungen beschrieben. Der Index t bezeichnet den Messzeitpunkt, in dem die jeweilige Ausprägung der Variable gemessen wurde. Binär definierte Dummy-Variablen erhalten im folgenden das Präfix D .

Tabelle 1: Übersicht relevanter Variablen und deren Eigenschaften

Variable	Abkürzung	Skalierung	Ausprägungen
Mathematical Confidence	$MCON_t$ $\forall t \in \{1; 4\}$	Metrisch	(1): [0; 90] (2): [0; 75] (3): [0; 120]
Schwerpunkt	$DMANFIN$	Nominal	{Finance & Accounting = 1; Management = 0}
Geschlecht	$DGENDER$	Nominal	{weiblich = 1; männlich = 0}
Geburtsjahr	$BIRTH$	Metrisch	
Migrationshintergrund	$DMIGR$	Nominal	{ja = 1; nein = 0}
Studierendenstatus	$DFIRST$	Nominal	{Erstsemester: ja = 1; Erstsemester: nein = 0}
Note der Hochschul- Zugangsberechtigung (Abitur-Note)	ABI	Metrisch	[1, 0; 4, 0]
Abitur-Note Mathematik (in Punkten)	$MABI$	Metrisch	[5; 15]
Selbsteinschätzung (in Punkten)	SE_t $\forall t \in \{1; 4\}$	Metrisch	(1): [0; 90] (2): [0; 75] (3): [0; 120]
Klausur-Ergebnis (in Punkten)	RES_t $\forall t \in \{1; 4\}$	Metrisch	(1): [0; 90] (2): [0; 75] (3): [0; 120]
Erfahrung erfolgreiche / missglückte Selbsteinschätzung T_{t-1}	$DGOODNEWS_{t-1}$ $\forall t \in \{2, 4\}$	Nominal	{Erfolg = 1; Enttäuschung = 0} (1)=Mathe-Vorkurs-Klausuren (2)=Probeklausur (3)=OMAT-Klausur

Außerdem werden die Variablen *Schwerpunkt*, *Migrationshintergrund*, *Studierendenstatus*, die *Selbsteinschätzungen* und die *Klausur-Ergebnisse* pro Messzeitpunkt, sowie die *Mathematical Confidence* folgendermaßen konstruiert:

Schwerpunkt ($DMANFIN$)

$DMANFIN$ stellt eine Variable dar, die den Wert eins annimmt, falls der Studierende den voraussichtlichen Schwerpunkt *Finance & Accounting* angibt, oder den Wert null annimmt, falls der Studierende voraussichtlich *Management* studieren wird. Diese beiden Spezialisierungen sind innerhalb des betriebswirtschaftlichen Schwerpunktes wählbar

(s. Abschnitt 3). Hierbei ist der Zusatz „voraussichtlich“ maßgebend, denn die Studierenden werden zwar in der Befragung bereits in ihrem ersten Fachsemester nach der favorisierten Schwerpunktwahl gefragt, die endgültige Entscheidung zwischen betriebs- oder volkswirtschaftlichem Schwerpunkt fällt jedoch erst zu Beginn des dritten Semesters. Die Wahl des spezialisierenden Schwerpunktes *Finance & Accounting* gegenüber *Management* wird erst im fünften Semester getroffen. Weil nur sehr wenige Daten zu Studierenden, die voraussichtlich den volkswirtschaftlichen Schwerpunkt wählen^x, vorhanden sind, werden diese von der vorliegenden Untersuchung ausgeschlossen, da eine statistisch valide Inferenz aufgrund zu geringer Stichprobengröße nicht möglich wäre.

Migrationshintergrund (*DMIGR*)

DMIGR stellt eine Variable dar, die den Wert eins annimmt, falls der Studierende einen Migrationshintergrund aufweist, oder den Wert null annimmt, falls der Studierende keinen Migrationshintergrund aufweist. Als Studierende mit Migrationshintergrund gelten diejenigen, die entweder eine ausländische oder eine doppelte Staatsangehörigkeit besitzen oder bei denen mindestens ein Elternteil im Ausland geboren wurde.

Studierendenstatus (*DFIRST*)

DFIRST ist binär kodiert. Für den Fall, dass der Studierende ein tatsächlicher Erstsemester ist, nimmt die Variable den Wert eins an, sonst nimmt sie den Wert null an.

Relative Selbsteinschätzung (*relSE_t*) und relatives Klausurergebnis (*relRES_t*)

Die Selbsteinschätzungen (*SE_t*) und die Klausurergebnisse (*RES_t*) sind metrisch skalierte Variablen, die durch Angabe einer Punktzahl quantifiziert werden. Weil in den verschiedenen Klausuren zu den Zeitpunkten $t \in \{1; 4\}$ unterschiedliche Höchstpunktzahlen erreichbar waren, ist es nicht sinnvoll, die absoluten Punktzahlen in der Analyse zu vergleichen, sondern Verhältnisse zu bilden. Die individuelle Selbsteinschätzung pro Klausur wird im Verhältnis zur maximal möglichen Gesamtpunktzahl der jeweiligen Klausur betrachtet, was der relativen Selbsteinschätzung (*relSE_t*) entspricht. Entsprechendes gilt für die Klausurergebnisse (*relRES_t*). So sind die Klausuren zwar bezüglich der Selbsteinschätzungen und Ergebnisse vergleichbar, allerdings verzerren weiterhin die unterschiedlichen Schwierigkeitsgrade der Klausuren den Vergleich. Bevor der Lösungsansatz zur Milderung dieser Verzerrung dargestellt wird, wird auf die letzte relevante Variable eingegangen.

Erfahrung einer erfolgreiche/ missglückte Selbsteinschätzung in $T - 1$ (*DGOODNEWS_{T-1}*)

DGOODNEWS_{T-1} ist eine binär definierte Variable, die den Wert eins annimmt, falls in der zuvor geschriebenen Klausur eine bessere Leistung erbracht wurde als in der Selbsteinschätzung erwartet. Im Fall einer schlechteren Leistung als erwartet, nimmt die Variable den Wert null an. *DGOODNEWS_{T-1}* zeigt die Erfahrung, die der jeweilige Studierende in der vorigen Klausur bezüglich seiner Kalibrierung der Selbsteinschätzung gemacht hat.

Mathematical Confidence (*MCON_t*)

Die Mathematical Confidence (*MCON_t*) ist eine kontinuierliche Variable, die das Verhältnis aus Selbsteinschätzung und Klausurergebnis pro Messzeitpunkt $t \in \{1; 4\}$ abbildet. Pro Studierenden i und Klausur in T wird der Quotient aus der jeweiligen Selbsteinschätzung und dem zugehörigen Klausurergebnis gebildet.

$$MCON_{i,t} = \frac{\text{Selbsteinschätzung}_{i,t}}{\text{Klausurergebnis}_{i,t}} = \frac{SE_{i,t}[\text{Pkt}]}{ERG_{i,t}[\text{Pkt}]}$$

Nimmt die Mathematical Confidence einen Wert größer eins an, so ist die Selbsteinschätzung im Verhältnis zu dem tatsächlichen Klausurergebnis größer. Folglich liegt Overconfidence vor und der Studierende schätzt seine mathema-

^xLediglich 14 von 274 Studierenden geben an, voraussichtlich den volkswirtschaftlichen Schwerpunkt zu wählen.

tischen Fähigkeiten zu optimistisch ein. Liegt allerdings ein Wert kleiner eins vor, so ist die Selbsteinschätzung im Vergleich zu dem tatsächlichen Klausurergebnis zu pessimistisch. Es liegt Underconfidence vor und der Studierende bewertet seine mathematischen Fähigkeiten schlechter als sie sind. Falls die Mathematical Confidence einen Wert um eins annimmt, so entspricht die Selbsteinschätzung dem tatsächlichen Klausurergebnis. Es liegt weder Under- noch Overconfidence vor, sondern das Confidence-Level unterliegt *perfekter Kalibrierung* (*perfect calibration*, vgl. [31]) . Der Studierende verfügt sodann über ein akkurates Selbstbild bezüglich seiner mathematischen Fähigkeiten und die Fähigkeit zur realistischen Selbstreflexion. In 6.1 wird die Mathematical Confidence pro Schwerpunkt und Messzeitpunkt $t \in \{1; 4\}$ graphisch dargestellt und beschrieben.

5.4 Trimmung des Datensatzes

Zur Milderung der Verzerrung durch die marginal verschiedenen Schwierigkeitsgrade der Klausuren, wurde der Datensatz bezüglich der Mathematical Confidence $MCON_t$ und den absoluten Klausurergebnissen ERG_t um 1 Prozent getrimmt. Die Trimmung nach Variable ERG_t eliminiert nicht-repräsentative Ausreißer und schließt diejenigen Studierenden aus, die die Höchstpunktzahlen erreicht haben. Dies ist bei der Analyse von Overconfidence sinnvoll, da sich diese per Definition nicht überschätzen können (vgl. [27]). Es ist zu beachten, dass die relative Selbsteinschätzung nach wie vor im Verhältnis zur maximal erreichbaren Höchstpunktzahl gemessen wird, da angenommen wird, dass die Studierenden sich bewusst sind, dass sie bei einer fairen Klausur die Höchstpunktzahl erreichen könnten.

Die zweite Trimmung nach $MCON_t$ schließt diejenigen Studierenden aus, die ein vergleichsweise hohes Confidence-Level $MCON_t$ besitzen. Ergo weisen diese Studierende eine hohe Diskrepanz zwischen Selbsteinschätzung und Klausurergebnis auf: Entweder die Selbsteinschätzung ist unverhältnismäßig hoch bei gegebenem Klausurergebnis oder das Klausurergebnis ist unverhältnismäßig schlecht bei gegebener Selbsteinschätzung.

Insgesamt wurden durch die doppelte Trimmung 37 von 743 Beobachtungen aus der Analyse ausgeschlossen. Dabei ist wichtig, dass die Verzerrungen durch die einerseits verschiedenen Schwierigkeitsgrade der Klausuren und andererseits außerordentliche Fehleinschätzung der eigenen Leistung im Zuge der Trimmung zwar abgeschwächt, allerdings nicht eliminiert werden können. Da jegliche deskriptive und statistische Analysen auf Basis des doppelt getrimmten Datensatzes durchgeführt werden, sind die Ergebnisse vor diesem Hintergrund vorsichtig zu interpretieren.

6 Deskriptive Statistiken

In Tabelle 3 werden die für die Auswertungen relevanten Variablen aus Tabelle 1 deskriptiv beschrieben. Es erfolgt eine separate Betrachtung nach Schwerpunkten. Insgesamt haben 274 Studierende eine Schwerpunktwahl abgegeben, darunter entschieden sich 60 Studierende für den Schwerpunkt *Finance & Accounting* und 71 Studierende für den Schwerpunkt *Management*^{xi}. Betrachtet man die einzelnen Beobachtungsmerkmale, so fällt auf, dass weibliche Studierende durchschnittlich häufiger den Schwerpunkt *Management* als *Finance & Accounting* wählen (Mittelwertvergleich: 0,6333 vs. 0,5070), was der Median zusätzlich verdeutlicht. Die *Management*-Studierenden sind durchschnittlich jünger (Mittelwertvergleich: 1992,7670 vs. 1993,9300) und weisen wesentlich seltener einen Migrationshintergrund auf (Mittelwertvergleich: 0,5135 vs. 0,3636), auch dies bestätigt wiederum der Median. Bezüglich der Leistungen im Abitur unterscheiden sich die Studierenden der beiden Schwerpunkte nur marginal: *F-&A*-Studierende schnitten sowohl im Abitur (Mittelwertvergleich: 2,0500 vs. 2,1000), als auch im Mathe-Abitur (Mittelwertvergleich: 10,9032 vs. 10,7073) etwas besser ab. Allerdings schwanken die Leistungen der *Management*-Studierenden im Abitur und im Mathe-Abitur weniger stark als unter den *F-&A*-Studierenden. Auf die Confidence-Level $MCON_1 - MCON_4$ der Schwerpunkte über die Messzeitpunkte wird im folgenden Abschnitt eingegangen.

^{xi}Von den übrigen 143 Studierenden haben lediglich 14 den volkswirtschaftlichen Schwerpunkt gewählt, weshalb dieser von der Analyse ausgeschlossen wurde. Die verbleibenden 129 Studierende haben die Angabe gemacht, dass sie bezüglich ihrer Schwerpunktwahl noch unentschieden sind.

Tabelle 3: Deskriptive Statistiken der relevanten Variablen

Variable	Abkürzung	Finance & Accounting				Management				
		Abs. Hfk.	Arithm. Mittel	Median	SE	Abs. Hfk.	Arithm. Mittel	Median	SE	
Kriterium	Mathematical Confidence	36	0,8611	0,8349	0,3009	41	0,9689	1,0311	0,4650	
	Mathematical Confidence	37	0,8755	0,8660	0,2383	53	1,0429	0,9936	0,4491	
	Mathematical Confidence	57	0,9052	0,8842	0,3141	66	0,9168	0,8727	0,3142	
	Mathematical Confidence	32	1,1118	1,0142	0,3922	40	1,2471	1,0785	0,4618	
Prädiktoren	Geschlecht	60	0,6333	0	0,4860	71	0,5070	1	0,5035	
	Geburtsjahr	60	1992,7670	1994	4,1675	71	1993,93	1994	1,8618	
	Migrationshintergrund	37	0,5135	1	0,5067	55	0,3636	0	0,4855	
	Studienabschnitt	58	0,6552	1	0,4795	70	0,6857	1	0,4676	
	Note der Hochschulzugangsberechtigung	34	2,0559	2,0500	0,6066	51	2,1039	2,1	0,3821	
	Note im Mathe-Abitur (in Punkten)	31	10,9032	11	2,7490	41	10,7073	11	2,5223	
	Erfahrung erfolgreiche/missglückte Selbsteinschätzung T ₁	DGOODNEWS _{1,1}	44	0,4545	0	0,5037	48	0,375	0	0,4892
	Erfahrung erfolgreiche/missglückte Selbsteinschätzung T ₃	DGOODNEWS ₃	58	0,6207	1	0,4895	69	0,6377	1	0,4842

6.1 Das Kriterium $MCON_t$

Wie in 5.3 bereits erwähnt, wurden innerhalb der Messzeitpunkte $t \in \{1; 4\}$ insgesamt drei Übungsklausuren und die OMAT-Klausur geschrieben, sodass die Studierenden, im Falle vollständiger Teilnahme, vier Selbsteinschätzungen abgeben und vier Klausurergebnisse erhalten. Um die Entwicklung der Confidence-Level $MCON_1 - MCON_4$ der betrachteten Schwerpunkte über die Messzeitpunkte hinweg zu vergleichen, werden diese einander in dem folgenden Abschnitt gegenübergestellt.

