

# Erklärung von schulischem (Miss-)Erfolg im Senegal mit Hilfe von fixed effects Modellen

ABSCHLUSSARBEIT

zur Erlangung des akademischen Grades  
Bachelor of Science  
im Studiengang Statistik

an der Wirtschaftswissenschaftlichen Fakultät der  
Humboldt-Universität zu Berlin

vorgelegt von

**Sebastian Fehrl**

Matrikel-Nr. 3764760 (FU)

**1. Prüfer: Dr. Bernd Droge**

**2. Prüfer: Prof. Dr. Bernd Rönz**

Berlin, den 1. September 2005

# Inhaltsverzeichnis

<b>Einführung</b>	<b>5</b>
Fragestellung der Untersuchung . . . . .	6
Bedeutung und Wahrnehmung des Problems . . . . .	7
Forschungsinteresse . . . . .	7
<b>I. Theorie</b>	<b>9</b>
<b>1. Stand der Forschung</b>	<b>10</b>
1.1. Wiederholproblematik . . . . .	10
1.2. Schulabbruch . . . . .	12
1.3. Unterrichtsformen . . . . .	13
<b>2. Senegal</b>	<b>15</b>
2.1. Datenbasis . . . . .	15
2.2. PASEC-Studie . . . . .	15
2.3. Weitere Studien zum Bildungssystem Senegals . . . . .	18
<b>3. Mängel bisheriger Studien</b>	<b>21</b>
3.1. Omitted Ability . . . . .	21
3.2. Abhilfe . . . . .	21
<b>II. Daten und Methoden</b>	<b>23</b>
<b>4. Noch einmal zur Datenbasis</b>	<b>24</b>
4.1. Datensatz . . . . .	24
4.2. Messung der Fähigkeit der Schüler . . . . .	24
<b>5. Methoden</b>	<b>27</b>
5.1. Paneldaten-Modelle für Lernfortschritt . . . . .	27
5.2. Modelle für Wiederholwahrscheinlichkeit . . . . .	31
<b>III. Analyse</b>	<b>33</b>

<b>6. Datenüberprüfung</b>	<b>34</b>
<b>7. Analysen Lernfortschritt</b>	<b>39</b>
7.1. Lernfortschritt Mathematik . . . . .	39
7.2. Positionsveränderung . . . . .	44
7.3. Interpretation der Ergebnisse . . . . .	48
<b>8. Analysen Wiederholwahrscheinlichkeit</b>	<b>51</b>
8.1. Zeitraum 95/96 - 97/98 . . . . .	51
8.2. Querschnitt 95/96 . . . . .	55
8.3. Interpretation der Ergebnisse . . . . .	59
<b>9. Fazit</b>	<b>62</b>
<b>Anhang</b>	<b>64</b>
<b>Literatur</b>	<b>67</b>

# Tabellenverzeichnis

0.1. Einschulungsquote für die Grundschule (v.H.) . . . . .	5
0.2. Wiederholraten nach Regionen in Afrika . . . . .	6
0.3. Wiederholraten europäischer Länder . . . . .	7
4.1. Fallzahlen pro Welle . . . . .	24
6.1. Vergleich von robusten Schätzern und standardmäßigen . . . . .	35
7.1. Tests auf Heteroskedastizität . . . . .	40
7.2. Korrelation der Residuen (Signifikanz) . . . . .	40
7.3. Schätzergebnisse fixed effects mit robusten Standardfehlern . . . . .	41
7.4. Hausman Test . . . . .	42
7.5. Schätzergebnisse random effects GLS . . . . .	42
7.6. Multikollinearitätsüberprüfung . . . . .	43
7.7. Tests auf Heteroskedastizität . . . . .	45
7.8. Korrelation der Residuen (Signifikanz) . . . . .	45
7.9. Schätzergebnisse fixed effects mit robusten Standardfehlern . . . . .	46
7.10. Hausman Test . . . . .	46
7.11. Schätzergebnisse random effects GLS . . . . .	47
7.12. Multikollinearitätsüberprüfung . . . . .	48
8.1. Schätzergebnis r.e. Probit Modell . . . . .	51
8.2. Schätzergebnisse gepooltes Probit Modell . . . . .	52
8.3. Schätzergebnisse gepooltes Logit-Modell . . . . .	53
8.4. Vorhersagen anhand des Modells . . . . .	54
8.5. Schätzergebnisse Logit Modell . . . . .	56
8.6. Schätzergebnisse mit Quartil-Dummies . . . . .	57
8.7. Schätzergebnisse mit Leistungsvariable . . . . .	57
8.8. Vorhersagen anhand des Modells . . . . .	58
8.9. Schätzergebnisse Logistische Regression . . . . .	60
9.1. Variablenbeschreibung . . . . .	64
9.2. Variablenbeschreibung . . . . .	65

# Abbildungsverzeichnis

2.1. PASEC-Panel Schema . . . . .	16
4.1. ICC . . . . .	25
5.1. Individuelle Intercepts ohne Korrelation . . . . .	27
5.2. Individuelle Intercepts mit Korrelation . . . . .	28
6.1. Mathematik . . . . .	34
6.2. Französisch . . . . .	35
6.3. Verteilung Leistungsvariable Anfang 95/96 . . . . .	36
6.4. Verteilung Leistungsvariable Ende 95/96 . . . . .	36
6.5. Verteilung Leistungsvariable Ende 96/97 . . . . .	37
6.6. Verteilung Leistungsvariable Ende 97/98 . . . . .	37
6.7. Verteilung Positionsveränderung . . . . .	38
7.1. Residuen Boxplots über Jahr . . . . .	40
7.2. Residuen Boxplots über die Jahre . . . . .	44
8.1. H.-L. Plot: Veränderung in $\chi^2$ . . . . .	54
8.2. H.-L. Plot: Veränderung in Devianz . . . . .	55
8.3. H.-L. Plot: Veränderung in $\chi^2$ . . . . .	58
8.4. H.-L. Plot: Veränderung in Devianz . . . . .	59

# Einführung

Möchte man die Qualität der Schulbildung in Entwicklungsländern verbessern, so kommt es auf zwei wichtige Fragestellungen an:

1. Wie kann der Lernerfolg der Kinder erhöht werden?
2. Wie erreicht man, dass möglichst viele Kinder überhaupt erst in den Genuss einer Schulbildung kommen?

In vielen Ländern Afrikas bekommt bei weitem nicht jedes Kind einen Platz in der Schule (Tab.0.1). Gleichzeitig ist die Wiederholrate in diesen Ländern (insbesondere im frankophonen Sub-Sahara-Afrika) sehr hoch (Tab.0.2). Im Senegal haben 65% der Personen, die die Grundschule beendet haben, mindestens eine Klasse wiederholt, oftmals mehrere [Dumas and Lambert, 2005]. Das hat zur Folge, dass viele Plätze in den Klassen von Wiederholern blockiert sind und nicht für andere Kinder zur Verfügung stehen. Somit ist insbesondere in diesen Ländern die Frage von Relevanz, ob Wiederholen ein effektives Mittel ist, um schwächeren Schülern zu helfen. Ist es das nicht, wie etliche Studien gezeigt haben, stellen die hohen Raten eine enorme Ressourcenverschwendung dar. Eine Sichtung der Befunde der Studien zur Effektivität des Wiederholens, ergab bei insgesamt 63 Studien, 54 negative Urteile und nur 9 positive [Barnes, 1999].

Tabelle 0.1.: Einschulungsquote für die Grundschule (v.H.)

Länder	Jungen	Mädchen
Mali	32	19
Mozambique	69	52
Niger	37	21
Nigeria	100	79
Rwanda	78	76
Senegal	65	47
Sudan	59	45
Tanzania	71	69
Uganda	78	63
Zaire	78	61
Zambia	85	79

Quelle: Japan International Cooperation Agency (JICA) (World Education Report, 1995 / UNESCO Statistical Yearbook, 1995)

Tabelle 0.2.: Wiederholraten nach Regionen in Afrika

Region	Wiederholrate
Afrika insgesamt	17.5 v.H.
Anglophones Afrika	10.3 v.H.
Frankophones Afrika	19.8 v.H.
Lusophones Afrika	23.4 v.H.
OECD (zum Vergleich)	2.2 v.H.

