

# Masterprüfung

**Fach:** Panel- und Evaluationsverfahren

**Prüfer:** Prof. Regina T. Riphahn, Ph.D.

## Vorbemerkungen:

**Anzahl der Aufgaben:** Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen.

**Bewertung:** Es können maximal 60 Punkte erworben werden. Die Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.

**Erlaubte Hilfsmittel:**

- Tabellen der statistischen Verteilung (sind der Klausur beigefügt)
- Taschenrechner
- Fremdwörterbuch

**Wichtige Hinweise:**

- Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten exakten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert.
- Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

**Aufgabe 1:****[12 Punkte]**

Sie interessieren sich dafür, ob sich das von Personen im Laufe ihrer Betriebszugehörigkeit erworbene betriebsspezifische Humankapital in höherer Entlohnung niederschlägt. Dazu möchten Sie unter Verwendung eines Datensatzes von Beschäftigten folgenden Zusammenhang schätzen:

$w$  = *Bruttostundenlohn, gemessen in €.*

$educ$  = *Schulbildung, gemessen in Jahren.*

$exp$  = *Berufserfahrung, gemessen in Jahren.*

$tenure$  = *Betriebszugehörigkeit im aktuellen Anstellungsverhältnis, gemessen in Jahren.*

- a) Erklären Sie kurz anhand eines Beispiels, warum  $tenure$  in der obigen Schätzgleichung eine endogene erklärende Variable sein könnte. (2 Punkte)
- b) Sie erwägen, für  $tenure$  zwei Instrumente zu nutzen, die sich auf das letzte Beschäftigungsverhältnis beziehen:

$quit$  = *1, wenn Person innerhalb der letzten 10 Jahre ihren Job gekündigt hat, sonst 0.*

$closure$  = *1, wenn Person innerhalb der letzten 10 Jahre aufgrund einer Betriebschließung ihren Job verloren hat, sonst 0.*

Diskutieren Sie die Exogenität der beiden Instrumente am Beispiel. (4 Punkte)

- c) Sie gehen davon aus, dass  $closure$  die Exogenitätsbedingung erfüllt und möchten die Gültigkeit der Exogenitätsbedingung für  $quit$  mit Hilfe eines Overidentifying Restriction Tests prüfen.
- Erklären Sie, wie Sie im vorliegenden Beispiel zur Ermittlung der Teststatistik vorgehen müssten. (4 Punkte)
  - Erklären Sie, warum der Overidentifying Restriction Test nicht bei genau identifizierten Modellen angewendet werden kann. (2 Punkte)

**Aufgabe 2:****[15 Punkte]**

Im Potential Outcome Approach bezeichnet man den Treatmentindikator mit  $D_i$  und die Potential Outcomes mit  $y_i(\cdot)$ , wobei  $y_i(1)$  die Outcomevariable mit Treatment und  $y_i(0)$  die Outcomevariable ohne Treatment ist.

- a) Erläutern Sie, was ein Average Treatment Effect (ATE) und was ein Average Treatment Effect on the Treated (ATT) ist. Stellen Sie zusätzlich beide Effekte mit Hilfe der oben genannten Notation formal dar. (6 Punkte)
- b) Was wird durch  $E[y_i(1) | D_i=1] - E[y_i(0) | D_i=0]$  allgemein abgebildet? Welche Bedingung muss gelten, damit der angegebene Ausdruck dem ATT entspricht? Zeigen Sie die Bedingung formal durch geeignete Umformung des Ausdrucks. Erläutern Sie die Bedingung inhaltlich. (6 Punkte)
- c) Welcher Treatmenteffekt kann identifiziert werden, wenn  $y_i(0)$  und  $y_i(1)$  unabhängig von  $D_i$  sind? Begründen Sie kurz und beschreiben Sie knapp, wie eine Situation herbeigeführt werden kann, in der diese Annahme zutrifft. (3 Punkte)

**Aufgabe 3:****[17 Punkte]**

Ihnen stehen Daten zu 15000 Personen zur Verfügung, die 2010 arbeitslos waren. Einige dieser Arbeitslosen haben während ihrer Arbeitslosigkeit an IT-Fortbildungskursen teilgenommen. Sie möchten nun prüfen, ob sich die Teilnahme an diesen Kursen positiv auf die Beschäftigungswahrscheinlichkeit im Jahr 2011 ausgewirkt hat. Sie möchten für die Fragestellung ein Matching-Verfahren verwenden.