In den folgenden Grafiken wird die auf der Abszisse befindliche relative Selbsteinschätzung dem zugehörigen, auf der Ordinate abgetragenen, relativen Klausurergebnis gegenübergestellt. Die winkelhalbierende Referenzlinie indiziert eine perfekte Kalibrierung (in Anlehnung an den Brier Score [5]), das heißt eine eingeschätzte Punktzahl von beispielweise 50 Punkten wird auch tatsächlich erreicht. Dabei ist wichtig, dass eine perfekte Kalibrierung bezüglich der Selbsteinschätzung eines Studierenden keine Aussage über die tatsächliche Leistung trifft. Befindet sich ein bestimmter Datenpunkt oberhalb der Winkelhalbierenden, so ist die zugehörige Selbsteinschätzung zu gering im Vergleich zu dem tatsächlich erreichten Klausurergebnis. Folglich schätzt der betreffende Studierende seine mathematischen Fähigkeiten schlechter ein als diese sind (*Underconfidence*). Liegt ein Datenpunkt hingegen unterhalb der Winkelhalbierenden, so ist das Confidence-Level des entsprechenden Studierenden zu hoch im Vergleich zu dem tatsächlich erreichten Klausurergebnis. Dieser Studierende schätzt seine mathematischen Fähigkeiten zu selbstbewusst ein (*Overconfidence*).

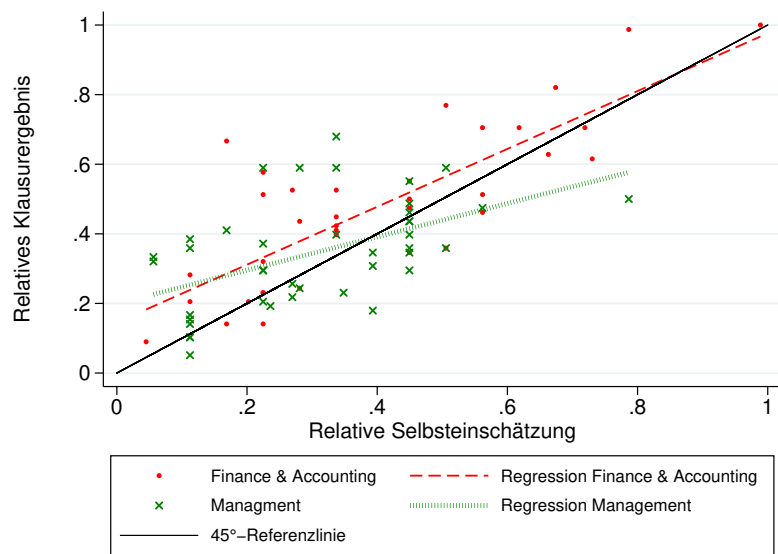
Die Regressionsgeraden fassen den Trend der Mathematical Confidence pro betrachtetem Schwerpunkt zusammen. Die zugehörigen Ergebnisse der einzelnen Regressionen sind in Tabelle 9 im Anhang A.2 zu finden.

6.1.1 Eingangsklausur Mathe-Vorkurs (T_1)

Abbildung 2 zeigt die Mathematical Confidence der beiden Schwerpunkte zum Zeitpunkt der Eingangsklausur des Mathe-Vorkurses T_1 . Die Leistung, die die Studierenden hier einschätzen und zeigen, besteht ausschließlich aus mathematischen Vorkenntnissen, beispielsweise aus Schulbildung oder bestehender Vorbildung.

Die gestrichelte Regressionsgerade der *Finance- & Accounting*-Studierenden befindet sich größtenteils oberhalb der Winkelhalbierenden. Erst bei einer selbsteingeschätzten Leistung von circa 90 Prozent der insgesamt erreichbaren Gesamtpunktzahl schneidet die Regressionsgerade die Winkelhalbierende. Schätzt also ein *Finance- & Accounting*-Studierender weniger als 90 Prozent der erreichbaren Gesamtpunktzahl zu erzielen, so unterschätzt er seine mathematischen Fähigkeiten im Durchschnitt. Somit befinden sich die Wertepaare des Streudiagramms der *F-&A*-Studierenden überwiegend oberhalb der Winkelhalbierenden, lediglich 10 der 36 Wertepaare liegen unterhalb der Winkelhalbierenden. Diese zehn Studierenden überschätzen sich. Die Regressionsgerade der *Management*-Studierenden hingegen verläuft wesentlich flacher und befindet sich zum Großteil unterhalb der Regressionsgerade der *F-&A*-Studierenden. Relativ zu *F-&A*-Studierenden schätzen die *Management*-Studierenden ihre Leistung also durchschnittlich besser ein. Ganz realistisch ist diese Einschätzung allerdings nicht, da sich Studierende, die angeben mindestens 40 Prozent der Gesamtpunktzahl zu erreichen, tendenziell überschätzen. Dies zeigt der Schnittpunkt der gepunkteten Regressionsgeraden mit der Winkelhalbierenden. Zudem ist auffällig, dass die Streuung der Wertepaare der Streudiagramme von den *F-&A*-Studierenden wesentlich breiter ist als die der *Management*-Studierenden: Während sich die Wertepaare der *F-&A*-Studierende über das Intervall $[0, 04; 1]$ erstrecken, befinden sich die Wertepaare der *Management*-Studierenden lediglich auf dem Intervall $[0, 05; 0, 8]$.

Abbildung 2: MCON₁, Eingangsklausur des Mathe-Vorkurses

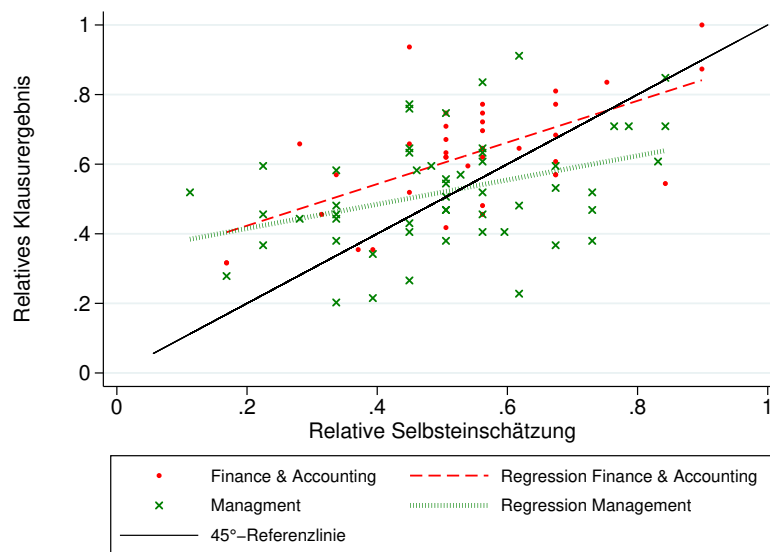


Basis für diese Grafik ist der nach MCON und ERG getrimmte Datensatz.

6.1.2 Ausgangsklausur Mathe-Vorkurs (T₂)

Abbildung 3 zeigt die Mathematical Confidence der Schwerpunkte zum Zeitpunkt der Abschlussklausur des Mathe-Vorkurses T₂. Im Vergleich zur Eingangsklausur des Mathe-Vorkurses ist anhand der Positionen der Wertepaare beider Schwerpunkte zu erkennen, dass sich die Leistung nach Vorkurs-Besuch verbessert, da sich die Punktelcke insgesamt vom Koordinatenursprung entlang der Winkelhalbierenden verschiebt. Dies äußert sich in einer Verschiebung der beiden Regressionsgeraden entlang der Ordinate, sodass die Studierenden durchschnittlich bei gleichbleibender relativer Selbsteinschätzung eine höhere Punktzahl erreichen. Die Steigung der Regressionsgeraden der *F-&A*-Studierenden verläuft flacher im Vergleich zu T₁ und schneidet die Winkelhalbierende bereits bei einer relativen Selbsteinschätzung von circa 78 Prozent (Vergleich T₁: 90 Prozent). Somit liegt im Durchschnitt eine Überschätzung seitens derjenigen *F-&A*-Studierenden vor, die vermuten, mehr als 78 Prozent der Gesamtpunktzahl erreichen zu können. Die Steigung der Regressionsgeraden der *Management*-Studierenden hat sich hingegen nicht sichtlich verändert. Lediglich der Schnittpunkt mit der Winkelhalbierenden hat sich verschoben, sodass eine Überschätzung der Leistung durchschnittlich erst bei einer relativen Selbsteinschätzung von circa 50 Prozent eintritt (Vergleich T₁: 40 Prozent).

Abbildung 3: MCON₂, Ausgangsklausur des Mathe-Vorkurses

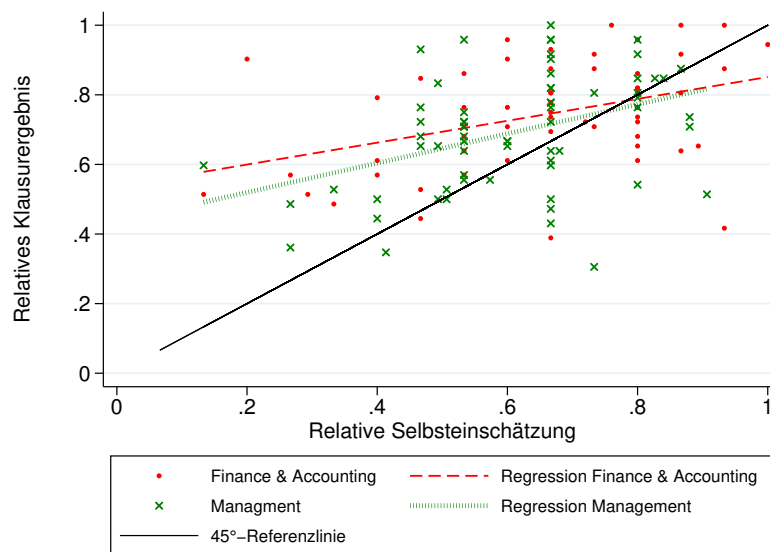


Basis für diese Grafik ist der nach MCON und ERG getrimmte Datensatz.

6.1.3 Probeklausur der OMAT-Vorlesung (T₃)

Abbildung 4 zeigt die Mathematical Confidence der Schwerpunkte zum Zeitpunkt der OMAT-Probeklausur T₄. Im Vergleich zu T₁ und T₂ haben sich die Wertepaare der beiden Schwerpunkte weiter von dem Koordinatenursprung entfernt. Es liegen vermehrt Wertepaare oberhalb der Winkelhalbierenden, sodass sich auch die Regressionsgeraden weiter entlang der Ordinate nach oben verschieben. Bei gleichbleibender Selbsteinschätzung zeigen die Studierenden also durchschnittlich eine höhere Leistung. Die Steigung der Regressionsgeraden der *F-&A*-Studierenden ist zwar vergleichsweise flacher geworden, befindet sich allerdings weiterhin oberhalb der Regressionsgeraden der *Management*-Studierenden. Die *F-&A*-Studierenden schneiden also bei gleicher Selbsteinschätzung durchschnittlich besser ab als die *Management*-Studierenden und unterschätzen sich gleichzeitig tendenziell. Da die Regressionsgeraden beider Schwerpunkte die Winkelhalbierende bei einer relativen Selbsteinschätzung von circa 80 Prozent schneiden, liegt eine durchschnittliche Überschätzung der Leistung erst bei einer erwarteten Leistung von 80 Prozent der Gesamtpunktzahl vor. Dies ist für *F-&A*-Studierende vergleichbar mit den vorherigen Messzeitpunkten; für *Management*-Studierende aber ein weiterer Fortschritt hin zur realistischen Selbsteinschätzung (Vergleich T₂: 50 Prozent und T₁: 40 Prozent).

Abbildung 4: MCON₃, Probeklausur der OMAT-Vorlesung

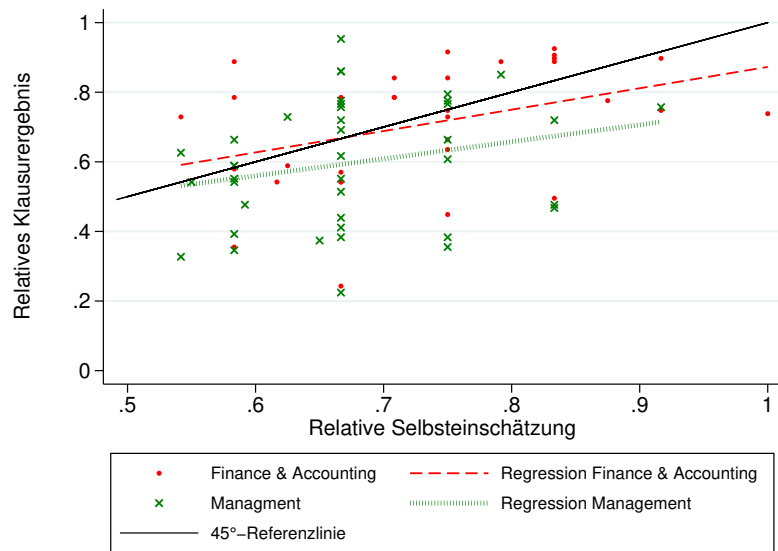


Basis für diese Grafik ist der nach MCON und ERG getrimmte Datensatz.

6.1.4 Klausur OMAT-Vorlesung (T₄)

Abbildung 5 zeigt die Mathematical Confidence der Schwerpunkt zum Zeitpunkt der OMAT-Klausur. Dabei ist wichtig, dass die Selbsteinschätzung in der letzten Vorlesung abgegeben wurde, ohne die Klausur und deren Schwierigkeitsgrad zu kennen. Auffällig ist, dass die befragten Studierenden der beiden Schwerpunkte erwarten, mindestens 50 Prozent der Gesamtpunktzahl erreichen zu können. Das heißt sie nehmen an, die Klausur im Durchschnitt zu bestehen. Es sind entsprechend lediglich Wertepaare vorhanden, die eine relative Selbsteinschätzung größer als 50 Prozent aufweisen, weshalb die Skalierung der Abszisse zugunsten der Übersichtlichkeit angepasst wurde. Wie auch in den Messzeitpunkten T₁ – T₃ beobachtet, befindet sich die Regressionsgerade der *F-&A*-Studierenden oberhalb der Regressionsgeraden der *Management*-Studierenden, mit dem Unterschied, dass die beiden Regressionsgeraden hier erstmals keinen Schnittpunkt aufweisen. Die Steigung der Regressionsgeraden ähnelt der in T₃ beobachteten Steigung der *F-&A*-Studierenden. Die Regressionsgerade der *F-&A*-Studierende liegt größtenteils unterhalb der Winkelhalbierenden und die der *Management*-Studierenden vollständig. Die Studierenden insgesamt weisen im Durchschnitt ein teilweise erheblich zu hohes Confidence-Level im Vergleich zu ihrer Leistung auf. Der Grad der Überschätzung ist höher als der Grad der Unterschätzung, vor allem unter *Management*-Studierenden: Die Wertepaare unterhalb der Winkelhalbierenden weisen einen größeren Abstand zu dieser auf als diejenigen Wertepaare oberhalb der Winkelhalbierenden.