Quelle: Association for the development of education in Africa (ADEA), (UNESCO Statistics (2002), MINEDAF (2002))

Auch für den Senegal, auf den sich diese Untersuchung bezieht, gibt es bereits eine Studie, die zu einer negativen Beurteilung der Effektivität kommt [PASEC, 2004]. Dieser Frage wird hier nicht erneut nachgegangen.

Die sich an diesen Befund anschließende Frage ist nun, wie sich diese missliche Lage ändern lässt. An dieser Stelle kann man sich natürlich fragen, weshalb die Regierung nicht einfach die Rate per Gesetz senken kann. Dies scheint jedoch nicht so einfach zu sein. So gibt es im Senegal bereits eine Regierungsanweisung das Niveau auf 10% zu fixieren. In einer Studie, die weiter unten ausführlich dargestellt wird (PASEC-Studie), ist jedoch trotzdem von einer aktuellen Rate von 14% die Rede. Entscheidend ist also, zunächst die Ursachen des Problems herauszufinden, um sie bekämpfen zu können. Wie in verschiedenen Studien gezeigt, treten hohe Wiederholraten besonders in den strukturschwächsten Regionen auf, so dass sie nicht nur ein Problem an sich darstellen, sondern auch als Indikator für weitere Probleme angesehen werden können.

## Fragestellung der Untersuchung

In dieser Arbeit sollen folgende Fragestellungen bearbeitet werden:

- Zunächst soll geklärt werden, welche Faktoren sich auf die schulischen Leistungen auswirken, um hier Verbesserungsvorschläge machen zu können.
- Welche weiteren Faktoren über die Schülerleistung hinaus die Wahrscheinlichkeit sitzen zu bleiben erhöhen, soll in einem zweiten Schritt untersucht werden.

Im ersten Teil wird die Suche nach Faktoren hauptsächlich auf Lehreigenschaften gerichtet, da diese Variablen für Reformvorschläge am interessantesten sind und über die Zeit variieren. Das ist von Relevanz, da fixed effects Modelle verwandt werden, die zeitinvariante Variablen eliminieren. Diese Modelle sollen eingesetzt werden, um unbeobachtete Heterogenität zwischen den Schülern auszuschalten. Insbesondere in der Bildungsforschung ist dies sinnvoll, um die Effekte der beobachtbaren Variablen von den nicht beobachtbaren kognitiven Fähigkeiten der Schüler (Ability) zu trennen und konsistent schätzen zu können.

## Bedeutung und Wahrnehmung des Problems

Im Vergleich zu vielen Entwicklungsländern, sind die Wiederholraten westlicher Industrieländer weit niedriger (Tab. 0.3). Trotzdem wird selbst in Frankreich die, im europäischen Vergleich leicht überdurchschnittliche, Wiederholrate als Problem angesehen [Forestier, 2004].

Tabelle 0.3.: Wiederholraten europäischer Länder

Länder	Wiederholrate
Deutschland	2 v.H.
Frankreich	4 v.H.
Finnland	0 v.H.
Schweden	0 v.H.

Quelle: [Paul and Troncin, 2004]

Will man also mehr Kindern den Schulzugang ermöglichen, ohne die Bedingungen für die anderen Kinder zu verschlechtern, drängt sich das Absenken der Wiederholrate geradezu auf. Die UNESCO hat das Problem bereits erkannt und als Ziel für die Entwicklungsländer ausgegeben, die Wiederholrate auf höchstens 10% zu senken.

UNICEF stellt fest, dass hohe Wiederholraten besonders in armen Regionen mit schlecht ausgebildeten Lehrern und Eltern auftritt [Cortina, 1999]. Somit sprechen noch weitere Gründe dafür, die hohen Wiederholraten als Problem anzusehen. Es wäre zu einfach, schlicht staatliche Verbote hoher Raten zu fordern, da Systeme mit automatischem Versetzen u.a. eine höhere Kompetenz der Lehrer im Umgang mit heterogenen Klassen erfordern.

Eine Strategie von UNICEF zur Verbesserung der Grundschulausbildung setzt bei der Lehrerweiterbildung an, sie lautet [Cortina, 1999](S.57):

A widespread belief among teachers and parents is that repetition is necessary for children. Uninspired teaching styles, lack of pedagogical training, and the need to improve the training of teachers working in the schools are variables that have direct influence on high repetition rates. An important element of reform is to change the pedagogical practices of teachers through the use of in-service training.

Weitere Ansätze zur Verbesserung der Schulsysteme werden weiter unten diskutiert.

## Forschungsinteresse

Die Konferenz der Bildungsminister der französischsprachigen Länder (CONFEMEN) hat im Rahmen des Programms zur Analyse der Bildungssysteme der CONFEMEN (PASEC) Datenerhebungen in afrikanischen Staaten durchgeführt, mit der Zielsetzung die größten

Probleme der Bildungssysteme und deren Ursachen herauszuarbeiten, um Lösungsansätze entwickeln zu können. Neben anderen Instituten ist auch das Hamburgische Welt-Wirtschafts Archiv (HWWA) an der Analyse der Daten beteiligt. Und zwar in Person von Katharina Michaelowa, die im wissenschaftlichen Beirat von PASEC sitzt. Über ein Praktikum bei Frau Michaelowa, bin ich auf das Themengebiet und an die Daten gekommen und die Ergebnisse fließen möglicherweise über diesen Weg in die PASEC-Forschung mit ein.

Teil I.  
Theorie

# 1. Stand der Forschung

## 1.1. Wiederholproblematik

Wie in der Einführung erwähnt, gab es bereits eine ganze Reihe von Studien zu der Fragestellung, ob Wiederholenlassen eine effektive pädagogische Maßnahme ist, oder nicht. Der überwiegende Anteil der Studien kam zu dem Ergebnis, dass es nicht effektiv ist, Schüler wiederholen zu lassen. Um zu verstehen, weshalb dies der Fall ist, betrachten wir einige Studien etwas näher.

### Ineffizient und Ungerecht

In einer aktuellen Studie aus Frankreich [Paul and Troncin, 2004] werden die wichtigsten Gründe gegen das Wiederholen aufgezeigt.

- Ein Schüler, der die erste Grundschulklasse wiederholt, erzielt zu Beginn der 3. Klasse bestenfalls gleich gute oder schlechtere Ergebnisse als ein Schüler gleichen Ausgangsniveaus, der nicht wiederholt. Zudem verliert der betroffene Schüler ein ganzes Jahr.
- Als Gründe für die schwachen Leistungen werden genannt, dass Wiederholer weniger motiviert sind und mit einem Stigma des Versagens behaftet sind, welches ihr Bild bei den Lehrern negativ beeinflusst.
- Wiederholen ist sehr ungerecht, da es keinen objektiven Maßstab gibt, sondern die Position des Schülers in der Klasse den Ausschlag gibt. D.h. ein Schüler in einer guten Klasse bleibt eher sitzen, als ein Schüler in einer schwächeren Klasse, obwohl beide das gleiche Leistungsniveau haben.
- Zudem bleiben Schüler mit schwachem sozialen Hintergrund häufiger sitzen. Besonders selten bleiben Lehrerkinder sitzen. Auch dies jeweils im Vergleich zu gleich starken Schülern betrachtet.
- Die PISA-Studie hat gezeigt, dass Länder mit geringerer Wiederholerrate besser abschneiden als solche mit hoher.
- Zudem verursacht das Sitzenbleiben enorme Kosten für das jeweilige Bildungssystem (in Frankreich ca. 2 Mrd. Euro jährlich).

Sicherlich ist die Situation Frankreichs aber nicht mit der in Entwicklungsländern direkt vergleichbar. Zum Beispiel könnte man davon ausgehen, dass ein Stigmatisierungseffekt sich in Entwicklungsländern nicht so stark auswirkt, wo z.B. im frankophonen Afrika

etwa 20% der Schüler einer Klasse sitzen bleiben und die Klassengröße häufig zwischen 60 und 100 Kindern liegt. Der motivationszerstörenden Stand als einsamer Versager, sollte so hier nicht Auftreten.