Folgende Variablen stehen Ihnen dabei zur Verfügung:

- teilnahme* = 1, wenn Person an der Fortbildungsmaßnahme teilgenommen hat, sonst 0.
- job* = 1, wenn Person 2011 nicht mehr arbeitslos, sonst 0.
- age* = Alter der Person in Jahren
- educ* = Anzahl der Jahre im Bildungssystem
- dist* = Entfernung zum nächsten Fortbildungszentrum in Fahrminuten
- kidsLT6* = Anzahl der Kinder unter 6 Jahren.

- a) Erläutern Sie die Grundidee des Matching-Verfahrens und die strenge CIA Annahme. (4 Punkte)
- b) Erläutern Sie den Unterschied zwischen exaktem Matching und Propensity Score Matching. (3 Punkte)
- c) Erläutern Sie, wie Nearest Neighbor Matching mit Zurücklegen funktioniert und was man dabei unter einem Caliper versteht. (4 Punkte)
- d) Die nachfolgende Tabelle zeigt Ergebnisse eines Nearest Neighborverfahrens ohne Zurücklegen und ohne Caliper. Prüfen und diskutieren Sie die Qualität des Matchings. (Hinweis: Die Spalte „%bias“ misst den standardisierten Bias und die Spalte „t-Test“ gibt das Testergebnis eines Mittelwertvergleichs zwischen Kontroll- und Treatmentgruppe wieder.) Unterbreiten Sie zwei Vorschläge, wie im Beispiel die Matchingqualität verbessert werden kann. (6 Punkte)

Variable	Sample	Mean		%reduct		t-test	
		Treated	Control	%bias	bias	t	p> t
age	Unmatched	31.527	29.415	72.2		46.69	0.000
	Matched	30.922	30.943	-0.7	99.0	-0.38	0.707
educ	Unmatched	12.243	12.904	-25.4		-16.42	0.000
	Matched	12.406	12.388	0.7	97.3	0.36	0.717
dist	Unmatched	36.655	40.329	-31.7		-20.71	0.000
	Matched	38.918	39.265	-3.0	90.5	-1.71	0.087
kidsLT6	Unmatched	.13657	.01764	38.4		25.37	0.000
	Matched	.04153	.04153	0.0	100.0	-0.00	1.000

#### Aufgabe 4:

[16 Punkte]

Sie interessieren sich dafür, ob Arbeitslose bei längerer Dauer der Arbeitslosigkeit ihre Lohnerwartungen senken. Ihnen liegen Paneldaten zu 466 Arbeitslosen auf Monatsbasis vor. Der Datensatz enthält folgende Informationen:

- w* = *Anspruchslohn des Arbeitslosen, gemessen in €.*
- dal* = *Dauer der Arbeitslosigkeit, gemessen in Monaten.*
- male* = *1, wenn männlich, sonst 0.*
- income* = *Nichtarbeitseinkommen des Arbeitslosen (z.B. Zinseinkommen), gemessen in €.*

Sie möchten ein lineares Modell mit *w* als abhängiger Variable und den übrigen drei Variablen als erklärenden Variablen schätzen.

- a) Insgesamt enthält der Datensatz 1887 Beobachtungen. Handelt es sich damit um ein Balanced Panel? Begründen Sie Ihre Aussage. (2 Punkte)

- b) Erläutern Sie anhand eines Beispiels, warum *dal* eine endogene erklärende Variable sein könnte. Geben Sie auch ein Beispiel. (2 Punkte)
- c) Sie möchten den Zusammenhang zwischen *w* und *dal* mit einem Fixed Effects Schätzer bestimmen. Stellen Sie für das gegebene Beispiel die konkrete Schätzgleichung einer LSDV Regression mit allen nutzbaren Regressoren auf. (Hinweis: Neu gebildete Variablen müssen definiert werden.) (4 Punkte)
- d) Welche Schätzung könnten Sie alternativ durchführen, um ebenfalls den Fixed Effects Schätzer zu erhalten? Stellen Sie auch hierfür die Schätzgleichung mit allen nutzbaren Regressoren auf. (Hinweis: Neu gebildete Variablen müssen definiert werden.) (4 Punkte)
- e) Welche Annahme ist nötig, damit eine Fixed Effects Schätzung konsistente Ergebnisse liefert? Erläutern Sie diese Annahme kurz verbal. Ist diese Annahme im vorliegenden Fall erfüllt? Begründen Sie Ihre Antwort. (4 Punkte)