Abbildung 5: MCON₄, Klausur der OMAT-Vorlesung

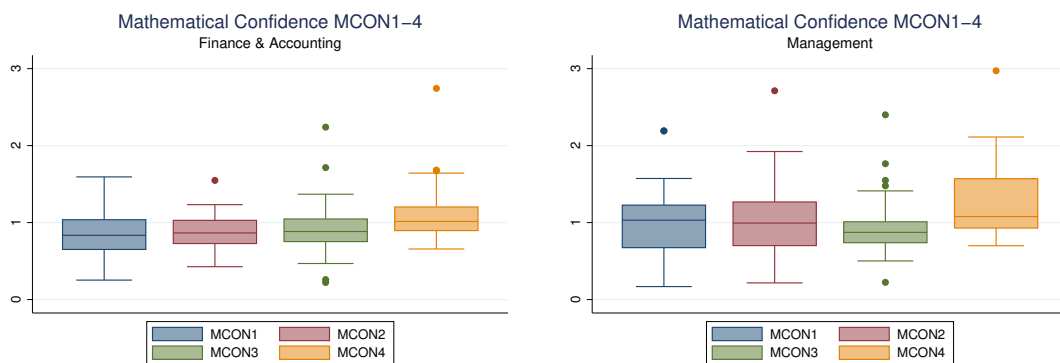


Basis für diese Grafik ist der nach MCON und ERG getrimmte Datensatz.

6.1.5 Fazit

Tabelle 2 zeigt MCON₁ – MCON₄ in einer Gegenüberstellung der Schwerpunkte *Finance & Accounting* und *Management*. Pro Schwerpunkt wurde für jeden Messzeitpunkt $t \in \{1; 4\}$ ein Boxplot generiert, der die Streuung der Beobachtungen, sowie die Quartile und den Median zeigt. Auf den ersten Blick ist auffällig, dass die Confidence-Level der *F-&A*-Studierenden insgesamt weniger schwanken.

Tabelle 4: MCON₁ – MCON₄ im Vergleich



Zusammenfassend wird festgehalten, dass *Management*-Studierende im Durchschnitt eine schlechtere Leistung in den Mathematik-Klausuren zeigen und sich gleichzeitig häufiger überschätzen als *F-&A*-Studierende. Beim Vergleich der Messzeitpunkte $T_1 - T_3$ zeigt der durchschnittliche *Management*-Studierende jedoch einen Lerneffekt bezüglich akkurater Selbsteinschätzung: Der Anteil der Wertepaare, die unterhalb der Winkelhalbierenden liegen, ist abnehmend, somit überschätzen sich weniger Studierende. In T_3 ist bei *Management*-Studierenden sogar eine Tendenz zur Underconfidence erkennbar. Dieser Trend ist allerdings in T_4 nicht mehr beobachtbar. Auch bei den *F-&A*-Studierenden scheint T_4 nicht repräsentativ zu sein, da die Regressionsgerade größtenteils im Bereich der Overconfidence liegt im Gegensatz zu $T_1 - T_3$. Betrachtet man den durchschnittlichen *F-&A*-Studierenden über $T_1 - T_3$, fällt auf, dass sich

dieser eher unterschätzt als überschätzt. Diese Inkonsistenz der Beobachtungen sowohl bei *F-&A*- als auch bei *Management*-Studierenden weist womöglich darauf hin, dass die OMAT-Klausur bezüglich des Schwierigkeitsgrades nicht mit den vorigen Übungsklausuren vergleichbar ist (trotz Nutzung des getrimmten Datensatzes). Die Beobachtung, dass sowohl *F-&A*-Studierende als auch *Management*-Studierende in der als leicht empfundenen Probeklausur tendenziell Underconfidence und in der schwierig empfundenen OMAT-Klausur Overconfidence zeigen, entspricht dem *Hard-Easy Effect* (vgl. [22]). Bei der Betrachtung aller Messzeitpunkte fällt insgesamt auf, dass Studierende beider Schwerpunkte im Fall einer niedrigen, relativen Selbsteinschätzung tendenziell Underconfidence und im Fall einer hohen, relativen Selbsteinschätzung Overconfidence zeigen.

6.2 Mittelwertvergleich

6.2.1 Auswahl des Testverfahrens

Die deskriptive Analyse aus 6.2.2 lässt vermuten, dass ein Gruppenunterschied hinsichtlich der Mathematical Confidence zwischen den Schwerpunkten besteht. Im nächsten Schritt wird statistisch überprüft, ob diese Unterschiede in den Messzeitpunkten $T_1 - T_4$ signifikant sind. Um zu entscheiden, welches Testverfahren auf Basis des vorliegenden Datensatzes angewendet werden darf, wird im folgenden bestimmt, ob die Messungen $MCON_t$ pro Schwerpunkt unabhängig sind, diese der Normalverteilung folgen und Varianzhomogenität vorliegt.

Dadurch, dass sich die Studierenden lediglich für einen Schwerpunkt entscheiden können, eine Doppelwahl also nicht möglich ist, befinden sich in den beiden Stichproben ausschließlich verschiedene Merkmalsträger, sodass von unverbundenen Stichproben ausgegangen wird. Aufgrund der kleinen Stichprobengrößen^{xii} wird im nächsten Schritt anhand des Shapiro-Wilk-Tests^{xiii} [10, S.121f] überprüft, ob die Messungen normalverteilt sind. Die Testergebnisse^{xiv} zeigen, dass nicht zu jedem Zeitpunkt T pro Schwerpunkt normalverteilte Daten vorliegen. Laut Bühner und Ziegler (2009, S. 261) sei eine Verletzung der Normalverteilungsannahme für den Zweistichproben-t-Test nicht allzu problematisch, solange die zu vergleichenden Stichprobenumfänge weitestgehend gleich groß sind und Varianzhomogenität vorliegt. Wie in 5.1 aus Tabelle 3 entnehmbar ist, sind die Stichprobengrößen bezüglich der Variablen $MCON_1 - MCON_4$ für die beiden Schwerpunkte zu keinem Messzeitpunkt gleich. Auch die Voraussetzung für Varianzhomogenität ist laut der Ergebnisse des Levene-Tests^{xv} nicht zu jedem Zeitpunkt erfüllt.

Um ein einheitliches Testverfahren für die Mittelwertvergleiche anzuwenden und wegen Nichterfüllung der Normalverteilungsannahme, der Varianzhomogenität und der gleichen Stichprobenumfänge wird im folgenden ein verteilungsunabhängiger Test angewendet [4, S.130]. Es empfiehlt sich, die nicht-parametrische Alternative des Zweistichproben-t-Tests, den Mann-Whitney-U-Test für unabhängige Stichproben, durchzuführen. Mithilfe von *Stata* wird der Rangsummentest von Wilcoxon^{xvi} durchgeführt, der mathematisch äquivalent zum Mann-Whitney-U-Test ist [4, S. 133].

Die Testergebnisse des Wilcoxon-Rangsummentests, die auf Basis der Approximation der Teststatistik durch die Normalverteilung generiert wurden, sind der folgenden Tabelle zu entnehmen:

^{xii}Die genaue Anzahl der Beobachtungen pro Messzeitpunkt ist in der Variablenübersicht in Kapitel 5.1 zu finden. Insgesamt schwankt die Beobachtungszahl pro Messzeitpunkt und Schwerpunkt zwischen 32 und 66 Beobachtungen.

^{xiii}Der Shapiro-Wilk-Test ist ein konservativer Signifikanztest, der unabhängig von der Stichprobengröße überprüft, ob die den Beobachtungen zugrundeliegende Zufallsvariable der Normalverteilung folgt. Eine kurze Erläuterung des Testverfahrens und die Testergebnisse sind in Anhang B.1 zu finden.

^{xiv}Die Ergebnisse des Shapiro-Wilk-Tests befinden sich in Tabelle 10 in Anhang B.1.

^{xv}Der Levene-Test prüft die Gleichheit zweier Populationsvarianzen, indem pro Gruppe die Abweichungen der Beobachtungen von den jeweiligen Gruppenmittelwerten verglichen werden Bortz and Schuster [4, S. 129]. Eine kurze Erläuterung des Testverfahrens und die Testergebnisse sind in Anhang B.2 zu finden.

^{xvi}Eine Beschreibung des Testverfahrens ist in Anhang B.3 zu finden.

Tabelle 5: Ergebnisse des Wilcoxon-Rangsummentest $MCON_1 - MCON_4$

$H_0 : F_{F\&A} = F_M$ gegen $H_1 : F_{F\&A} \neq F_M$								
Variable	Z	P($MCON_{t,M} > MCON_{t,F\&A}$)	mit $\alpha = 0,05$			mit $\alpha = 0,1$		
			$z_{\frac{\alpha}{2}}$	$z_{1-\frac{\alpha}{2}}$	Entscheidung	$z_{\frac{\alpha}{2}}$	$z_{1-\frac{\alpha}{2}}$	Entscheidung
$MCON_1$	1,103	0,573	-1,960	1,960	H_0 nicht ablehnen	-1,645	1,645	H_0 nicht ablehnen
$MCON_2$	1,652	0,603	-1,960	1,960	H_0 nicht ablehnen	-1,645	1,645	H_0 ablehnen
$MCON_3$	-0,195	0,490	-1,960	1,960	H_0 nicht ablehnen	-1,645	1,645	H_0 nicht ablehnen
$MCON_4$	1,332	0,592	-1,960	1,960	H_0 nicht ablehnen	-1,645	1,645	H_0 nicht ablehnen

Die Nullhypothese, dass sich die Schwerpunkte *Finance & Accounting* und *Management* in ihren durchschnittlichen Confidence-Level pro Messzeitpunkt T unterscheiden, kann jeweils zum Signifikanzniveau von 5 Prozent nicht abgelehnt werden. Wird hingegen ein Signifikanzniveau von 10 Prozent angenommen, ergibt sich für T₂ ein statistisch signifikanter Unterschied.

7 Multivariate Analysen

In den folgenden Regressionsanalysen soll der Einfluss von möglichen Prädiktoren auf das Kriterium $MCON_t$ analysiert werden. Von den vier Messzeitpunkten, wird lediglich Messzeitpunkt T₂ mit Messzeitpunkt T₄ verglichen. Von der Analyse des Messzeitpunktes T₁ wird im folgenden abgesehen, da dieser T₂ nur um circa drei Wochen vorgelagert ist. Es wird angenommen, dass T₂ für die Messung von $MCON_t$ auf Basis des wiederholten Schulwissens repräsentativer als T₁ ist, weil die Datenerhebung unterschiedlich lange Übergangszeiten zwischen Abitur und Studienbeginn vernachlässigt und auf Grundlage dessen kein unverzerrter Vergleich der Confidence-Levels möglich wäre. Des Weiteren kann so der in 6.2 vermutete Gruppenunterschied zwischen *F-&A*- und *Management*-Studierenden in T₂ auf signifikante Einflussfaktoren untersucht werden. In T₂ gelten die Studierenden immer noch als Studienanfänger, die keine Hochschulbildung im engeren Sinne erhalten sondern nur an dem Mathe-Vorkurs teilgenommen haben.

Auch Messzeitpunkt T₃ wird von den folgenden Regressionsanalysen ausgeschlossen: Die Probeklausur wird bereits sechs Wochen nach Semesterbeginn geschrieben. Zu diesem Zeitpunkt verfügen die Studierenden erst geringfügig über Studienerfahrung. In Messzeitpunkt T₄ hingegen wurden im Regelfall bereits zwei universitäre Prüfungen abgelegt, sodass die Studierenden sowohl Studien- als auch Prüfungserfahrung sammeln konnten. Somit wird angenommen, dass T₄ ein repräsentativer Maßstab für das Confidence-Level von Studierenden mit Studienerfahrung ist.

Der Vergleich von $MCON_2$ mit $MCON_4$ repräsentiert eine Gegenüberstellung des Confidence-Levels von Studierenden ohne Studienerfahrung nach Auffrischung der mathematischen Schulbildung (T₂) und der Confidence von Studierenden mit Studienerfahrung nach erfolgter mathematischer Hochschulbildung (T₄).

Zunächst wird pro betrachtetem Messzeitpunkt von einem simplen Basismodell ausgegangen, das im Zuge einer *hierarchischen* Regressionsanalyse schrittweise um weitere mögliche Prädiktoren erweitert wird (vgl. Bühner2009, S. 688). Im Basismodell wird angenommen, dass $MCON_t$ mit $t \in \{2, 4\}$ eine Funktion der unabhängigen Variablen *Schwerpunkt* ($DMANFIN$), *Geschlecht* ($DGENDER$) und *Migrationshintergrund* ($DMIGR$) ist.

$$MCON_t = f(DMANFIN, DGENDER, DMIGR) \forall t \in \{2, 4\}$$

In weitergehenden Analysen soll der jeweilige Einfluss der *zusätzlichen* Prädiktoren *Geburtsjahr (BIRTH)*, *Studierendenstatus (FIRST)*, *Abitur-Note (ABI)*, *Abitur-Note im Fach Mathematik (MABI)* und die *Erfahrung* einer erfolgreichen respektive missglückten Selbsteinschätzung bezüglich der jeweils vorherigen Klausur ($DGOODNEWS_{t-1}$) untersucht werden.

$$MCON_t = f(DMANFIN, DGENDER, DMIGR, BIRTH, DFIRST, ABI, MABI, DGOODNEWS_{t-1}) \\ \forall t \in \{2, 4\}$$

Um diese Zusammenhänge empirisch zu analysieren wird jeweils ein zugehöriges ökonometrisches Modell hergeleitet [35, S. 3f]. Dabei werden multiple lineare Regressionsmodelle entwickelt, deren Parameter mithilfe der *Methode der Kleinsten Quadrate (Ordinary Least Squares)* geschätzt werden. Hierbei werden ausschließlich Log-Level Spezifikationen genutzt, sodass die geschätzten Koeffizienten sinnvoll interpretiert werden können. Die Nutzung einer Level-Level-Spezifikation ist hier nicht geeignet, da der Einfluss einer Änderung des Prädiktors in Einheiten des Kriteriums gemessen wird; das Kriterium allerdings per Definition nicht in Einheiten gemessen wird (s. 6.2.1), weshalb die geschätzten Koeffizienten schwer interpretiert werden könnten. Bei der Log-Level-Spezifikation werden die geschätzten Koeffizienten der Prädiktoren als Semi-Elastizitäten interpretiert. Der Koeffizient β_i entspricht der Veränderung in $\ln(MCON)$, falls $\Delta x_i = 1$ gilt. Dabei ist $100 * \beta_i$ die approximative prozentuale Veränderung in der abhängigen Variable $MCON_t$ [35, S. 183].

$$\ln(MCON_t) = f(DMANFIN, DGENDER, DMIGR, BIRTH, DFIRST, ABI, MABI, DGOODNEWS_{t-1}) \\ \forall t \in \{2, 4\}$$

Die OLS-Methode setzt voraus, dass die zugrundeliegende Stichprobe eine Zufallsstichprobe ist (vgl. [35, S. 80]) und fehlende Beobachtungen des Kriteriums keinem systematischen Muster folgen. Um dem Risiko einer falschen Inferenz aufgrund verzerrter geschätzter Koeffizienten wegen nichtbeachteter Stichprobenauswahl (*Sample-Selection*) zu entgehen, wird den folgenden OLS-Regressionen ein *Heckit-Modell* gegenübergestellt. Das Heckit-Modell kontrolliert für Verzerrungen aufgrund nicht zufälliger Determinanten der Teilnahme an der Studie, dass bestimmte Gruppierungen systematisch nicht an den Befragungen oder Klausuren teilnehmen (*Incidental Selection*, vgl. [8]; *Incidental Truncation*, vgl. [35, S. 593]). In der folgenden Auswertung wird der *Heckman two-step estimator* verwendet, da dieser im Gegensatz zu dem *Full Information Maximum Likelihood estimator* weniger sensitiv auf eine Verletzung der Normalverteilungsannahme reagiert [8]. Im ersten Schritt des Heckit-Modells wird ein Probit-Modell geschätzt, auf Basis dessen die Wahrscheinlichkeit der Teilnahme an der Studie und die *inverse Mill's Ratio* berechnet wird. Anhand dieser Kennzahl wird bestimmt, ob eine Stichprobenselektion vorliegt (vgl. [28]). Falls dies der Fall ist, wird die *inverse Mill's Ratio* als Prädiktor in die OLS-Regression aufgenommen, um für den Selektionseffekt zu kontrollieren (vgl. [28]). Es wird für jede Regression der Messzeitpunkte T_2 und T_4 geprüft, ob eine Stichprobenselektion vorhanden ist. Hierbei seien y_j die jeweils zu schätzende Regressionsgleichungen der Spezifikationen j für T_2 und T_4 (Tabelle 6 respektive 7) :

$$y_j = \mathbf{x}_j \beta + u_{1j}$$

Die abhängige Variable $MCON_2$ bzw. $MCON_4$ sei annahmegemäß dann beobachtbar^{xvii}, falls die folgende Relation des *Partizipationsmodells*^{xviii} erfüllt ist:

$$\mathbf{z}_j \gamma + u_{2j} > 0 \\ \text{mit } \mathbf{z}_j = \{DMANFIN, DGENDER, DMIGR, MABI, DFIRST, BIRTH, DGOODNEWS_{t-1}\}$$

^{xvii}Beobachtbarkeit der abhängigen Variable bedeutet, dass die Studierenden sowohl die jeweilige Klausur t bearbeitet haben, als auch eine Selbsteinschätzung zu dieser abgegeben haben.