Zunächst aber noch ein Blick nach Lateinamerika, wo das Problem hoher Wiederholraten schon länger im Bewusstsein der Politiker ist und entsprechend Programme zur Verbesserung der Situation aufgelegt wurden.

In einem Artikel [Schiefelbein and Schiefelbein, 1999] zu dem Problem des Wiederholens und der Schulqualität in Chile werden folgende Gründe für hohe Wiederholraten genannt:

- Mangel an Vorschulerziehung. Kinder, die vor der Einschulung nicht in die Vorschule gehen können, erzielen deutlich schwächere Leistungen.
- Starre Lehrformen, die sich an den 'Durchschnittsstudent' richten, obwohl in den Klassen sehr heterogene Alterszusammensetzungen und somit auch unterschiedliche geistige Entwicklungsstadien vorliegen.
- Starre Schulzeiten auch in ländlichen Regionen, die einen großen Nachteil für Schüler bedeuten, die bei der Ernte helfen müssen und dann nicht in der Lage sind den verpassten Stoff nachzuholen.
- Schlechte Qualität des Unterrichts, was das Eingehen auf Vorwissen der Schüler, Verdeutlichen von Zusammenhängen und Realitätsbezügen angeht (Frontalunterricht).
- Schlechte Lehrerbezahlung, die zur Folge hat, dass der Lehrerberuf für gut ausgebildete Absolventen unattraktiv ist.
- Schlechte Unterrichtsmaterialien und Anleitungen für die Lehrer
- Unterrichtsausfall, der selten durch Vertretung ersetzt wird.
- Kultur des Wiederholens. Viele Lehrer sind von Sinn und Notwendigkeit des Wiederholens überzeugt.
- Schlechte Ernährung und Gesundheit armer Schüler, die die Konzentrationsfähigkeit herabsetzt
- Zu strenge Standards für die Versetzung

Diese Auflistung der Gründe macht deutlich, dass hohe Wiederholraten im wesentlichen Folgen sozialer Ungleichheit, schlechter Lehrerbildung und dem damit verbundenen schlechten Unterricht sind. Die Situation in Chile kommt der in Afrika schon viel näher, als jene in Frankreich. Als Konsequenzen aus den hohen Wiederholraten ergeben sich in Chile [Schiefelbein and Schiefelbein, 1999]:

- Heterogenität in der Alterszusammensetzungen der Klassen.

- Schlechter Einfluss der unmotivierten Wiederholer auf das Lernklima in der Klasse.
- Schlechte Resultate der Wiederholer und reduzierte Wahrscheinlichkeit, die Grundschule abzuschließen.
- Höhere Kosten pro Schüler für den Staat. Chile gibt jährlich 300 Millionen Dollar für Schulplätze von Wiederholern aus.
- Soziale Ungleichheit, da insbesondere Kinder aus armen schlecht gebildeten Familien vom Wiederholen betroffen sind.

Die Auflistung der Konsequenzen macht deutlich, dass hohe Wiederholraten zu schlechteren Lernbedingungen an den Schulen und zu mehr sozialer Ungleichheit führen. Zudem bedeuten sie eine enorme Ressourcenvergeudung, gerade in Entwicklungsländern mit knappen Budgets. Ähnliche Befunde finden sich in Berichten über Brasilien, Mexiko und Argentinien [Randall et al., 1999].

## 1.2. Schulabbruch

Eine weitere schwerwiegende Folge des Sitzenbleibens ist häufig Schulabbruch. So ist eine positive Korrelation zwischen Wiederholrate und Abbruchrate in verschiedenen Ländern aufgezeigt worden [Barnes, 1999]. Ob ein Schüler wiederholen muss, oder nicht, ist ein guter Indikator dafür, ob er die Schule abbricht oder nicht. So bleiben in Lateinamerika 40% derer, die sitzen bleiben im darauffolgenden Jahr der Schule fern, während nur 10% der Nichtwiederholer im selben Jahr abbricht [Barnes, 1999]. Im Senegal sieht es genauso dramatisch aus. Von den Schülern, die sitzen bleiben, brechen 46% später die Grundschule ab, von denen die mehrfach sitzen bleiben sogar 60%, im Gegensatz zu allerdings auch sehr hohen 30%, die auch ohne sitzen zu bleiben abbrechen [PASEC, 2004]. Dieser Zusammenhang ist nicht besonders verwunderlich. Zum einen, stellt es für die Kinder eine Abschreckung dar, wenn sie wiederholen müssen, zum anderen spielen in armen Familien die Kosten, die der Verzicht auf die Arbeitskraft des Kindes für ein zusätzlichen Jahr bedeuten, eine Rolle. Dies ist insbesondere in ländlichen Regionen, in denen die Kinder auf dem Feld helfen müssen, der Fall.

Schulabbruch ist deshalb ein wichtiges Problemfeld in den Entwicklungsländern. Insbesondere in Lateinamerika, das die höchste Rate von Kindern hat, die die Grundschule nicht beenden, nämlich über 50% [Cortina, 1999], ist dieses Thema bekannt und einige Programme wurden ins Leben gerufen, um die Abbruchrate zu senken.

So startete beispielsweise Brasilien 1995, in einem Bundesstaat (dem Hauptstadt-Distrikt), ein Programm, dass die Abbruchrate unter armen Familien senken sollte. Jeder Familie, die noch alle ihre Kinder zwischen 7 und 14 Jahren in der Schule hatte, wurde ein Mindestgehalt bezahlt. Kinder aus Familien, die in das Programm aufgenommen wurden, hatten eine deutlich niedrigere Abbruchrate und eine um 10 % niedrigere Wiederholrate, als Kinder aus Familien, die nicht aufgenommen wurden [Pino and Koslinski, 1999].

Ein weiteres erfolgreiches Programm stellte das von internationalen Organisationen finanzierte Programm Escuela Nueva in Kolumbien dar, dass u.a. mit automatischem Versetzen von der ersten in die zweite Klasse arbeitete [Cortina, 1999].

Im Senegal liegt die Schulabbruchrate in der Grundschule, wenn man den neuesten Zahlen von UNICEF glauben kann, mittlerweile unter einem Drittel. So beträgt der Anteil der Schüler, die die 5.Klasse erreicht haben an allen Schülern zwischen 1998 und 2001 68%, nach einer anderen UNICEF-Quelle zwischen 1997 und 2003 sogar 93% (beide Angaben stammen von der UNICEF-Homepage). Beide Zahlen sind relativ hoch verglichen mit den lateinamerikanischen Angaben, allerdings weichen sie stark von einander ab. In der PASEC-Panelstudie (1995-2000) ist die Abbruchrate deutlich höher (s.o.), wobei möglicherweise nicht alle Schüler, die aus dem Panel gefallen sind, die Schule tatsächlich abgebrochen haben und in der Studie somit zu hohe Angaben gemacht werden.<sup>1</sup>

### 1.3. Unterrichtsformen

Aus erziehungswissenschaftlicher Sicht hat Schulversagen viel mit der Philosophie der Erziehung zu tun [Barnes, 1999]. Barnes unterscheidet zwei idealtypische Methoden. Zum einen die 'transmission' Methode und zum anderen die 'transaction' Methode. Erstere bedeutet, kurz gesagt, das Lehren von Fakten, die zweite die Erziehung zum selbstständigen lernen, durch Förderung der Verknüpfung neuer Inhalte mit bereits Gelerntem und die Kreativität der Schüler stimulierende Unterrichtsmethoden.

In der 'transmission' Methode besteht die strikte Aufteilung zwischen dem Lehrer, der das Wissen hat und dem Schüler, der sich dieses aneignen soll. Dies geschieht im Wesentlichen durch Auswendiglernen durch den Lehrer vorgegebener Wissenspakete, oft isoliert von Anwendungsmöglichkeiten. Umfang und Niveau der Wissenspakete ist organisiert aufgeteilt in verschiedene Schulstufen. Der Unterrichtsstil ist autoritär und kennt nur die Richtung Lehrer-Schüler. Wenn die Standards einer Stufe nicht erreicht werden, soll der Schüler sitzen bleiben, um den Rückstand aufzuholen. Von den Schülern wird erwartet, dieselben Dinge im selben Tempo zu lernen. Dieser Ansatz führt zu höheren Wiederholraten, da die unterschiedlichen Schüler de facto nicht im selben Tempo lernen können.

Das 'transaction' Erziehungsmodell geht vom Lernenden aus, der sich sein eigenes Wissen aneignet durch aktives Auseinandersetzen mit neuen Informationen und der Verbindung mit bereits Bekanntem. Lesen und Schreiben wird beispielsweise nicht mit isolierten Textbuchübungen gelehrt, sondern als Mittel um Geschichten zu erzählen und Briefe zu schreiben. Der Lehrer ist bemüht eine vielfältige Lernumgebung mit möglichst viel Realitätsbezug zu kreieren, die die Schüler zu eigener Kreativität und Assoziationsleistungen animiert. Da in diesem Ansatz keine starren Standards erfüllt werden sollen, sondern Lernen als individueller Prozess verstanden wird, der bei den Schülern unterschiedlich

---

<sup>1</sup>Dass die angegebenen Zahlen teilweise stark von einander abweichen, obwohl sie z.T. von der selben Quelle stammen, ist ein Indiz für die schlechte Datenbasis. Sie sollen deshalb nicht für einen exakten Vergleich von Afrika und Lateinamerika herhalten, sondern lediglich in ihrer ungefähren Größenordnung die Parallelen aufzeigen.

abläuft, treten hier niedrigere Wiederholraten auf.

Mit Klassengrößen von häufig weit über 50 Schülern lässt sich ein Unterrichtsstil, der stärker auf die individuelle Entwicklung einzelner Schüler eingeht, schwerer umsetzen. So dominiert in vielen Entwicklungsländern auch aus diesem Grund Frontalunterricht, nach dem Schema der 'transmission' Methode. Hinzu kommt die schlechte Lehreraus- und fortbildung.

## 2. Senegal

### 2.1. Datenbasis

Bevor wir einen Blick auf die Ergebnisse verschiedener Studien zum Bildungssystem Senegals werfen, sind ein paar Informationen zur Datenbasis hilfreich, auf der alle Studien basieren. PASEC hat im Senegal zwischen 1995 und 2000 eine Datenbasis aufgebaut.

Es liegt ein Paneldatensatz über 5 Jahre vor, der 1995/96 mit 1975 Schülern startete, von denen allerdings bis 99/00 995 verloren gegangen sind (Abb.2.1). Schulabbruch ist wie oben diskutiert ein weiteres großes Problem in Entwicklungsländern.

- 95/96 waren alle Schüler in der ersten Klasse. 99/00 sind diejenigen Schüler, die nie wiederholt haben, in der 6. Klasse.
- Es wurden alle Wiederholer weiterverfolgt, so dass deren Entwicklung untersucht werden kann.
- In jedem Jahr wurden die Schüler in Mathematik und Französisch getestet.
- Zudem gab es (leider inkonsequenterweise nicht in jedem Jahr) einen Schülerfragebogen, einen Lehrerfragebogen und einen Direktorenfragebogen.
- Die Stichprobe ist mit einem geschichteten Cluster Sampling Verfahren gezogen worden, wobei in der letzten Stufe ganze Klassen die Klumpen dargestellt haben.

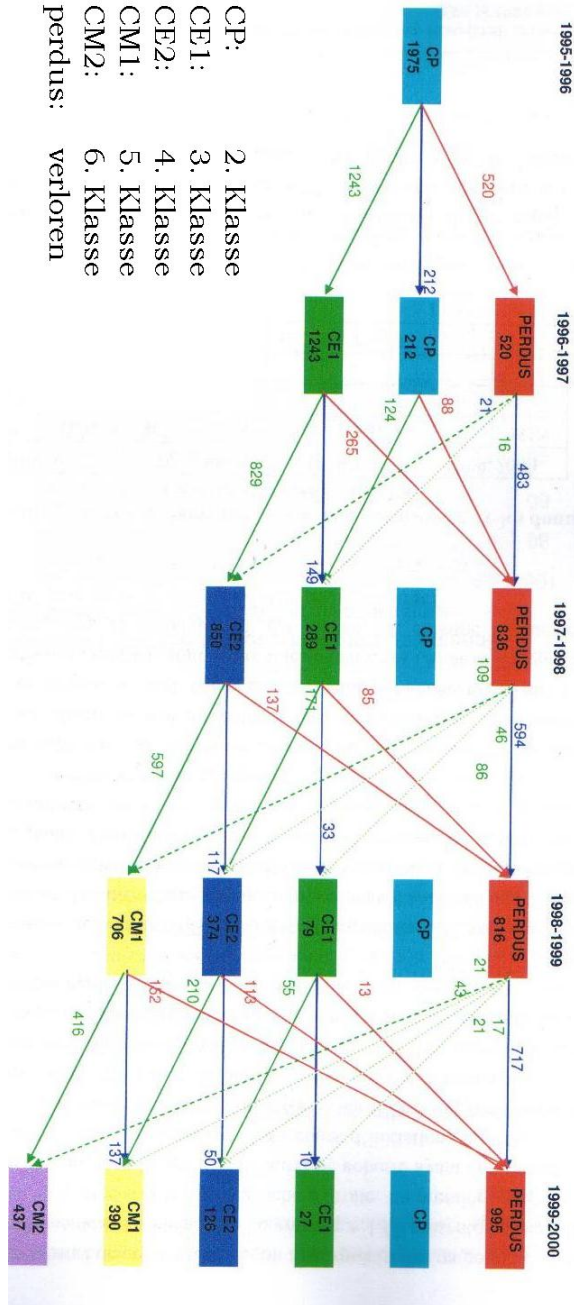
### 2.2. PASEC-Studie

Betrachten wir nun die Ergebnisse, die das PASEC-Team selbst aus den Daten gewonnen hat [PASEC, 2004]. Ziel dieser Untersuchung war es, die Effizienz des Wiederholens zu prüfen.

Die Studie ist in mehrere Schritte unterteilt. Zunächst werden die Faktoren, die sitzen bleiben erklären untersucht. Hierfür werden, für jede Stufe einzeln, Probit-Modelle spezifiziert und geschätzt mit der abhängigen Variable 'sitzen geblieben ja/nein'. Als signifikant negative Einflussfaktoren auf die Wahrscheinlichkeit sitzen zu bleiben haben sich dabei folgende gezeigt:

- Anzahl der richtigen Antworten in den Tests
- ländliche oder Stadtrandgend im Gegensatz zu Stadt

Abbildung 2.1.: PASEC-Panel Schema



Quelle: PASEC-Studie

Als signifikant positive Einflüsse ergaben sich:

- Rang in der Klasse
- Höhe der Schulbildung des Lehrers
- Höhe der Berufsausbildung des Lehrers

Im einem weiteren Teil der Studie wird im Querschnitt deskriptiv herausgestellt, dass etwa 30 % der Wiederholer gar nicht zu den schwächsten Schülern gehören, sondern hier eher Rang des Schülers in einer guten Klasse entscheidend war. Zudem wurde gezeigt, dass Kinder mit sozial schwachem Hintergrund häufiger sitzen bleiben und dass Lehrer mit längerer Ausbildung und höheren Abschlüssen mehr Schüler sitzen lassen. Außerdem glauben 77% der Lehrer, dass Wiederholen effektiv oder sogar sehr effektiv ist.

Hierauf folgt der interessanteste Teil der Studie, der sich in 2 Schritte unterteilt. Im ersten Schritt wird untersucht, welche Auswirkungen das Wiederholen einer Klasse auf die Position des Schülers innerhalb der Klasse hat.

- Hier stellte sich für alle Klassenstufen heraus, dass die Wiederholer ihre Position verbesserten und im Leistungsmittelfeld der Klasse angelangten. Im darauf folgenden Jahr verlieren sie jedoch wieder etwas an Boden gegenüber ihren Klassenkameraden.