^{xviii}Die Prädiktoren der zu schätzenden Regression sind eine echte Teilstichprobe der Prädiktoren des Partizipationsmodells [35, S. 595]. Somit sind mindestens die aufgeführten Prädiktoren in der Selektionsgleichung enthalten.

Sei u_1 das jeweilige Residuum der zu schätzenden Regressionenszusammenhänge, dann misst ρ die Korrelation zwischen u_1 und u_2 mit $u_1 \sim N(0, \sigma)$ und $u_2 \sim N(0, 1)$. Mithilfe des Standardfehlers des Residuums u_1 der jeweiligen Regressionsgleichung, berechnet *Stata* die inverse Mill's Ratio $\lambda = \rho * \sigma$, auch bekannt als *Selektivitätseffekt* [32, *College Station, TX: Stata Press.*]. Ist der geschätzte Koeffizient λ signifikant, so wird die Nullhypothese $H_0 : \rho = 0$ abgelehnt und es liegt Sample Selection vor (vgl.[35, S. 596]). In diesem Fall sollte $\hat{\lambda}$ in die zu schätzende Regressionsgleichung einbezogen werden. Zudem sollten bei vorliegender Sample-Selection die über die Heckit-Methode geschätzten Koeffizienten bevorzugt werden, da die gewöhnlichen OLS-Schätzer der Populationsvarianzen zu gering sind und nicht für die Berechnung der Standardfehler nach OLS genutzt werden können [8, 18]. Falls in den jeweiligen Spezifikationen in Tabelle 6 und 7 laut der Heckman Two-Stage Methode eine Stichprobenselektion vorliegt, werden die zugehörigen Heckman-Schätzer mit dem Zusatz „H“ angegeben.

7.1 Regressionsergebnisse

In Tabelle 6 und 7 werden die Regressionsergebnisse für die Messzeitpunkte T_2 und T_4 präsentiert. Es wird zunächst eine Regression für das Basismodell geschätzt, das in der weitergehenden Analyse um zusätzliche Prädiktoren ergänzt wird. Die Auswahl und Reihenfolge der ergänzenden Prädiktoren erfolgt gemäß ihrer Korrelation mit dem jeweiligen Kriterium. Die Interkollinearitäten zwischen den einzelnen möglichen Prädiktoren i sind bei dieser Auswahl weitestgehend zu vernachlässigen, da insgesamt nur geringfügige Interkollinearität ($\rho_{ij} < 0,2$ mit $i \neq j$) (s. Tabelle 11 in C.2) zwischen den einzelnen Prädiktoren vorliegt. Insbesondere unter den dichotomen Prädiktoren konnte entweder keine Kollinearität oder nur minimale Kollinearität festgestellt werden. Auch die Korrelationen zwischen dichotomen und intervallskalierten Variablen sind ausnahmslos sehr gering. Lediglich die intervallskalierten Variablen untereinander weisen eine stärkere Korrelation auf: Die Variable *BIRTH* korreliert jeweils mittelstark mit den Variablen *ABI* und *MABI*. Entsprechend gilt, je jünger der Studierende ist, desto besser ist seine Leistung sowohl im Abitur selbst, als auch in der Mathematik-Prüfung im Rahmen des Abiturs. Die Variable *MABI* weist ebenfalls eine mittelstarke Korrelation mit der Variable *ABI* auf, was nicht verwunderlich ist, da die Note der Mathematikprüfung Bestandteil der Abiturnote insgesamt ist. Bei Einbezug korrelierter Prädiktoren in die Regressionsanalyse sollten die Koeffizienten unter Beachtung des Zusammenhangs der Prädiktoren interpretiert werden.

Im folgenden werden lediglich die Ergebnisse derjenigen Regressionsmodelle gezeigt, in denen mindestens ein Prädiktor signifikanten Einfluss auf die *Mathematical Confidence* nimmt. Der absolute Betrag der geschätzten Koeffizienten variiert teilweise aufgrund verschiedener zugrundeliegender Spezifikationen, die Richtung der geschätzten Effekte ist jedoch sowohl in T_2 als auch T_4 weitestgehend dieselbe.

7.1.1 OLS-Regressionen T_2

Tabelle 6 zeigt die Regressionen, die das Kriterium $\ln(\text{MCON}_2)$ auf mögliche Prädiktoren regressieren. Im Folgenden wird zunächst auf die verschiedenen Signifikanzlevel der Prädiktoren in den unterschiedlichen Modellen eingegangen und im Anschluss werden die zugehörigen geschätzten Koeffizienten interpretiert.

Ein robustes Ergebnis aller Spezifikationen in T_2 ist, dass die Prädiktoren des Basismodells einen zum 5 Prozent Signifikanzniveau relevanten Einfluss auf das durchschnittliche Confidence-Level der Studierenden nehmen (zutreffend für alle Modelle außer I.VII).

Wird zusätzlich für das Geburtsjahr *BIRTH* und den Studierendenstatus *DFIRST* bzw. nur für *DFIRST* kontrolliert (Modell I.III bzw. Modell I.IV), zeigt sich, dass sowohl der voraussichtlich gewählte Schwerpunkt, als auch das Geschlecht des Studierenden hochsignifikanten Einfluss auf das Confidence-Level nehmen. Der Migrationshintergrund der Studierenden hat außer in Modell I.III in allen Spezifikationen einen zum 5 Prozent Niveau signifikanten Einfluss auf das Confidence-Level. Das Ausmaß des Einflusses schwankt nur geringfügig.

Die Ergänzung des Basismodells durch den Prädiktor *DGOODNEWS*₁ (Modell I.II) zeigt, dass auch die Erfahrung einer erfolgreichen respektive missglückten Selbsteinschätzung bezüglich der Eingangsklausur des Mathe-Vorkurses

einen zum 5 Prozent Niveau signifikanten Einfluss auf das Confidence-Level zum Zeitpunkt T_2 nimmt. Die Prädiktoren *BIRTH*, *DFIRST*, *MABI* und *ABI* zeigen keinen signifikanten Einfluss auf die Confidence der Studierenden. Auch die Ergänzung des Basismodells (I.I) um Interaktionsvariablen zwischen *DMANFIN*, *DMIGR*, *DGENDER*, *DFIRST* und *BIRTH* zeigen keine signifikanten Einflüsse. Lediglich die Interaktionsvariable zwischen *DGENDER* und *DGOODNEWS₁* hat ergänzend zu dem Basismodell (I.I) einen zum 5%-Niveau signifikanten Einfluss auf die Confidence, während die beiden einzelnen Prädiktoren *DGENDER* und *DGOODNEWS₁* bei Interaktion dieser wiederum insignifikant werden: Der Einfluss, den ein Erfolgserlebnis aus der Eingangsklausur auf die Confidence des Studierenden zum Zeitpunkt der Abschlussklausur des Mathe-Vorkurses hat, ist abhängig von dem Geschlecht des Studierenden.

In T_2 zeigen sowohl die folgenden Regressionen als auch die Korrelationen zwischen *DMANFIN* und *MCON₂* (s. Tabelle 11 in 6.2), dass sich *F-&A*-Studierende (*DMANFIN=1*) weniger selbstbewusst einschätzen und deren Confidence somit im Durchschnitt geringer ist als die der Management-Studierenden (*DMANFIN=0*). Je nach Spezifikation des Modells schwankt der Einfluss des Schwerpunktes auf die Confidence. In dem Basismodell I.I, bei dem 15,48 Prozent der Varianz in *MCON₂* erklärt wird, weist der durchschnittliche *F-&A*-Studierende ein um 20,44 Prozent geringeres Confidence-Level auf als der durchschnittliche Management-Studierende. In Modell I.VI, in dem überdies 36,89 Prozent der Variation in *MCON₂* erklärt wird, hat die Schwerpunktwahl einen hochsignifikanten Einfluss, sodass der durchschnittliche *F-&A*-Studierende ein um 23,06 Prozent geringeres Confidence-Level aufweist als der durchschnittliche Management-Studierende. Einen ähnlich großen Einfluss übt *DGENDER* auf *MCON₂* aus: Das Basismodell I.I zeigt, dass weibliche Studierende (*DGENDER=1*) ein im Durchschnitt um 24,86 Prozent geringeres Confidence-Level aufweisen als männliche Studierende (*DGENDER=0*). Auch hier schwankt der Einfluss von *DGENDER* auf *MCON₂* über verschiedene Spezifikationen hinweg. Allerdings zeigt sich grundsätzlich, dass sich weibliche Studierende durchschnittlich weniger selbstbewusst einschätzen.

Die Ergänzung des Basismodells I.I um *DMIGR* verdeutlicht, dass das Vorhandensein eines Migrationshintergrundes das Confidence-Level positiv beeinflusst: Ein Studierender mit Migrationshintergrund (*DMIGR=1*) ist durchschnittlich um 22,65 Prozent selbstbewusster als der durchschnittliche Studierende ohne Migrationshintergrund (*DMIGR=0*). Die Größe dieses Effekts verändert sich nur marginal in den verschiedenen Spezifikationen, während der Effekt insgesamt in keinem Modell als hochsignifikant einzustufen ist.

Bei Ergänzung des Basismodells I.I um *DGOODNEWS₁* zeigt sich, dass ein Erfolgserlebnis aus der Eingangsklausur (*DGOODNEWS₁ = 1*) einen negativen Einfluss auf das Confidence-Level hat. Schneidet ein durchschnittlicher Studierender in der Eingangsklausur des Mathe-Vorkurses besser ab als ursprünglich erwartet, so zeigt dieser ein um 16,08 Prozent geringeres Confidence-Level als ein Studierender mit einem die ursprüngliche Erwartung unternetzenden Ergebnis (*DGOODNEWS₁ = 0*). Bei Betrachtung der Interaktionsvariable zwischen *DGOODNEWS₁* und *DGENDER* zeigt Modell I.III, dass insbesondere das besagte Erfolgsmoment einen in Abhängigkeit vom Geschlecht unterschiedlichen Einfluss auf die Confidence aufweist: Weibliche Studierende, die in der Eingangsklausur des Mathe-Vorkurses besser als erwartet abgeschnitten haben, weisen durchschnittlich ein um 26,95 Prozent geringeres Confidence-Level auf als männliche Studierende mit selbigem Erfolgserlebnis.

7.1.2 Heckman-Zweistufen-Verfahren T_2

Bei Durchführung des Zweistufen-Verfahrens nach Heckman wird deutlich, dass die inverse Mill's Ratio in keinem der in Tabelle 6 aufgeführten Modelle signifikant ist. Insofern scheint keine endogene Sample-Selection vorzuliegen. Betrachtet man allerdings die nach Heckman geschätzten Koeffizienten, so zeigen sich in einigen Spezifikationen erhebliche Unterschiede zu den OLS-Schätzern. Vor allem fällt auf, dass in jeder Spezifikation alle Prädiktoren im Zuge der Korrektur für Stichprobenselektion insignifikant werden. Nimmt man also trotz der insignifikanten inversen Mill's Ratio an, dass Sample-Selection vorliegt, so zeigt lediglich das Geschlecht einen hochsignifikanten Einfluss: Unabhängig von einer vermeintlichen Stichprobenselektion unterschätzen sich weibliche Studierende im Durchschnitt um 25 bis 30 Prozent. Außerdem hätte die inverse Mill's Ratio zwar einen insignifikanten, aber negativen Effekt. Das heißt diejenigen Studierenden, die die Abschlussklausur des Mathe-Vorkurses bearbeitet und eine Selbsteinschätzung

zu dieser abgegeben haben, weisen im Durchschnitt ein geringeres Confidence-Level auf als eine Zufallsstichprobe von vergleichbaren Studierenden.

7.1.3 OLS-Regressionen T_4

Tabelle 7 enthält Regressionsergebnisse aus Messzeitpunkt T_4 . Im Folgenden wird zunächst auf die verschiedenen Signifikanzlevel der Prädiktoren in den unterschiedlichen Modellen eingegangen und im Anschluss werden die zugehörigen geschätzten Koeffizienten interpretiert.

Im Vergleich zu T_2 ist auffällig, dass die gleichen Prädiktoren in den gleichen Modellen grundsätzlich in geringerem Ausmaß signifikant sind, sodass womöglich essentielle Prädiktoren zur Erklärung der Varianz in $MCON_4$ fehlen.

Das Basismodell II.I zeigt, dass *DMANFIN*, *DGENDER* und *DMIGR* signifikanten Einfluss auf das Confidence-Level der Studierenden nehmen. Allerdings unterliegt der Einfluss von *DGENDER* und *DMIGR* im Vergleich zu T_2 einer geringeren Signifikanz (Vergleich T_2 : $\alpha = 0,05$ vs. T_4 : $\alpha = 0,1$). Auch der in geringem Ausmaß erklärte Anteil der Varianz in $MCON_4$ ist in der Basis-Spezifikation in T_4 wesentlich geringer als in T_2 (Vergleich T_2 : $R^2 = 0,1548$ vs. T_4 : $R^2 = 0,0958$), was die Vermutung fehlender Prädiktoren bestärkt.