Wie die PASEC Leute auch betonen, steigert der Wiederholer sein Wissen in dem Jahr des Wiederholens weit weniger stark als seine neuen Klassenkameraden. Seine vergleichsweise trotzdem gute Position im Klassenschnitt, erklärt sich lediglich durch seinen Wissensvorsprung zu Beginn des Schuljahres.

Der Effekt, dass der Schüler seine Position innerhalb der Klasse deutlich verbessern kann, ist für die Lehrer der offensichtlichste und könnte die Ursache dafür sein, dass die überwiegende Zahl der Lehrer Wiederholen befürwortet und als effektiv einschätzt.

Im 2. Schritt wird ausgenutzt, dass die Testbögen der Französisch- und Mathematiktests für die aufeinander folgenden Stufen einen gemeinsamen Teil an Testproblemen enthielten. Somit ist ein Vergleich im Lernfortschritt der Schüler, die wiederholten, mit denen die trotz gleichen Leistungsstandes, versetzt wurden möglich. Es wurden nun Modelle geschätzt, die als abhängige Variable das Abschneiden im gemeinsamen Teil des Tests im Jahre  $(t+1)$  hatten und als unabhängige Variablen das Testergebnis des Jahres  $(t)$ , in dem noch alle in derselben Stufe waren, sowie die Variable 'Wiederholer ja/nein'. Zusätzlich wurden in einigen Modellen Kontrollvariablen eingebaut, die sich aber als nicht besonders erklärungsstark herausstellten. Für jede Stufe wurden solche Modelle geschätzt.

- Ergebnis war, dass das Wiederholen in keiner Stufe einen signifikant positiven Effekt auf die Leistung im gemeinsamen Testteil hatte, in einigen Stufen sogar einen signifikant negativen.

Hieraus ließ sich dann schließen, dass Wiederholen nicht effektiv für das Fortkommen der Schüler ist und die Lehrer einer 'optischen Täuschung' (S.102) aufsitzen.

## 2.3. Weitere Studien zum Bildungssystem Senegals

Einige weitere Studien wurden mit den PASEC-Daten im Querschnitt durchgeführt. Zunächst betrachten wir drei Studien, die sich mit Faktoren schulischen Erfolgs befassen, die am HWWA erstellt wurden. Die erste befasst sich ganz explizit mit der Suche nach Faktoren von Lernfortschritt [Michaelowa, 2001], die zweite spezifischer mit sogenannten 'peer effects', also Effekten auf das Lernen durch andere Schüler, z.B. durch Bücherbesitz anderer Schüler in der Klasse [Michaelowa and Fröhlich, 2005]. Die dritte Studie beschäftigt sich, unter anderem, mit Lehrerzufriedenheit und dem Effekt auf die Leistung der Schüler [Michaelowa, 2002].

Folgende Faktoren haben sich in den Studien als relevante Einflüsse auf die Leistung der Schüler herausgestellt:

### Auf Schülerebene

- Zugang zu Büchern
- Alphabetisierung der Eltern
- Zugang zu Medien
- Anzahl wiederholter Klassen

### Auf Klassenebene

- Ausbildungsgrad des Lehrers
- Art des Anstellungsverhältnis Lehrer (staatlich, freiwillig etc.)
- Motivation Lehrer
- Fortbildungen Lehrer
- Fehltage Lehrer
- Klassengröße
- Kenntnis der lokalen Sprache Lehrer
- Peer Effect: Wieviele Bücher insgesamt in Klasse (bzw. pro Kopf)

### Auf Schulebene

- Ausstattung mit Bibliothek
- weitere Ausstattung

- Region (ländlich/ urban)
- Inspektion durch Behörde

Sowohl die PASEC-Studie, als auch diese Studien geben Hinweise darauf, welche Faktoren bei weiteren Modellen in jedem Fall zu berücksichtigen sind. Die folgende und letzte Studie, die ich hier vorstelle, schlägt bereits einen ähnlichen Weg ein, wie ich ihn im Analyseteil begehe.

## **Aktuellste Senegalstudie [Dumas and Lambert, 2005]**

Diese wahrscheinlich neueste Studie basiert sowohl auf den PASEC-Daten, als auch auf Daten einer Haushaltsbefragung, die 2003 im Senegal durchgeführt wurde. An letzterer Befragung haben auch viele Eltern der PASEC Kinder teilgenommen, so dass daraus ein einziger Datensatz erzeugt werden konnte. So konnten die Autoren den Einfluss weiterer Variablen auf die Schulleistung untersuchen. Dies waren z.B. die Bildung und der Wohlstand der Eltern. Es wurde als abhängige, zu erklärende Variable jedoch nicht die Schulleistung, gemessen anhand der Ergebnisse in den PASEC-Tests, herangezogen, sondern der erreichte Schulabschluss. Die Frage, auf welchem Wege Familienhintergrundvariablen, wie die oben genannten, die Höhe des Schulabschlusses beeinflussen, lässt sich auf 2 verschiedenen Weisen beantworten [Dumas and Lambert, 2005]:

- Zum einen über den Weg der besseren Lerneffektivität, d.h. über den Weg besserer Schulleistungen, die z.B. durch bessere Förderung durch gebildete Eltern, erreicht werden. Dieser Vorteil müsste sich auch in besseren Resultaten in den PASEC-Tests niederschlagen.
- Zum anderen über den Weg höherer Präferenzen zugunsten von Bildung, die sich in erhöhten Anforderungen an das Kind niederschlagen, was die Dauer des Schulbesuchs betrifft. Diese führt dann letztlich zu einem höheren Abschluss als bei einem Kind mit gleich guter Leistung aber mit Eltern, die z.B. einen früheren Einstieg in das Berufsleben einem höheren Bildungsabschluss vorziehen.

Der erste Weg ist sicherlich wichtiger, wenn man die Leistung der Schule oder des gesamten Bildungssystems beurteilen will. Ein gutes Schulsystem sollte die ungleichen Chancen versuchen auszugleichen und alle Kinder nach ihren Fähigkeiten fördern. Zwei Kinder mit gleichen kognitiven Fähigkeiten sollten in der Lage sein, gleichen Erfolg in der Schule zu haben. Leisten Variablen wie Bildung der Eltern einen signifikant positiven Erklärungsbeitrag zu der Leistung des Kindes unter Kontrolle des kognitiven Niveaus, so bedeutet dies ein ungerechtes Schulsystem, was die Chancengleichheit angeht.

Um also den ersten Weg zu untersuchen, sollte anstelle des erreichten Abschlusses die Leistung anhand der Testergebnisse gemessen werden, was die Autoren zwar vorschlagen, aber nicht ausführen.

Stattdessen begehen Sie einen anderen Weg, um die Präferenzen der Eltern aus dem Spiel zu lassen. Sie untersuchen den Effekt verschiedener Familienhintergrundvariablen

(insbesondere Bildung der Eltern und Wohlstand) auf die Anzahl wiederholter Jahre in der Grundschule. Sie unterstellen, plausibler Weise, gleiche (geringe) Präferenzen der Eltern gegenüber dem Sitzenbleiben ihres Kindes. Sie meinen, dass die Variablen nur noch über den ersten Weg, also über eine höhere Effektivität des Lernens, auf die Anzahl der wiederholten Jahre wirken kann. Kontrolliert wird dabei das kognitive Niveau der Schüler zu Beginn der Schullaufbahn, durch Aufnahme der Testergebnissen zu Beginn der Grundschule als erklärende Variable.

Die Autoren finden dann, dass lediglich Wohlstand einen signifikanten negativen Einfluss auf die Anzahl der wiederholten Jahre hat und glauben, so gezeigt zu haben, dass Wohlstand sich positiv auf die Lerneffektivität auswirkt.

Dies ist sicherlich falsch, da unter anderen die oben dargestellte Frankreichstudie gezeigt hat, dass die Chance, sitzen zu bleiben, bei Kindern gleichen Niveaus im entsprechenden Schuljahr, stark von ihrer sozioökonomischen Stellung abhängt, bzw. der der Eltern, welche wiederum bedeutend durch den Wohlstand der Eltern bestimmt ist. So bleiben Kinder wohlhabender und gebildeter Eltern bei gleichem Niveau weniger oft sitzen, als andere. Deshalb ist die Schlussfolgerung der Autoren unzulässig.