In Modell II.II wird das Basismodell um den insignifikanten Prädiktor *DGOODNEWS₃* erweitert. Im Zuge dessen ändern sich die Koeffizienten der Prädiktoren aus dem Basismodell II.I und die Güte der Regression zwar nur marginal, allerdings verbessert sich die Signifikanz von *DMIGR*: Durch die Ergänzung von *DGOODNEWS₃* wird für die in *DMIGR* von *DGOODNEWS₃* abhängige Varianz kontrolliert, sodass der Einfluss von *DMIGR* auf $MCON_4$ von *DGOODNEWS₃* bereinigt ist.

Wird das Basismodell II.I um die Prädiktoren *BIRTH* und *FIRST* erweitert, zeigt sich weder eine Verbesserung der Signifikanzen der Prädiktoren des Basismodells, noch weisen die hinzugefügten Prädiktoren selbst einen signifikanten Einfluss auf das Confidence-Level auf. Auf Basis der Regressionsergebnisse kann also wie in T_2 davon ausgegangen werden, dass weder das Alter noch der Studierendenstatus einen signifikanten Einfluss auf das Confidence-Level der Studierenden aufweist. Diese empirische Erkenntnis steht im Gegensatz zu denen von Nowell und Alston (2007), die älteren Studierenden ein geringeres Maß an Overconfidence zuschreiben. Die hier vorliegende Insignifikanz des Alters der Studierenden ist möglicherweise auf die geringe Variation in *BIRTH* zurückzuführen, da die Studierenden einer sehr homogenen Altersgruppe entstammen.

Modell II.III zeigt die Auswirkungen der Abiturleistung auf das Basismodell. Diese scheint selbst keinen signifikanten Einfluss auf das Confidence-Level aufzuweisen. Durch deren Kontrolle zeigt sich allerdings, dass *DMANFIN* und *DGENDER* hochsignifikanten Einfluss auf $MCON_4$ aufweisen. Um diese Beziehung zwischen *ABI* und sowohl *DMANFIN* als auch *DGENDER* genauer zu analysieren, werden in den Modellen II.IV und II.V Interaktionsvariablen aufgenommen. Modell II.IV enthält die Interaktion zwischen *DMANFIN* und *ABI* und Modell II.V die Interaktion zwischen *DGENDER* und *ABI*. So wird einerseits angenommen, dass die Abiturleistung unter Studierenden abhängig vom Schwerpunkt unterschiedlich ist und andererseits in Abhängigkeit vom Geschlecht ebenfalls verschieden ist. In beiden Modellen beeinflusst *ABI* und der jeweilige Interaktionsterm^{xix} das Confidence-Level auf dem 5 Prozent Signifikanzniveau. Auch *DMANFIN* und *DGENDER* bleiben unter Einbezug der Interaktionen signifikante Einflussfaktoren; nur *DMIGR* weist in dieser Kombination keinen signifikanten Erklärungsgehalt für $MCON_4$ auf. Ähnliches ist bei Aufnahme des Prädiktors *MABI* in das Basismodell II.I zu beobachten: Modell II.VI zeigt, dass die alleinige Ergänzung des Basismodells durch *MABI* keinen signifikanten Erklärungsgehalt bezüglich des Confidence-Levels enthält, sondern sogar die Prädiktoren des Basismodells insignifikant werden lässt. Wird nun zusätzlich die Interaktion zwischen *MABI* und *DGENDER* berücksichtigt, so haben sowohl die reinen Prädiktoren *DGENDER* und *MABI* als auch der Interaktionsterm Einfluss auf dem Signifikanzniveau von 1 Prozent, sodass die Leistung der Abiturprüfung in Mathematik wie auch die reine Abiturleistung geschlechterspezifisch ist.

^{xix}Interaktionsterme sind künstliche Prädiktoren, die dem Produkt zweier Prädiktoren entsprechen (vgl. [35, S. 844]) und somit einen nichtlinearen Zusammenhang zwischen den beiden besagten Variablen implizieren.

Tabelle 6: OLS vs. Heckit, Kriterium: $\ln(MCON_2)$

Prädiktoren	(I.I)	(I.II)	(I.III)	(I.IV)	(I.V)	(I.VI)	(I.VII)	(I.VIII)
<i>DMANFIN</i>	-0,2044** (0,0837) H: 0,0508	-0,1571* (0,0897) H: -0,0730	-0,1621* (0,0873) H: -0,0821	-0,2141** (0,0868) H: -0,0869	-0,2372*** (0,0867) H: -0,1074	-0,2306*** (0,0840) H: -0,1048	-0,2198** (0,0895) H: -0,0687	-0,1001 (0,0954) H: -0,085
<i>DGENDER</i>	-0,2486** (0,0953) H: -0,1119**	-0,1759* (0,0947) H: -0,2128*	-0,0499 (0,1289) H: -0,0927	-0,2514** (0,0957) H: -0,2697***	-0,2661*** (0,0953) H: -0,3106***	-0,2645*** (0,0947) H: -0,3104***	-0,2804*** (0,0974) H: -0,2855***	-0,2543** (0,1057) H: -0,263***
<i>DMIGR</i>	0,2265** (0,0944) H: -0,0104	0,2114** (0,1009) H: 0,1484	0,2002** (0,0979) H: 0,1426	0,2239** (0,0939) H: 0,1652	0,2064** (0,0930) H: 0,1223	0,2078** (0,0930) H: 0,122	0,1985** (0,0973) H: 0,1493	0,0793 (0,1072) H: 0,1075
<i>BIRTH</i>			-0,0098 (0,0131) H: -0,0087	-0,0076 (0,0122) H: -0,0027				
<i>DFIRST</i>					-0,1345 (0,0878) H: -0,1674	-0,1389 (0,0892) H: -0,1717		
<i>ABI</i>							0,0133 (0,0768) H: -0,0411	-0,0261 (0,0194) H: -0,291
<i>MABI</i>								
<i>DGOODNEWS1</i>		-0,1608** (0,0768) H: -0,1448	-0,0325 (0,0971) H: -0,0579					
<i>DGOODNEWS1*DGENDER</i>			-0,2695* (0,1536) H: -0,2156					
<i>CONS</i>	0,0071 (0,0733)	-0,0064 (0,0843)	-0,0477 (0,0963)	19,6385 (26,1916)	15,2383 (24,2904)	0,1276 (0,1087)	0,0081 (0,1608)	0,2761 (0,222)
Beobachtungen	83	71	71	83	83	83	76	65
R ²	0,1548	0,1815	0,2120	0,1591	0,1819	0,3689	0,1692	0,1366
H: Inverse Mills Ratio $\hat{\lambda}$	-0,1833 (0,2961)	-0,1269 (0,378)	-0,1115 (0,2950)	-0,1507 (0,3044)	-0,1294 (0,3003)	-0,1384 (0,2937)	-0,2408 (0,3386)	0,0718 (0,35)

Anmerkungen: Angegeben sind die unstandardisierten Regressionskoeffizienten sowie die robusten Standardfehler (in Klammern). In den rechteckigen Klammern befinden sich die über die Heckit-Methode (two-stage method) generierten Koeffizienten. Signifikante Koeffizienten sind folgendermaßen hervorgehoben: * / ** / *** = $p < 0,1/0,05/0,01$.

Das Basismodell II.I zeigt, dass der durchschnittliche *F-&A*-Studierende ($DMANFIN = 1$) im Gegensatz zum durchschnittlichen Management-Studierenden ($DMANFIN = 0$) ein um 18,26 Prozent reduziertes Confidence-Level aufweist. Des Weiteren zeichnen sich weibliche Studierende ($DGENDER = 1$) durchschnittlich über ein um 15,25 Prozent geringeres Confidence-Level aus als männliche Studierende ($DGENDER = 0$). Diejenigen Studierenden mit Migrationshintergrund ($DMIGR = 1$) zeigen außerdem ein um 14,47 Prozent höheres Ausmaß an Confidence als Studierende ohne Migrationshintergrund ($DMIGR = 0$). Die Richtungen der Effekte und deren Größenordnung ist in Modell II.II bei Ergänzung des Basismodells um den Prädiktor $DGOODNEWS_3$ in Modell II.II weitestgehend äquivalent.

Erst durch Einbezug von Interaktionstermen ändern sich die absoluten Werte der geschätzten Koeffizienten maßgeblich im Vergleich zum Basismodell II.I. Modell II.IV berücksichtigt die Abiturleistung *ABI* und nimmt diesbezüglich an, dass sich das Confidence-Level der Studierenden in Abhängigkeit von der Abiturleistung und der Schwerpunktwahl ändert, je nach Ausprägung dieser beiden Variablen in unterschiedlichem Maß. Die Relation der Confidence-Level zwischen den Schwerpunkten verändert sich in Abhängigkeit der Abiturnote: Eine sehr gute Leistung im Abitur führt bei *F-&A*-Studierenden zu höherer Confidence als bei *Management*-Studierenden. Je schlechter die Leistung im Abitur, desto geringer ist das durchschnittliche Confidence-Level von *F-&A*-Studierenden und desto höher das der *Management*-Studierenden. Bei Nichtbeachtung der Interaktion würden diejenigen Studierenden, die ein schlechtes Abitur vorweisen, allgemein ein höheres Ausmaß an Confidence zeigen, insbesondere würden sich aber *Management*-Studierende durchschnittlich eher überschätzen als *F-&A*-Studierende.^{xx}

Nicht nur die Schwerpunktwahl beeinflusst den Effekt des Abiturs auf das Confidence-Level, sondern auch das Geschlecht. Dieser Interaktionseffekt wird in Modell II.V aufgezeigt. Der Einfluss von $DMANFIN$ auf das Ausmaß der Confidence ähnelt dem des Basismodells. Hingegen zeigt $DMIGR$ lediglich einen Einfluss von 1,91 Prozent, wobei dieser insignifikant ist. Die zum 5 Prozent Niveau signifikante Interaktion zwischen $DGENDER$ und *ABI* zeigt, dass die Abiturnote in Abhängigkeit vom Geschlecht einen unterschiedlichen Einfluss auf das Confidence-Level aufweist: Je besser der Abiturschnitt, desto höher das Maß an Confidence bei weiblichen im Gegensatz zu männlichen Studierenden. Je schlechter allerdings die Abiturnote, desto höher das Confidence-Level der männlichen im Vergleich zu weiblichen Studierenden.^{xxi}

Modell II.VI ergänzt das Basismodell II.I um den Prädiktor *MABI*, der Leistung in der Abiturprüfung Mathematik, und zeigt, dass diese keinen signifikanten Einfluss auf das Confidence-Level der Studierenden nimmt. Wird das Modell II.VI allerdings um die Interaktion zwischen Geschlecht und der Leistung im Mathematik-Abitur erweitert, zeigt sich ein hochsignifikanter Einfluss der beiden besagten Variablen und der Interaktion. Die Größe des Effekts von *MABI* allein ändert sich abgesehen vom Signifikanzlevel nicht und ist nach wie vor sehr gering (vgl. [3, S. 141]).

^{xx}Beispiel: Unter Berücksichtigung dieser Interaktion wird ein durchschnittlicher *F-&A*-Studierender ($DMANFIN = 1$), der sein Abitur mit 1,0 absolviert hat, betrachtet und dessen Confidence-Level mit dem analogen Management-Studierenden verglichen. Ceteris paribus weist dieser *F-&A*-Studierende durch seine Studienwahl und Abiturnote ein um 62,68 Prozent ($0,7461 * 1 + 0,3436 * 1,0 - 0,4629 * 1 * 1,0$) erhöhtes Confidence-Level auf, während der analoge Management-Studierende nur ein um 34,36 Prozent ($0,7461 * 0 + 0,3436 * 1,0 - 0,4629 * 0 * 1,0$) erhöhtes Confidence-Level zeigt. Wird nun der gleiche Fall mit dem Unterschied, dass die beiden Studierenden einen Abiturschnitt von 2,0 erreicht haben, untersucht, zeigt sich, dass der *F-&A*-Studierende ein im Durchschnitt um 50,75 Prozent ($0,7461 * 1 + 0,3436 * 2,0 - 0,4629 * 1 * 2,0$) erhöhtes Confidence-Level aufweist, der Management-Studierende hingegen ein um 68,72 Prozent ($0,7461 * 0 + 0,3436 * 2,0 - 0,4629 * 0 * 2,0$) erhöhtes Ausmaß an Confidence.

^{xxi}Beispiel: Eine durchschnittliche weibliche Studierende ($DGENDER = 1$) wird mit dem durchschnittlichen männlichen Studierenden ($DGENDER = 0$) verglichen, wobei angenommen wird, dass beide ihr Abitur mit einem Schnitt von 1,0 ($ABI = 1,0$) absolvierten. Unter diesen Voraussetzungen weist ceteris paribus die weibliche Studierende ein um 63,95 Prozent ($0,7771 * 1 + 0,3304 * 1,0 - 0,4680 * 1 * 1,0$) erhöhtes Confidence-Level vor, während der analoge männliche Studierende ein um 33,04 Prozent ($0,7771 * 0 + 0,3304 * 1,0 - 0,4680 * 0 * 1,0$) erhöhtes Confidence-Level zeigt. Wird von einem Abiturdurchschnitt von 2,0 ausgegangen, so zeigt die durchschnittliche weibliche Studierende 50,19 Prozent ($0,7771 * 1 + 0,3304 * 2,0 - 0,4680 * 1 * 2,0$), hingegen der durchschnittliche männliche Studierende ein um 66,08 Prozent ($0,7771 * 0 + 0,3304 * 2,0 - 0,4680 * 0 * 2,0$) erhöhtes Ausmaß an Confidence.

7.1.4 Heckman-Zweistufen-Verfahren T_4

Auch bei Betrachtung von T_4 wird im Zuge des Zweistufen-Verfahrens nach Heckman deutlich, dass die inverse Mill's Ratio in keinem der in Tabelle 6 aufgeführten Modelle signifikant ist. So scheint auch zu diesem Messzeitpunkt keine endogene Sample-Selection vorzuliegen. Allerdings fällt bei dem Vergleich der OLS-Schätzer und Heckman-Schätzer für die Koeffizienten der Prädiktoren auf, dass sich diese auch in T_4 teilweise erheblich unterscheiden, was wiederum ein Indiz für Sample-Selection ist. Lediglich das Geschlecht und die Leistung im Mathe-Abitur, sowie deren Interaktion sind nach der Heckman-Korrektur noch signifikant. So unterschätzen sich weibliche Studierende im Durchschnitt im Gegensatz zu den männlichen Studierenden. Zudem weisen diejenigen Studierenden, die eine bessere Leistung im Mathe-Abitur aufweisen zugleich ein geringeres Confidence-Level auf. Die übrigen Prädiktoren werden insignifikant bei Kontrolle für Sample-Selection. Wie in T_2 weist die inverse Mill's Ratio einen insignifikanten aber negativen Effekt auf. Das heißt diejenigen Studierenden, die die Abschlussklausur des Mathe-Vorkurses bearbeitet und eine Selbsteinschätzung zu dieser abgegeben haben, weisen im Durchschnitt ein um 10 bis 20 Prozent geringeres Confidence-Level auf als eine Zufallsstichprobe von vergleichbaren Studierenden.