Interessant ist jedoch die Idee, dass Ausgangsniveau der Schüler, ihr kognitives Niveau, zu Beginn der Grundschule, als Kontrollvariable in die Modelle mit einzubeziehen. Als Variablen werden hierfür die Testergebnisse der ersten PASEC-Tests, zu Beginn der 2. Klasse, genommen.

Die Idee dabei ist, unbeobachtete Variablen auszuschalten, die alle Testergebnisse über die Zeit in gleicher Weise beeinflussen, also quasi 'fixed effects'. Der Ansatz ist ähnlich dem der Wahl von einem 'fixed effects' Modell, bei dem solche Variablen ebenfalls herausfallen.

## 3. Mängel bisheriger Studien

### 3.1. Omitted Ability

Weshalb es von Interesse ist, Faktoren für schulischen Erfolg und Wiederholwahrscheinlichkeit zu finden, obwohl dies offenbar in einigen Studien schon geschehen ist:

In allen genannten Studien, außer der letzten, wurde nichts darüber gesagt, welchen Einfluss ausgelassene relevante erklärende Variablen haben könnten. Dabei liegt dieses Thema gerade in der Bildungsforschung nahe. So ist eindeutig erwiesen, dass kognitive Fähigkeiten (Ability) einen starken Einfluss auf Lernerfolg und Wiederholwahrscheinlichkeit haben. Sogar in einigen Ökonometrie Lehrbüchern wird als Beispiel für 'Omitted Variable Bias' Ability angeführt, also angeborene, oder in den ersten Lebensjahren erworbene, Fähigkeiten, sich Wissen anzueignen und zu verstehen [Arellano, 2003].

Die Vermutung liegt daher nahe, dass alle Ergebnisse der Studien, die mit Querschnittsdaten erhoben wurden, verzerrt sein könnten.

Es gibt zwei verschiedene Auffassungen über die Natur des Zusammenhangs von Ability und Testergebnissen, die Hansen et al. darstellen [Hansen et al., 2004]. Die eine Ansicht ist die, dass kognitive Ability sich früh festigt und danach nicht mehr verändert. Mit IQ Tests kann die Ability gemessen werden, unabhängig von Länge und Qualität der Schulbildung. Eine andere Auffassung meint, dass IQ-Testergebnisse deutlich von Schulbesuch beeinflusst werden. Die Kontroverse verläuft aber im Wesentlichen über die Güte von IQ Tests und ist hier nicht interessant, da wir es ja mit Tests zu tun haben, die den Lernfortschritt untersuchen und eben nicht die kognitiven Fähigkeiten.

Auch die zweite Auffassung, die von den Autoren vertreten wird, geht nämlich von einer latenten, früh gefestigten Ability aus, die sie von der manifesten Ability unterscheiden zu der sich die durch Bildung erworbene gesellt. Folgen wir dieser Argumentation, so gilt es in den Analysen die latente Ability auszuschalten.<sup>1</sup>

### 3.2. Abhilfe

Mein Ziel ist es unverzerrte Schätzer für die Lernerfolgskfaktoren und Faktoren für die Wiederholwahrscheinlichkeit zu bekommen. Dazu verwende ich 'fixed-effects' Modelle. Dies bedeutet zugleich, dass ich mich auf die über die Zeit variierenden Faktoren konzentriere, da zeitinvariante Faktoren, wie Region der Schule oder Familiencharakteristika, durch das Verfahren eliminiert werden. Da die Lehrer der Schüler jedoch häufig wechseln, sind die Eigenschaften der Lehrer variabel.

---

<sup>1</sup>Im folgenden wird der Einfachheit halber immer von Ability gesprochen, wo exakt genommen latente Ability gemeint ist.

Im folgenden wird für die 'Produktion' von Lernfortschritt von einem 'konventionellen value-added Modell' ausgegangen, welches einem Artikel von Rivkin entnommen ist [Rivkin et al., 2005]:

Das Modell beschreibt den Lernzuwachs ( $\Delta A_{ijs}$ )<sup>2</sup> eines Schülers  $i$  mit Lehrer  $j$  in Schule  $s$  im Jahr  $t$ :

$$\Delta A_{ijst} = A_{ijs,t} - A_{ijs,t-1} \quad (3.1)$$

$$\Delta A_{ijst} = X_{it}\beta_X + T_{jst}\beta_T + S_{st}\beta_S + f_s + \varepsilon_{ijst} \quad (3.2)$$

Der Lernzuwachs hängt vom Familienhintergrund ( $X$ ), Lehrereigenschaften ( $T$ ), Schuleigenschaften ( $S$ ), Schüler Ability ( $f$ ) und einem Zufallsfehler ( $\varepsilon$ ) ab. Hierbei ist zu beachten, dass die Schüler Ability betrachtet wird als:

Set of cognitive skills, motivation, and personal traits that affect the rate of achievement growth but that do not change during the school years considered. [Rivkin et al., 2005]

Wie erwähnt, variieren lediglich die Lehrereigenschaften über die Zeit, so dass bei einem einfachen fixed effects Modell nur der Einfluss der Lehrereigenschaften zu schätzen ist, auf die ich mich auch konzentrieren möchte.

---

<sup>2</sup>A steht für Achievement nicht für Ability

Teil II.

## Daten und Methoden

## 4. Noch einmal zur Datenbasis

### 4.1. Datensatz

Da die wichtigsten Variablen aus den Lehrerfragebögen kommen, müssen die Jahre 99/00 und 98/99 herausgenommen werden, da es im ersten gar keinen Lehrerfragebogen und 98/99 nur einen sehr kleinen gab. Die meisten, nicht so relevanten, Variablen aus den Schülerfragebogen sind zeitinvariant und fliegen bei den fixed-effects Modellen raus. Deshalb ist es nicht weiter schlimm, dass es 96/97 und 98/99 keinen Schülerfragebögen gab. Dasselbe gilt für die Direktorenfragebögen.

Für die Analyse wird nicht der von PASEC zusammengesetzte Paneldatensatz verwendet. Stattdessen wurde der Datensatz am HWWA neu aus den Rohdaten zusammengesetzt und mit neuen standardisierten Fähigkeitsvariablen versehen (s.u.). Der Datensatz wurde erst vor kurzem fertig und noch für keine Analyse benutzt, so dass nicht auszuschließen ist, dass Fehler im Datensatz sind.

Für die Analysen stehen folgende Fallzahlen zur Verfügung, die leicht von den Fallzahlen der PASEC-Studie(Abb.2.1) abweichen:

Tabelle 4.1.: Fallzahlen pro Welle

Welle	Fallzahl
95/96	1963
96/97	1450
97/98	1144

Auf Grund von fehlenden Werten fallen bei den unterschiedlichen Untersuchungen jeweils noch einige Fälle weg.

### 4.2. Messung der Fähigkeit der Schüler

Anders als Variablen wie z.B. Einkommen lassen sich die Leistungen der Schüler nicht direkt von Jahr zu Jahr vergleichen, da das fortschreitende Niveau der Schüler unterschiedlich schwere Tests erfordert. Da in dem Datensatz jedoch so genannte 'Anchor-Items', also Items die in zwei aufeinander folgenden Jahren in den Tests auftauchen, vorkommen, ist es möglich mit Hilfe der Item Response Theorie eine solche vergleichbare Variable im Nachhinein zu konstruieren.

Die Item Response Theorie umfasst Verfahren, die sich, von Psychometrikern entwickelt, in der Bildungsforschung durchgesetzt haben. Alle großen Studien wie PISA oder

TIMSS stellen, mit diesen Verfahren erzeugte, Schülerleistungsfähigkeits-Variablen zur Verfügung.

Da diese Theorie verschiedene spezielle Verfahren beinhaltet, die dicke Lehrbücher füllen, und die Anwendung dieser Verfahren nicht Teil dieser Arbeit ist, werden lediglich die Grundzüge, nach Hambleton, kurz dargestellt [Hambleton et al., 1991].