Tabelle 7: OLS vs. Heckit, Kriterium: $\ln(MCON_4)$

Prädiktoren	(II.I)	(II.II)	(II.III)	(II.IV)	(II.V)	(II.VI)	(II.VII)
<i>DMANFIN</i>	-0,1826** (0,0856) H: -0,1061	-0,211** (0,0873) H: 0,0852	-0,2087*** (0,0724) H: -0,1174	0,7461* (0,3954) H: 0,1098	-0,1593** (0,07) H: -0,1141	-0,0772 (0,247) H: -0,1032	-0,034 (0,0568) H: -0,0441
<i>DGENDER</i>	-0,1525* (0,0863) H: -0,1434	-0,1757* (0,0878) H: -0,1061	-0,1884*** (0,0704) H: -0,1872**	-0,1669** (0,0650) H: -0,1881**	0,7771* (0,4255) H: 0,0997	-0,0860 (0,0762) H: -0,1420	-0,7210*** (0,2157) H: -0,8863***
<i>DMIGR</i>	0,1447* (0,0752) H: 0,0252	0,1841** (0,0821) H: -0,009	0,0992 (0,0977) H: -0,0084	0,0494 (0,0926) H: -0,0045	0,0191 (0,0911) H: -0,0012	-0,0282 (0,0629) H: 0,0210	-0,0704 (0,056) H: -0,0394
<i>ABI</i>			0,1061 (0,1324) H: 0,101	0,3436** (0,1471) H: 0,163	0,3304** (0,1425) H: 0,1594		
<i>MABI</i>						-0,0037 (0,0137) H: -0,002	-0,0396*** (0,0122) H: -0,0434**
<i>GOODNEWS₃</i>		0,0800 (0,0766) H: -0,0475		-0,4629** (0,1881) H: -0,1136			
<i>DMANFIN * ABI</i>							
<i>DGENDER * ABI</i>					-0,4680** (0,2049) H: -0,1441		
<i>DGENDER * MABI</i>							0,0609*** (0,0197) H: 0,0717***
<i>CONS</i>	0,1932** (0,0791)	0,1533* (0,0808)	-0,0093 (0,2488)	-0,5146* (0,2838)	-0,4701* (0,2049)	0,1403 (0,1633)	0,5209*** (0,1434)
Beobachtungen	59	58	51	51	51	45	45
R ²	0,0958	0,1151	0,2241	0,3367	0,3370	0,0743	0,1960
Heckit: Inverse Mills Ratio $\hat{\lambda}$	-0,1645 (0,1912)	-0,1094 (0,2111)	-0,1865 0,1853	-0,1948 (0,1842)	-0,2024 (0,1862)	-0,1598 0,1938	-0,1391 (0,1563)

Anmerkungen: Angegeben sind die unstandardisierten Regressionskoeffizienten sowie die robusten Standardfehler (in Klammern). In den rechteckigen Klammern befinden sich die über die Heckit-Methode (two-stage method) generierten Koeffizienten. Signifikante Koeffizienten sind folgendermaßen hervorgehoben: * / ** / *** = $p < 0,1/0,05/0,01$.

7.2 Annahmen der multiplen Regression

Um die Regressionsergebnisse der Stichprobe auf die zugrundeliegende Population übertragen und generalisieren zu können, müssen die Gauss-Markov-Annahmen der multiplen Regression erfüllt sein (vgl. [12, S. 220]). Im Folgenden werden die entsprechenden Annahmen geprüft.

Da die Regressionsgleichungen so konzipiert sind, dass lediglich lineare Parameter geschätzt werden, ist die Annahme der *Linearität von Parametern (MLR1)* [35, S. 79] erfüllt.

Des Weiteren wird davon ausgegangen, dass pro Messzeitpunkt eine Zufallsstichprobe der Population der Studierenden der Wirtschaftswissenschaften vorliegt und von diesen wiederum zufallsabhängige Teilstichproben der Populationen der *F-&A-* und *Management-*Studierenden vorliegen. Es wird angenommen, dass die Stichproben unabhängig sind und jeder Studierende der Population mit der gleichen Wahrscheinlichkeit für die Stichprobe hätte ausgewählt werden können, sodass auch die Annahme der *Zufallsstichprobe (MLR2)* erfüllt ist.

Um zu prüfen, ob perfekte Kollinearität vorliegt, werden die Korrelationen aller unabhängigen Variablen betrachtet. Ist eine unabhängige Variable durch eine Linearkombination einer anderen unabhängigen Variable oder durch eine Linearkombination durch sich selbst darstellbar, so liegt perfekte Kollinearität vor und das Modell kann nicht mithilfe der OLS-Methode geschätzt werden. Die Korrelationsmatrix der unabhängigen Variablen (s. Tabelle 13, Anhang C.2) zeigt, dass keine perfekten Korrelationen vorliegen. Daher und weil keine der unabhängigen Variablen konstant ist, ist die Annahme, dass *keine perfekte Kollinearität (MLR3)* vorliegt, erfüllt^{xxii} [35, S. 80]. Falls korrelierte Variablen als unabhängige Prädiktoren in eine Regressionsanalyse einbezogen werden, so sollten zusätzlich die partiellen Einflüsse der einzelnen Prädiktoren auf die abhängige Variable betrachtet werden, um eine sinnvolle Inferenz zu gewährleisten. Die letzte Voraussetzung, um die Unverzerrtheit der OLS-Schätzer zu gewährleisten, ist die *Zero-Conditional-Mean-Annahme (MLR4)*[35, S. 82]:

$$E(u|x_1, \dots, x_k) = 0 \text{ oder } Cov(u|x_1, \dots, x_k) = 0$$

Im Erwartungswert nehmen die Residuen den Wert null an, unabhängig von den Ausprägungen der Prädiktoren. Diese Annahme ist axiomatischer Bestandteil der OLS-Methode und ist somit für die vorliegende Stichprobe erfüllt, sodass *exogene* unabhängige Variablen vorliegen und kein Endogenitätsproblem vorliegt.

Um effiziente Schätzer zu erhalten, muss die Annahme der *Homoskedastizität (MLR5)* erfüllt werden:

$$Var(u|x_1, \dots, x_k) = 0$$

Homoskedastizität bedeutet, dass die Varianz der Residuen konstant und für alle möglichen Kombinationen von Ausprägungen der unabhängigen Variablen gleich ist [35, S. 89]. In den Regressionen werden robuste Standardfehler (*White-Standardfehler*)^{xxiii} genutzt, um für (vermeintliche) Heteroskedastizität zu korrigieren und somit die asymptotische Gültigkeit der Teststatistiken und der resultierenden statistischen Inferenz zu gewährleisten.

8 Konklusion

Overconfidence ist insbesondere in wirtschaftlichen Berufsfeldern ein verbreitetes Phänomen, das Karrierelaufbahnen häufig aufgrund von Fehleinschätzungen und -leistungen aus übertriebenem Optimismus negativ beeinträchtigt. Angesichts der teilweise schwerwiegenden Konsequenzen^{xxiv} forscht die Psychologie und insbesondere die Wirtschaftspsychologie nach Ursachen für Wahrnehmungsschwierigkeiten, die sowohl Berufseinsteiger als auch Manager betreffen. Der anhaltende Fachkräftemangel bietet qualifizierten Absolventen zwar vielfältige Perspektiven, allerdings

^{xxii}Wichtig zu beachten ist, dass Multikollinearität kein Problem darstellt, lediglich perfekte Kollinearität darf nicht vorliegen.

^{xxiii}Die Standardfehler einer Regression sind im Falle von Heteroskedastizität nicht konsistent, sodass konsistente Schätzer für die Standardfehler konstruiert werden.

^{xxiv}"No problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence." (Plous 1993, S. 217)

scheitern diese häufig an mangelnder Sozialkompetenz und Selbstüberschätzung [20]. In der vorliegenden Arbeit wird das Ausmaß der Selbstüberschätzung im Bereich Mathematik von Studierenden der Wirtschaftswissenschaften explorativ mithilfe einer an Moore (2008) angelehnten Kennzahl für *Overestimation*, die „*Mathematical Confidence*“, untersucht. Dabei wird insbesondere analysiert, ob *F-&A*-Studierende über das erste Semester hinweg ein durchschnittlich höheres Level an Confidence aufweisen als *Management*-Studierende. Diese Hypothese kann auf Basis dieser Arbeit nicht bestätigt werden. Die diese Arbeit einleitende Frage, ob Studierende des Schwerpunktes *Finance & Accounting* oder *Management* bereits zu Studienbeginn eine verzerrte Selbsteinschätzung bezüglich ihrer mathematischen Leistungsfähigkeit haben, kann hingegen empirisch bestätigt werden. Deskriptive Auswertungen zeigen, dass sich *F-&A*-Studierende im Durchschnitt insbesondere zu Semesterbeginn maßgeblich unterschätzen. Wohingegen sich *Management*-Studierende nicht nur zu Studienbeginn, sondern über das gesamte Semester hinweg, optimistischer einschätzen als *F-&A*-Studierende, obwohl - oder gerade weil - sie durchschnittlich schlechter abschneiden. Der Vergleich von der relativ simplen^{xxv} Probeklausur (T_3) mit der verhältnismäßig schwierigen finalen Klausur (T_4) zeigt, dass sich die Studierenden insgesamt bei einfachen Problemstellungen tendenziell unterschätzen und bei schweren Aufgaben überschätzen. Dieses empirische Ergebnis entspricht dem *Hard-Easy Effect* (vgl. [22]), der besagt, dass die subjektive Wahrnehmung des Schwierigkeitsgrad ist nicht objektiv mit dem zugrundeliegenden Test vereinbar ist.

Mithilfe des Rangsummentests nach Wilcoxon wird lediglich zu Semesterbeginn in T_2 ein Gruppenunterschied bezüglich des Confidence-Levels festgestellt. Im weiteren Verlauf des Semesters konnten keine weiteren statistisch signifikanten Unterschiede zwischen *F-&A*-Studierenden und *Management*-Studierenden identifiziert werden. Möglicherweise bestehen Unterschiede weiterhin, werden allerdings in der vorliegenden Untersuchung durch andere Effekte, wie potentiell dem *Hard-Easy Effect*, verborgen. In der weiterführenden Forschung von Gruppenunterschieden könnte in dem Zusammenhang beispielweise für die *explizite* Einschätzung des Schwierigkeitsgrads der Klausuren und das individuelle Lernverhalten als Einflussfaktoren zusätzlich kontrolliert werden.

Um die einleitende Frage, welche die Determinanten für Fehleinschätzungen sind, zu beantworten, wird die *Mathematical Confidence* auf soziodemographische, die Schulbildung und das wirtschaftswissenschaftliche Studium betreffende Prädiktoren regressiert. Bei Annahme einer zugrundeliegenden Zufallsstichprobe zeigen OLS-Regressionen über verschiedene Modellspezifikationen hinweg, dass das Confidence-Level zu Semesterbeginn und -ende von der Schwerpunktwahl und dem Geschlecht abhängt. Entgegen der ursprünglichen Erwartung weisen *F-&A-Studierende* ein im Schnitt um rund 20 Prozent geringeres Confidence-Level auf als *Management*-Studierende. Dies gilt sowohl zu Semesterbeginn als auch zu Semesterende. Auch weibliche Studierende schätzen sich durchschnittlich weniger optimistisch ein als männliche Studierende (vgl. [3, S. 140], [27]). Insbesondere zu Studienbeginn beeinflusst der vermeintliche Migrationshintergrund das Ausmaß an Confidence signifikant. Dieser Effekt ist in T_4 insignifikant, sodass zu Semesterende *ceteris paribus* Studierende mit und ohne Migrationshintergrund über ein ähnliches Confidence-Level verfügen. Für das Ausmaß an Confidence zu Semesterende ist außerdem die Leistung im Abitur relevant, wobei gilt, je schlechter die Leistung sowohl im Abitur allgemein als auch im Mathe-Abitur, desto höher das Confidence-Level. Das Mathe-Abitur hat dabei einen geringeren Einfluss als das Abitur. Der Einfluss des Abiturs hängt außerdem vom Geschlecht des Studierenden und dessen Schwerpunktwahl ab. Der Einfluss des Mathe-Abiturs hingegen ist ausschließlich abhängig von dem Geschlecht. Die zusätzliche Untersuchung, ob die zugrundeliegende Stichprobe eine Zufallsstichprobe ist, konnte über die Heckman-Korrektur anhand der inversen Mill's Ratio nicht abgelehnt werden. Kontrolliert man dennoch für das Problem *Sample-Selection* über das Heckman-Zweistufen-Verfahren, so zeigt sich, dass in T_2 lediglich das Geschlecht und in T_4 das Geschlecht und die Leistung im Mathe-Abitur signifikanten Einfluss auf das Confidence-Level haben. Dieses Ergebnis divergiert stark von dem der OLS-Regressionen. In weiterer Forschung könnten diese Ergebnisse mit denen anderer Methoden zur Korrektur von *Sample-Selection* verglichen werden, beispielsweise die Heckman-Korrektur unter Anwendung einer *Maximum-Likelihood*-Methode.

Bei allen Regressionen ist allerdings auffällig, dass das Bestimmtheitsmaß der verschiedenen Spezifikationen durchweg sehr gering ist. Dies ist ein Indikator für das Fehlen von Prädiktoren relevanten Erklärungsgehalts. In weiter-

^{xxv}Die Einschätzungen des Schwierigkeitsgrades der jeweiligen Klausur entspricht dem Feedback der Studierenden des betreffenden Semesters. Auch die Leistungen in den jeweiligen Klausuren können diese Einschätzungen bestätigen.

gehenden Untersuchungen könnten Daten zu folgenden weiteren Prädiktoren erhoben werden: Lernstundenanzahl zu Semesterbeginn und -ende, Ausmaß an Prüfungsangst, Ausmaß der Mathematik-Affinität oder -Aversion^{xxvi}, Einschätzung der Wichtigkeit der Mathematik für das wirtschaftswissenschaftliche Studium und den Vertrauensgrad in eigene Einschätzungen. Zudem sollte nach Bearbeitung jeder Klausur der empfundene Schwierigkeitsgrad abgefragt werden, um so besser für Unterschiede in den Schwierigkeitsgraden der Klausuren zu korrigieren.

Wichtig bei der Interpretation der Ergebnisse ist zu beachten, dass in dieser Arbeit ausschließlich die *Mathematical Confidence* betrachtet wird. Um eine allgemeine Aussage über das Confidence-Level der beiden Schwerpunkte treffen zu können, sollte die *Mathematical Confidence* in weitergehender Forschung beispielweise mit der *volks-* oder *betriebswirtschaftlichen Confidence* verglichen oder kombiniert werden. Des Weiteren sollte sich nicht ausschließlich auf eine Kennzahl basierend auf dem Konzept der Overestimation beschränkt werden. Es sollten zudem andere Kennzahlen in Betracht gezogen werden, um verschiedene Facetten der Overconfidence abzubilden: Better-than-average-Effekt, Brier-Score in Verbindung mit Wahrscheinlichkeitsurteilen und die Overprecision.