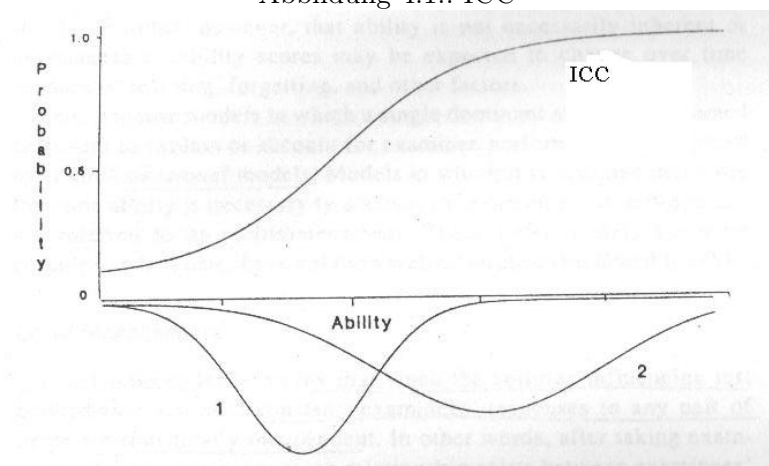
Um die Testergebnisse verschiedener Tests, oder unterschiedlicher Testteilnehmergruppen, vergleichen zu können, müssen zwei Bedingungen erfüllt werden:

- Die Leistungsfähigkeit der Teilnehmer muss unabhängig von der Testschwierigkeit geschätzt werden
- und die Schwierigkeit der Test-Items (Testfragen) muss unabhängig von der Teilnehmerschaft geschätzt werden.

Das sind die wesentlichen Unterschiede zur klassischen Testtheorie und Kern der Item Response Theorie.

Hierzu wird die Verbindung der Schülerleistung mit der Schwierigkeit der Items über die so genannte 'Item Characteristic Curve (ICC)' beschrieben (Abb.4.1). Diese Kurve gibt die Wahrscheinlichkeit einer korrekten Antwort in Abhängigkeit der Fähigkeit an. In der Regel wird eine logistische Verteilungsfunktion unterstellt, so dass die Schätzung der ICC bei bekannter Fähigkeit der Schüler einfach mit einer logistischen Regression zu bewältigen wäre. Wenn sich das jeweilige IRT Modell den Daten gut anpasst, wird unabhängig von der Verteilung der Fähigkeiten in der jeweiligen Testgruppe dieselbe ICC geschätzt. Dies ist die Eigenschaft der Invarianz, die auch für die andere Richtung gilt. Bei bekannter Item-Schwierigkeit (ICC), kann die Fähigkeit eines Schülers egal mit welchem Set von Items geschätzt werden.

Abbildung 4.1.: ICC



Quelle: [Hambleton et al., 1991]

In der Regel sind jedoch weder die Fähigkeiten bekannt, noch die Item-Schwierigkeiten (bzw. ICCs). In diesem Fall muss ein iteratives 'Joint Maximum Likelihood Estimation'

Verfahren angewandt werden, um beides gleichzeitig zu schätzen [Hambleton et al., 1991]. Um für verschiedene Tests einen gemeinsamen Maßstab zu bekommen werden die 'Anchor-Items' benötigt.

Für etliche Spezialfälle gibt es eigene Modelle, so wie z.B. für den Fall, in dem davon auszugehen ist, dass ein Teil der Items durch Raten richtig beantwortet werden konnte.

Im weiteren Verlauf werden die Fähigkeitsschätzer aus diesem Verfahren als gegeben angenommen. Eine Plausibilitätsanalyse der Werte steht zudem am Beginn der Analysen.

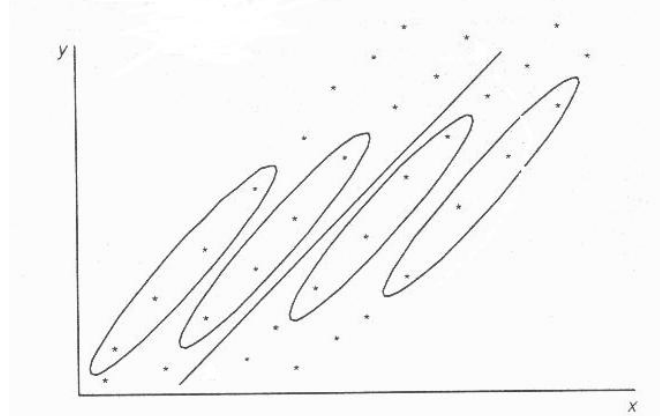
Da sich die Fähigkeit eines Schülers von Schuljahr zu Schuljahr steigert, sollten dies auch die Werte unserer neuen Variablen (je eine für Französisch- und Mathe-Fähigkeit). Somit können, wie in Modell 3.2, Lernfortschritt-Variablen als Differenz der Fähigkeiten zweier Jahre gebildet werden. Diese Variablen werden als abhängige Variablen in den Modellen im Analyseteil verwendet.

## 5. Methoden

### 5.1. Paneldaten-Modelle für Lernfortschritt

Im Vergleich zu Querschnittsdaten erlauben Paneldaten einen besseren Umgang mit Heterogenität in den Einheiten (hier Schülern), die durch eine Vielzahl nicht messbarer erklärender Variablen (Omitted Variables) verursacht wird [Kennedy, 2003]. Entscheidend für die Wahl eines geeigneten Schätzverfahrens ist die Art der Heterogenität. Sind die 'Omitted Variables' unkorreliert mit den im Model vorhandenen erklärenden Variablen, verursacht die Heterogenität keine Verzerrung der Schätzung. In Abb.5.1 ist dieser Fall dargestellt. Die Punkte innerhalb der Ellipsen gehören jeweils zu einer Einheit (z.B. Schüler), für die jeweils 4 Beobachtungen vorliegen. Wie zu erkennen ist, haben die Einheiten verschiedene Intercepts, die die Heterogenität widerspiegeln. Die Intercepts sind jedoch nicht korreliert mit der erklärenden Variable  $X$ .

Abbildung 5.1.: Individuelle Intercepts ohne Korrelation



Quelle: [Kennedy, 2003]

In diesem Fall führt auch eine einfache gepoolte OLS Schätzung, die die Heterogenität überhaupt nicht modelliert, zu unverzerrten Ergebnissen. Ein effizienteres Verfahren stellt unter der Voraussetzung, dass Heterogenität vorliegt, ein 'random effects' Modell dar, in dem für jede Einheit ein eigener Intercept zugelassen wird. Zudem werden diese Intercepts als aus einer Verteilung gezogen modelliert. Sie werden dann behandelt als ein Teil des Fehlerterms, der somit aus zwei Teilen besteht, dem individuellen Term und dem gemeinsamen.

Nun kann in der Schätzung nicht mehr von einer diagonalen Varianz-Kovarianz Matrix ausgegangen werden. Es gibt jedoch eine einfache Transformation, nach der dann per OLS

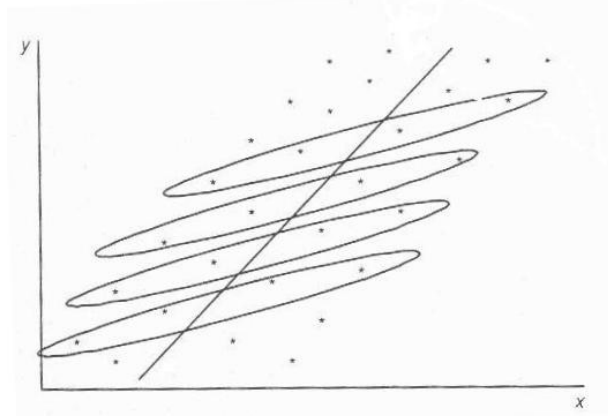
geschätzt werden kann [Kennedy, 2003].

Häufig kann jedoch nicht von der Orthogonalität der 'Omitted Variables' und der weiteren im Modell vorhandenen erklärenden Variablen ausgegangen werden.