Falls in Übungsklausuren anderer Module ebenfalls nach deren Bearbeitung eine Selbsteinschätzung abgefragt wird, könnte die somit explizite Äußerung einer eingeschätzten Punktzahl, die bewusste Reflexion gerade erbrachter Leistung und ein regelmäßiges Feedback eine akkurate Selbsteinschätzung begünstigen.

^{xxvi}Laut Bescherer (2003, S.93f) leiden 21 Prozent von insgesamt knapp 1000 befragten Studienfängerinnen und -anfängern nicht-mathematischer Studiengänge unter der Angst ihr Studium wegen mathematischer Inhalte nicht bewältigen zu können. Darüber hinaus sehen befragte Hochschullehrende, unter anderem der Wirtschaftswissenschaften, diese Angst als berechtigt an, da sie „sogar zwischen 30 und über 50 Prozent der Studierenden wenig oder gar nicht ausgesprägte Kenntnisse in Mathematik bescheinigen.“

Anhang

A Zusatz: Deskriptive Statistiken

A.1 Relative Selbsteinschätzung und Relatives Ergebnis

Im Folgenden werden die Variablen $RelSE_t$ und $RelRES_t$ erklärt und die zugehörigen deskriptiven Statistiken pro Messzeitpunkt $T_1 - T_4$ und Schwerpunkt aufgeführt. Die Variable $RelSE_t$ beschreibt die relative Selbsteinschätzung pro Klausur in $t \in \{1, 4\}$, das heißt die Einschätzung der eigenen Leistung im Verhältnis zur möglichen Höchstpunktzahl. Die Variable $RelRES_t$ repräsentiert das relative Klausurergebnis pro Klausur in $t \in \{1, 4\}$, das heißt das individuelle Ergebnis im Verhältnis zu der Höchstpunktzahl des getrimmten Datensatzes. Dabei ist wichtig zu unterscheiden, dass die Selbsteinschätzung vor Abgabe der jeweiligen Klausur abgegeben wird, ohne das Ergebnis zu kennen, sodass die Studierenden theoretisch davon ausgehen die Höchstpunktzahl erreichen zu können. Dahingegen wird $RelRES_t$ im Verhältnis zu der getrimmten Höchstpunktzahl berechnet, um für die verschiedenen Schwierigkeitsgrade zu korrigieren und Vergleichbarkeit zu fördern. Tabelle 8 zeigt die deskriptiven Statistiken der relativen Selbsteinschätzungen und Klausurergebnissen pro Schwerpunkt. Auffällig ist, dass die relative Selbsteinschätzung von *F-&A*-Studierenden zu jedem Zeitpunkt T größer ist als die der *Management*-Studierenden. Allerdings berechtigt, da die durchschnittliche Leistung der *F-&A*-Studierenden ebenfalls in jedem Zeitpunkt höher ist.

Tabelle 8: Deskriptive Statistiken $RelSE_t$ und $RelRES_t$

Variable	Finance & Accounting				Management			
	Abs. Hfk.	Arithm. Mittel	Median	SE	Abs. Hfk.	Arithm. Mittel	Median	SE
$RelSE_1$	36	0,3941	0,3333	0,2177	41	0,3035	0,2778	0,1621
$RelSE_2$	37	0,5348	0,5556	0,1673	53	0,5013	0,5	0,1713
$RelSE_3$	57	0,6573	0,6667	0,1957	66	0,6095	0,6667	0,1590
$RelSE_4$	32	0,7341	0,75	0,1122	40	0,6760	0,6667	0,0888
$RelRES_1$	44	0,4505	0,4423	0,2239	48	0,3640	0,3462	0,1886
$RelRES_2$	38	0,6189	0,6266	0,1693	57	0,5270	0,5190	0,1744
$RelRES_3$	58	0,7419	0,7569	0,1535	69	0,6985	0,7083	0,1676
$RelRES_4$	56	0,6699	0,7243	0,1961	65	0,5980	0,5701	0,1745

A.2 Regressionen der Mathematical Confidence

Die folgenden Regressionsergebnisse wurden pro Messzeitpunkt mithilfe der relativen Selbsteinschätzungen $RelSE_1 - RelSE_4$ als unabhängige Variablen und der relativen Klausurergebnisse $RelRES_1 - RelRES_4$ als abhängige Variablen generiert^{xxvii}. Die Regressionstabellen beschreiben die in 6.1 dargestellten Regressionsgeraden der Mathematical Confidence. Die Ergebnisse wurden mithilfe einer OLS-Regression generiert, dabei wurden robuste Standardfehler genutzt, um für Heteroskedastizität zu korrigieren.

Die jeweiligen Ergebnisse werden nicht weitergehend erläutert, zumal die Interpretierbarkeit der im Zuge der Regression geschätzten Koeffizienten aufgrund der geringen Anzahl an Beobachtungen fragwürdig ist. Die Regressionsergebnisse dienen lediglich als Ergänzung zu den Diagrammen in 6.1 und dahingehend zur Veranschaulichung der Dimension der Unterschiede in der Kalibrierung zwischen den Studierenden der Schwerpunkte *Finance & Accounting* und *Management*.

^{xxvii}Die deskriptiven Statistiken zu den Variablen $RelSE_{1-4}$ und $RelRES_{1-4}$ sind in Tabelle 8, Anhang A.1 zu finden.

Tabelle 9: Regressionen von $RelRES_t$ auf $RelSE_t$

Finance & Accounting				
Prädiktoren	$RelRES_1$	$RelRES_2$	$RelRES_3$	$RelRES_4$
$RelSE_1$	0,840*** (0,1024)			
$RelSE_2$		0,6048*** (0,1382)		
$RelSE_3$			0,3145** (0,1125)	
$RelSE_4$				0,6147** (0,2524)
CONS	0,1453*** (0,0507)	0,3036*** (0,0768)	0,5367*** (0,0757)	0,2581 (0,1951)
Anzahl Beobachtungen	36	37	57	32
R^2	0,6496	0,3812	0,1588	0,1535
Anmerkungen: Angegeben sind die unstandardisierten Regressionskoeffizienten sowie die robusten Standardfehler (in Klammern). Signifikante Koeffizienten sind folgendermaßen hervorgehoben: */**/** = $p < 0,1/0,05/0,01$.				
(Forts.)				
Management				
Prädiktoren	$RelRES_1$	$RelRES_2$	$RelRES_3$	$RelRES_4$
$RelSE_1$	0,4853*** (0,1053)			
$RelSE_2$		0,349** (0,109)		
$RelSE_3$			0,4218*** (0,1152)	
$RelSE_4$				0,4882* (0,2505)
CONS	0,1992*** (0,0421)	0,3451*** (0,0563)	0,4351*** (0,0699)	0,2672 (0,1683)
Anzahl Beobachtungen	41	53	66	40
R^2	0,2636	0,1392	0,1633	0,0588
Anmerkungen: Angegeben sind die unstandardisierten Regressionskoeffizienten sowie die robusten Standardfehler (in Klammern). Signifikante Koeffizienten sind folgendermaßen hervorgehoben: */**/** = $p < 0,1/0,05/0,01$.				

B Zusatz: Mittelwertvergleich

B.1 Der Shapiro-Wilk-Test

Der Shapiro-Wilk-Test überprüft die Annahme, dass die der Stichprobe zugrundeliegende Zufallsvariable normalverteilt ist, unabhängig von der Stichprobengröße. Er ist also auch anwendbar, wenn die Gruppengröße $n_j < 30$ ist. Der Test zählt zu den konservativen Testverfahren, das heißt die Ergebnisse werden weniger schnell signifikant Bühner and Ziegler [7, S. 96]. Die Nullhypothese besagt, dass eine Normalverteilung vorliegt. Sie wird verletzt, falls sich für die folgende Teststatistik zu kleine Werte ergeben:

$$W^2 = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Hierbei steht $x_{(i)}$ für das i -te Element der aufsteigend geordneten Stichprobe und a_i für die jeweilig tabellierten Gewichte. Die Nullhypothese wird zum Signifikanzniveau α verworfen, falls gilt: $W^2 \leq W_{\alpha}^2$.

Die folgende Tabelle zeigt die Teststatistiken W^2 und die zugehörigen kritischen Werte W_{α}^2 pro Schwerpunkt für alle Messzeitpunkte. Die Werte der Teststatistik wurden mithilfe der statistischen Software *Stata* generiert, die zugehörigen kritischen Werte wurden aus der Shapiro-Wilk-Tabelle entnommen.

Tabelle 10: Ergebnisse des Shapiro-Wilk-Tests *MCON1-MCON4*

Variable	Finance & Accounting			Management		
	W^2	$W_{\alpha=0,05}^2$	Entscheidung	W^2	$W_{\alpha=0,05}^2$	Entscheidung
<i>MCON</i> ₁	0,96584	0,941	H ₀ nicht ablehnen	0,9907	0,935	H ₀ nicht ablehnen
<i>MCON</i> ₂	0,9321	0,95	H ₀ ablehnen	0,9752	0,936	H ₀ nicht ablehnen
<i>MCON</i> ₃	0,8502	0,95	H ₀ ablehnen	0,8977	0,95	H ₀ ablehnen
<i>MCON</i> ₄	0,84267	0,94	H ₀ ablehnen	0,7529	0,93	H ₀ ablehnen

B.2 Der Levene-Test

Der Levene-Test ist ein Signifikanztest zur Überprüfung von Gleichheit der Fehlervarianzen (Varianzhomogenität) zweier (oder mehrerer) Gruppen, siehe Bühner and Ziegler [7, S. 379] sowie Bortz and Schuster [4, S.129]. Die Nullhypothese besagt, dass Varianzhomogenität vorliegt, das heißt, die Populationsvarianzen über die verschiedenen Gruppen sind konstant und gleich Bortz and Schuster [4, S.129]:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2.$$

Die Alternativhypothese lautet entsprechend, dass sich die Varianzen von (mindestens) zwei Stichproben unterscheiden, sodass Varianzheterogenität vorliegt:

$$H_1 : \exists (\sigma_i^2, \sigma_j^2) : \sigma_i^2 \neq \sigma_j^2 \text{ mit } i \neq j.$$

Falls sich für die folgende Teststatistik zu kleine Werte ergeben, wird die Nullhypothese abgelehnt:

$$W = \frac{(n-m) \cdot \sum_{j=1}^m n_j \cdot (\bar{Y}_j - \bar{Y})^2}{(m-1) \cdot \sum_{j=1}^m \sum_{i=1}^{n_j} (Y_{j,i} - \bar{Y}_j)^2} \text{ wobei}$$

- $Y_{j,i} = |X_{j,i} - \bar{X}_j|$ mit $j = 1, \dots, m$ und $i = 1, \dots, n_j$
- $\bar{Y}_j = \frac{1}{n_j} * \sum_{i=1}^{n_j} Y_{j,i}$ als die mittlere, absolute Abweichungen der Gruppe j
- $\bar{Y} = \frac{1}{n} * \sum_{j=1}^m n_j \bar{Y}_j$ als die mittlere Abweichung der mittleren gruppenspezifischen Abweichungen
- $n = \sum_{j=1}^m n_j$ als die Summe aller Beobachtungen der Gruppen j

Zur Berechnung der Teststatistik werden die arithmetischen Mittel verwendet, wie auch ursprünglich von Levene vorgeschlagen. Brown und Fortsynthe schlagen in ihrer Studie 1974 vor, statt des arithmetischen Mittels robustere Lagemaße zu verwenden: Dem Stichprobenmedian und den um 10 Prozent getrimmten Mittelwert. Die Verwendung des Stichprobenmedians schien insbesondere bei Vorliegen der Cauchy-Verteilung sinnvoll und die Nutzung des getrimmten Mittel bei einer sehr schiefen Verteilung [1]. Da in dem vorliegenden Datensatz weder eine Cauchy-Verteilung noch eine sehr schiefe Verteilung vorliegt, wird das arithmetische Mittel nach Levene genutzt.

Die Nullhypothese wird zum Signifikanzniveau α verworfen, falls gilt: $W < F_{m-1, n-m; 1-\alpha}$.

Tabelle 11: Ergebnisse des Levene-Tests $MCON1-MCON4$

Variable	W	$F_{m-1, n-m; 0,95}$	p-Wert	Entscheidung
$MCON_1$	4,8251	3,9600	0,0311	H_0 nicht ablehnen
$MCON_2$	8,8055	3,9470	0,0039	H_0 nicht ablehnen
$MCON_3$	0,0083	3,9360	0,9274	H_0 ablehnen
$MCON_4$	2,1124	3,9780	0,1506	H_0 ablehnen

B.3 Der Rangsummentest nach Wilcoxon

Der verteilungsunabhängige Rangsummentest nach Wilcoxon führt keinen Mittelwertvergleich im engeren Sinne durch, sondern untersucht Rangplatzunterschiede. Dabei beschreibt $MCON_{F\&A}$ das Confidence-Level von F-&A-Studierenden und $MCON_M$ das Confidence-Level der Management-Studierenden. Im Folgenden sei m_i der Stichprobenumfang der Zufallsvariable $MCON_{F\&A}$ und n_j der Stichprobenumfang der Zufallsvariable $MCON_M$. Die Nullhypothese besagt, dass jeder Rangplatz für beide Gruppen gleich wahrscheinlich ist, siehe Bühner and Ziegler [7, S. 282], dass die Populationen der beiden Stichproben also derselben Verteilung folgen [21, S. 170]:

$$H_0 : F_{F\&A} = F_M$$

Die Alternativhypothese besagt entsprechend, dass sich die Verteilungen der Populationen unterscheiden [21, S. 171]:

$$H_1 : F_{F\&A} > F_M \text{ oder } F_{F\&A} < F_M$$

wobei $F_{F\&A} > F_M \implies F(x) > F_M(x) \forall x \in \mathbb{R}$

Den Verteilungen der Grundgesamtheiten der Gruppen werden stetige Verteilungsfunktionen unterstellt. Um die Nullhypothese zu prüfen, werden die beiden Stichproben kombiniert und die Realisationen der Zufallsvariablen $MCON_{F\&A}$ ($x_{F\&A,i}$) und der Zufallsvariablen $MCON_M$ ($y_{M,j}$) aufsteigend in einer Reihe geordnet (Bsp.: $x_{F\&A,i(1)} < x_{F\&A,i(2)} < y_{M,j(1)} < x_{F\&A,i(3)} < y_{M,j(2)} < \dots < x_{F\&A,i(m)} < y_{M,j(n)}$). Den aufsteigend geordneten Messwerten der werden jeweils Ränge zugeordnet, siehe Bühner and Ziegler [7, S. 281]. Zur Berechnung der Teststatistik werden die Ränge der kleineren Stichprobe aufsummiert:

$$W_N = \sum_{i=1}^r R(X_i)$$

wobei X_i die Realisationen der kleineren Stichprobe seien

Wichtig zu beachten ist, dass Bindungen innerhalb einer Gruppe mit Durchschnittsrängen versehen werden, Bindungen zwischen den Rängen der verschiedenen Gruppen sind irrelevant für die Berechnung der Teststatistik [10, S. 169].

Die Nullhypothese wird bei dem vorliegenden zweiseitigen Test abgelehnt, falls gilt: $W_N \leq w_{\frac{\alpha}{2}}$ oder $W_N > w_{1-\frac{\alpha}{2}}$. Die kritischen Werte können aus der Tabelle für den Wilcoxon-Rangsummentest abgelesen werden.

Da sämtliche Stichproben für die beiden Schwerpunkte in jedem Messzeitpunkt einen größeren Stichprobenumfang als 25 Beobachtungen umfassen, kann die Teststatistik durch eine Normalverteilung approximiert werden [10, S. 169], sodass unter der Nullhypothese gilt:

$$E(W_N) = \frac{m(N+1)}{2}$$

$$Var(W_N) = \frac{mn(N+1)}{12}$$

Entsprechend berechnet sich die Teststatistik für $m, n \rightarrow \infty, \frac{m}{n} \rightarrow \lambda \neq 0$ asymptotisch:

$$Z = \frac{W_N - m(N+1)/2}{\sqrt{mn(N+1)/12}} \sim N(0, 1)$$

Falls Bindungen bei Beobachtungen innerhalb der Stichprobe der kleineren Gruppe vorliegen, muss die Varianz folgendermaßen korrigiert werden:

$$\text{Var}(W_N^*) = \frac{mn(N+1)}{12} - \frac{mn}{12N(N-1)} * \sum_{j=1}^s (b_j^3 - b_j)$$

wobei s = Anzahl der Bindungsgruppen und b_j = Ausmaß der Bindung^{xxviii}

Da hier keine Bindungen vorliegen, muss die Varianz nicht angepasst werden. Die Nullhypothese wird bei dem vorliegenden zweiseitigen Test abgelehnt, falls gilt: $Z \leq z_{\frac{\alpha}{2}}$ oder $Z > z_{1-\frac{\alpha}{2}}$.

C Zusatz: Multivariate Analysen

C.1 Korrelationen der Prädiktoren mit dem Kriterium

Um zu entscheiden, welche Prädiktoren in die Regressionsmodelle aufzunehmen sind, ist es sinnvoll, die Korrelationen zwischen den Prädiktoren und dem Kriterium $MCON_2$ bzw. $MCON_4$ zu betrachten. Zwar sind so keine Aussagen über Kausalzusammenhänge möglich, allerdings werden lineare Einflüsse sichtbar.

Tabelle 12: Korrelationen der Prädiktoren mit dem Kriterium

Kriterium	Prädiktoren				
	<i>DMANFIN</i>	<i>DGENDER</i>	<i>DMIGR</i>	<i>DFIRST</i>	<i>DGOODNEWS₁</i>
<i>MCON₂</i>	-0,2144	-0,2682	0,1351	-0,1914	-0,2694
<i>MCON₄</i>	-0,1547	-0,1303	0,0318	-0,0741	-0,2333

(Forts.)

Kriterium	<i>DGOODNEWS₃</i>	<i>BIRTH</i>	<i>ABI</i>	<i>MABI</i>
<i>MCON₂</i>	nicht sinnvoll	-0,0702	0,0624	-0,1785
<i>MCON₄</i>	-0,0564	-0,1507	0,3829	-0,1620

C.2 Korrelationen der Prädiktoren

Im Kontext der Überprüfung der Gauss-Markov-Annahmen zur Durchführung einer OLS-Regression werden in der folgenden Tabelle die Korrelationen der unabhängigen Variablen betrachtet. Hohe Interkorrelationen bzw. Multikollinearität verhindern signifikante Inferenz auf Basis der Regressionsanalyse, zumal die Schätzung der Parameter

^{xxviii}Im Falle einer Dreierbindung ergibt sich beispielsweise der Wert $b_j = 3$.

unter Multikollinearität nicht möglich ist. Im Falle von Kollinearität ist die Konstanthaltung der anderen Prädiktoren im Zuge der Messung des Einflusses eines einzigen Prädiktors nicht möglich, sodass die Interpretierbarkeit der Regressionskoeffizienten, der Standardfehler und Konfidenzintervalle beeinträchtigt wird (vgl. [29])

Wichtig zu beachten ist, dass abhängig von dem Skalenniveau der Variablen unterschiedliche Maßzahlen zur Bestimmung des statistischen Zusammenhangs zweier Variablen genutzt werden müssen. Werden im Folgenden ausschließlich dichotome Variablen (Dummies) verglichen, so wird der *Phi-Koeffizient* (ϕ) als Zusammenhangsmaß genutzt (vgl. [9, S. 92]). Für den Vergleich zweier binär definierten Variablen ist der Phi-Koeffizient äquivalent zu Cramer's V (vgl. [2]). Weil *Stata* lediglich die Option der Berechnung von Cramer's V bietet und wegen der Äquivalenz von Cramer's V und dem Phi-Koeffizienten, können laut Acock (2008) die von *Stata* berechneten Werte als Phi-Koeffizienten interpretiert werden. Dies macht in der folgenden Analyse Sinn, da negative Zusammenhänge auftreten, die über Cramer's V allein nicht erklärt werden können, da Cramer's V ausschließlich auf dem Intervall $[0; 1]$ definiert ist. Tabelle 9 enthält also die Phi-Koeffizienten der dichotomen Prädiktoren. In Abhängigkeit der Größe des Wertes ϕ kann auf die Stärke des Zusammenhangs geschlossen werden, das Vorzeichen ist dabei allerdings abhängig von der Definition der Variablen, dass heißt der Zuordnung der Werte 1 und 0, und sollte daher nicht interpretiert werden (vgl. [30, S. 211]).

Wert	Interpretation
$ \phi \approx 0$	kein Zusammenhang
$ \phi \approx 1$	starker Zusammenhang

In Tabelle 9 werden zudem die Korrelationen der intervallskalierten Prädiktoren mithilfe der *Produkt-Moment-Korrelation* (Pearsons r) berechnet. Dabei gelten die folgenden Interpretationen der Werte des Korrelationskoeffizienten r ([6, S. 346]):

Wert	Interpretation
$r \in [0, 0; 0, 2]$	keine bzw. sehr geringe Korrelation
$r \in (0, 2; 0, 5]$	geringe Korrelation
$r \in (0, 5; 0, 7]$	mittlere Korrelation
$r \in (0, 7; 0, 9]$	hohe Korrelation
$r \in (0, 9; 1, 0]$	sehr hohe Korrelation

Die Korrelationen einer intervallskalierten Variable mit einer dichotomen Variable werden im Folgenden mithilfe der *punktbiserialen Korrelation* (r_{pb}) berechnet ([7, S. 626]). Da der punktbiserialer Korrelationskoeffizient auf den Produkt-Moment-Korrelationskoeffizienten zurückzuführen ist, ist die Interpretation der Werte von r_{pb} äquivalent zu denen des Pearsons r (siehe oben).

Die folgende Tabelle enthält jegliche quantifizierte Zusammenhänge zwischen den relevanten Prädiktoren: Die Phi-Koeffizienten dichotomer Variablen, die Produkt-Moment-Korrelationen der intervallskalierten Variablen und die punktbiserialen Korrelationen von dichotomen und intervallskalierten Variablen.

Tabelle 13: Korrelationen der Prädiktoren

	<i>DMANFIN</i>	<i>DGENDER</i>	<i>DMIGR</i>	<i>DFIRST</i>	<i>DGOODNEWS₁</i>	<i>DGOODNEWS₃</i>	<i>BIRTH</i>	<i>ABI</i>	<i>MABI</i>
<i>DMANFIN</i>	1								
<i>DGENDER</i>	-0,1408	1							
<i>DMIGR</i>	0,1487	0,1341	1						
<i>DFIRST</i>	-0,0324	-0,1124	-0,0821	1					
<i>DGOODNEWS₁</i>	0,0807	0,1015	0,0008	-0,0539	1				
<i>DGOODNEWS₃</i>	-0,0175	0,1044	-0,0295	0,0230	0,0640	1			
<i>BIRTH</i>	-0,1824	-0,0290	-0,0980	0,0968	0,0929	-0,0039	1		
<i>ABI</i>	-0,0489	-0,1107	0,0398	-0,0345	0,0014	0,0613	-0,3951	1	
<i>MABI</i>	0,0372	-0,0015	-0,1945	0,1036	0,0365	0,0256	0,2340	-0,5439	1

Literatur

- [1] Levene Test for Equality of Variances. URL <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.htm>.
- [2] Alan C. Acock. *A Gentle Introduction to Stata*. Stata Press, 2 edition, 2008.
- [3] Christine Bescherer. Selbsteinschätzung mathematischer Studierfähigkeit von Studienanfängerinnen und Empirische Untersuchung und praktische Konsequenz. 2003.
- [4] Jürgen; Bortz and Christof Schuster. *Statistik für Human- und Sozialwissenschaftler*. SpringerMedizin, Springer Verlag GmbH, 7 edition, 2010. ISBN 9783642127694. URL 978-3-642-12769-4.
- [5] Glenn W. Brier. Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, 78 (1):1-3, 1950. doi: 10.1126/science.27.693.594. URL <http://docs.lib.noaa.gov/rescue/mwr/078/mwr-078-01-0001.pdf>.
- [6] Achim Bühl. *SPSS 16: Einführung in die moderne Datenanalyse*. Pearson Deutschland GmbH, 2008.
- [7] Markus; Bühner and Matthias Ziegler. *Statistik für Psychologen und Sozialwissenschaftler*. Pearson Deutschland GmbH, Hallbergmoos, 2009. ISBN 978-3-8273-7274-1.
- [8] Shawn; Bushway, Brian D.; Johnson, and Lee Ann Slocum. Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology. *Journal of Quantitative Criminology*, 23(2):151-178, 2007.
- [9] Thomas Cleff. *Deskriptive Statistik und moderne Datenanalyse: Eine computergestützte Einführung mit Excel, PASW (SPSS-9 und STATA)*. Gabler Verlag, Springer, 2011.
- [10] Christine Duller. *Einführung in die nichtparametrische Statistik mit SAS und R*, volume 1. Physica-Verlag, Springer, Heidelberg, 2008. ISBN 9788578110796. doi: 10.1017/CBO9781107415324.004.
- [11] Uwe; Engel and Oliver B. Schmidt. Unit- und Item-Nonresponse. In *Handbuch Methoden der empirischen Sozialforschung*, pages 331-348. Springer Fachmedien, Wiesbaden, 2014.
- [12] Andy Field. *Discovering Statistics Using SPSS (and sex and drugs and rock'n'roll)*. Sage Publications, 3 edition, 2009.
- [13] Lin Freitag. Manager halten sich für toller, als sie sind, dec 2015.
- [14] Robert Gillenkirch. Stichwort: Entscheidungstheorie, . URL <http://wirtschaftslexikon.gabler.de/Archiv/56961/entscheidungstheorie-v8.html>.
- [15] Robert Gillenkirch. Stichwort: Urteilsheuristik, . URL <http://wirtschaftslexikon.gabler.de/Archiv/830359610/urteilsheuristik-v1.html>.
- [16] Markus; Glaser and Martin Weber. Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, 32(1):1-36, 2007. doi: <http://www.jstor.org/stable/41953463>.
- [17] Daniela Grieco and Robin M. Hogarth. Overconfidence in absolute and relative performance: The regression hypothesis and Bayesian updating. *Journal of Economic Psychology*, 30(5):756-771, 2009. ISSN 01674870. doi: 10.1016/j.joep.2009.06.007. URL <http://linkinghub.elsevier.com/retrieve/pii/S0167487009000646>.

- [18] James J. Heckman. The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for such Models. *Annals of Economic and Social Measurement*, 5 (4):475–492, 1976.
- [19] Rüdiger Hossiep and Patrick Schardien. Der Sandwich- Chef : Kündigungsgrund Nummer 1, 2009.
- [20] Kienbaum and Human Resource & Management Consulting. High Potentials Studie 2011/2012. Technical report, 2012. URL <http://www4.kienbaum.de/shopweb/printview.aspx?product=375>.
- [21] Jürgen; Lehn and Helmut Wegmann. *Einführung in die Statistik*. Teubner Studienbücher Mathematik, 5 edition, 2005. ISBN 978-3-486-57687-0. doi: 10.1524/9783486710922.
- [22] S Lichtenstein and B Fischhoff. Do those who know more also know more about how much they know? *Organizational Behaviour and Human Performance*, 20(3052):159–183, 1977.
- [23] Julia; Mendzheritskaya, Nora; Dörmann, and Julia Boser. Auswirkung der Mathematik-Vorkurse auf Studienerfolg in der Studieneingangsphase des wirtschaftswissenschaftlichen Studiums. Technical report, Methodenzentrum Sozialwissenschaften, 2015.
- [24] Don A. Moore and Daylian M. Cain. Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition. *Organizational Behavior and Human Decision Processes*, 103(2): 197–213, 2007. ISSN 07495978. doi: 10.1016/j.obhdp.2006.09.002. URL <http://linkinghub.elsevier.com/retrieve/pii/S0749597806000896>.
- [25] Don A Moore and Paul J Healy. The Trouble With Overconfidence. *Psychological review*, 115(2):502–517, 2008. ISSN 0033-295X. doi: 10.1037/0033-295X.115.2.502.
- [26] Allan H Murphy. A New Vector Partition of the Probability Score, 1973. ISSN 0021-8952. URL [http://journals.ametsoc.org/doi/abs/10.1175/1520-0450\(1973\)012<0595:ANVPOT>2.0.CO;2%delimitter"026E30F\\$n\(null\)](http://journals.ametsoc.org/doi/abs/10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2%delimitter).
- [27] Clifford Nowell and Richard M Alston. I Thought I Got an A! Overconfidence Across the Economics Curriculum. *The Journal of Economic Education*, 38(2):131–142, 2007. ISSN 0022-0485. doi: 10.3200/JECE.38.2.131-142. URL <http://www.tandfonline.com/doi/abs/10.3200/JECE.38.2.131-142{\protect\T1\textbraceleft}{#}{\protect\T1\textbraceright}.UfgUlo21Fmc>.
- [28] Horst; Rottmann and Benjamin R. Auer. Stichwort: Heckman-Zweistufen-Verfahren. URL <http://wirtschaftslexikon.gabler.de/Archiv/89487/heckman-zweistufen-verfahren-v8.html>.
- [29] Christian Schendera. *Regressionsanalyse mit SPSS*. De Gruyter Oldenbourg, 2014.
- [30] Siegfried Schumann. *Repräsentative Umfrage: Praxisorientierte Einführung in die empirische Methoden und statistischen Analyseverfahren*. Oldenbourg Verlag, 2012.
- [31] Teddy Seidenfeld. Calibration, Coherence, and Scoring Rules. *Philosophy of Science*, 52(2):274–294, 1985.
- [32] StataCorp. 2015. heckman- Heckman selection model.
- [33] Ola Svenson. Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, 47(2):143–148, 1981.
- [34] Alex B.; Van Zant and Don A. Moore. Avoiding the Pitfalls of Overconfidence while Benefiting from the Advantages of Confidence. *California Management Review*, 55(2):5–23, 2013.
- [35] Jeffrey M. Wooldridge. *Introductory Econometrics - A Modern Approach*. Cengage Learning, 5 edition, 2013.