Bilden wir z.B. ein Modell, in dem Schulleistung modelliert wird und die erklärende Variable sozio-ökonomischer Status der Eltern nicht vorliegt. Eine im Modell vorkommende Variable Lehrerqualität ist jedoch mit der fehlenden Variable korreliert, z.B. dadurch, dass reiche Eltern (=hoher Status) bessere Schulen für ihr Kind wählen können. In diesem Fall ergibt die Schätzung eines einfachen gepoolten OLS Modells, wie auch die eines 'random effects' Modells verzerrte Schätzer (zum Bsp. des Koeffizienten für Lehrerqualität), vorausgesetzt, der Elternstatus würde einen Erklärungsbeitrag zur Schulleistung leisten.

Dieser Fall ist in Abb. 5.2 dargestellt, wo deutlich die Korrelation der Intercepts mit der erklärenden Variable X zu erkennen ist.

Abbildung 5.2.: Individuelle Intercepts mit Korrelation



Quelle: [Kennedy, 2003]

Hier kann mit Hilfe eines 'fixed effects' Modells unter einigen Annahmen eine unverzerrte Schätzung erreicht werden. So müssen die unbeobachteten Variablen innerhalb der Schüler über die Zeit konstant sein, damit ihr Einfluss eliminiert werden kann.

Wie bereits erwähnt, soll die Panelstruktur der PASEC-Daten ausgenutzt werden, um eine solche unbeobachtete Heterogenität auszuschalten. Im Falle der zu erklärenden Variable Schulleistung haben wir es mit einer metrischen abhängigen Variable zu tun und können ein lineares Modell spezifizieren.<sup>1</sup>

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, 2, \dots, T \quad (5.1)$$

mit  $c_i$ : unobserved heterogeneity

---

<sup>1</sup>Die folgende Panelmethoden-Darstellung folgt im wesentlichen Wooldridge [Wooldridge, 2002]

Wenn  $c_i$  mit den erklärenden Variablen  $x_{it}$  korreliert ist, verursacht die Schätzung des Modells

$$y_{it} = x_{it}\beta + u_{it} \quad (5.2)$$

verzerrte Schätzer  $\beta$ . Man spricht von random effects  $c_i$ , wenn diese nicht mit den erklärenden Variablen korreliert sind.

Meine Vermutung ist allerdings, dass die Schätzer verzerrt sind, da die ausgelassenen unbeobachteten Variablen, wie kognitive Fähigkeiten mit anderen erklärenden Variablen korrelieren, was zu testen ist (s.u.).

Deshalb müssen 'fixed effects' Modelle eingesetzt werden. Diese Modelle eliminieren die  $c_i$ -Effekte. Eine gängige Methode dafür ist die Mittelwertbereinigung durch die within oder fixed effects Transformation. Zunächst werden die Variablen über die Zeit gemittelt:

$$\bar{y}_i = \bar{x}_i\beta + c_i + \bar{u}_i \quad (5.3)$$

Nun wird diese gemittelte Gleichung von der Ursprungsgleichung abgezogen

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i) \quad (5.4)$$

Die  $c_i$ s fliegen heraus. Die neue Gleichung

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i) \quad (5.5)$$

oder

$$\check{y}_{it} = \check{x}_{it}\beta + \check{u}_{it} \quad \text{bzw.} \quad \check{y}_i = \check{X}_i\beta + \check{u}_i \quad (5.6)$$

kann per OLS konsistent geschätzt werden unter der Annahme

$$E[u_{it}|x_{it}, c_i] = 0, \quad t = 1, 2, \dots, T \quad (5.7)$$

Bei den in STATA implementierten fixed effects Schätzern wird bei der within Transformation zusätzlich der Mittelwert der jeweiligen Welle abgezogen, was die Aufnahme von Zeitdummies überflüssig macht.

Möchte man, bevor man sich für ein random effects oder fixed effects Modell entscheidet, testen ob exogene unbeobachtete Heterogenität vorliegt und also ein 'random effects' Modell angebracht ist, kann man einen Hausmann-Test durchführen. Hierzu sind zunächst beide Modelle zu schätzen.

Die Schätzung von 5.6 liefert die gleichen Schätzer wie die Schätzung einer OLS Schätzung mit Dummyvariablen für alle Schüler, d.h. wenn man die Individualeffekte mit-schätzt [Wooldridge, 2002]. Deshalb wird der fixed effects Schätzer bisweilen auch Dummyvariablen Schätzer genannt.

Der Vorteil bei der Mitschätzung der Individualeffekte ist, dass man auf diesem Weg Residuen erhält, die man wie üblich auf Heteroskedastizität und Autokorrelation überprüfen kann. Dies ist mit den Residuen der Schätzung von 5.6 ungleich schwieriger.

## Modellannahmen und Überprüfung

OLS ist für die Schätzung von 5.6 bzw. das Dummyvariablenmodell nur dann effizient, wenn zusätzlich zu 5.7 folgende Annahme erfüllt ist [Wooldridge, 2002]:

$$E[u_i u_i' | x_{it}, c_i] = \sigma_u^2 I_T \quad (5.8)$$

Dies bedeutet, dass die  $u_{it}$  über  $t$  die gleiche Varianz haben und unkorreliert sind. Wie erwähnt, sind diese Annahmen am besten mit den Residuen der Dummy-Variablen Schätzung zu testen.

Hinzu kommt die Überprüfung, ob Multikollinearität vorliegt, die auch im transformierten Modell durchgeführt werden kann. Hierzu sind z.B. die Variance Inflation Faktoren zu berechnen. Des weiteren sollte jedes Modell mit Hilfe von Werten wie Cook's Distance, Dfbetas und Leverages auf Ausreißer hin untersucht werden.

## Vorgehen bei Verletzung der Modellannahmen

Beim Vorliegen von Multikollinearität oder dem Vorliegen von Ausreißern, bei denen Datenfehler wahrscheinlich sind, ist das Vorgehen einfach. Im ersten Fall muss ein neues Set an erklärenden Variablen gefunden werden, welches keine Kollinearität zeigt. Im zweiten Fall kann man die Ausreißer aus dem Modell nehmen, allerdings wirklich nur, wenn von Datenfehlern ausgegangen werden kann.

Liegt eine Verletzung der Annahme 5.8 vor, so sind die Schätzer nicht verzerrt, jedoch werden die Standardfehler falsch angegeben. Nun bieten sich zwei Möglichkeiten an:

1. Man kann einen asymptotisch robusten Varianzmatrix-Schätzer benutzen, der sowohl gegen Autokorrelation, als auch bei Heteroskedastizität robust ist, unter der Voraussetzung, dass  $T$  klein ist im Verhältnis zu  $N$  [Wooldridge, 2002]:

$$A\hat{var}(\hat{\beta}_{FE}) = (\ddot{X}'_i \ddot{X}_i)^{-1} \left( \sum_{i=1}^N \ddot{X}'_i \hat{u}_i \hat{u}_i' \ddot{X}_i \right) (\ddot{X}'_i \ddot{X}_i)^{-1} \quad (5.9)$$

2. Eine weitere Möglichkeit im Falle von Autokorrelation stellt eine fixed effects GLS Schätzung dar. In diesem Verfahren werden die Daten transformiert. Es muss jedoch eine Welle eliminiert werden, was im Falle von nur drei Wellen für mich nicht in Frage kommt. Deshalb soll dieses Schätzverfahren hier nicht weiter ausgeführt werden. Im Falle des alternativen random effects Modells, welches für den Hausman Test spezifiziert werden soll, gibt es ebenfalls die Möglichkeit einer Transformation

der Daten, wenn Heteroskedastizität oder Autokorrelation vorliegt, ohne dass hier eine Welle eliminiert werden muss. Dieses Schätzer nennt sich Baltagi-Wu GLS Schätzer und ist in STATA implementiert [STATA, 2003]. Er wird für die Alternativmodelle verwendet.

## 5.2. Modelle für Wiederholwahrscheinlichkeit

Für das Modell zur Erklärung der Wiederholwahrscheinlichkeit wäre unter den gleichen Überlegungen ein Logit-’fixed effect’ Modell sinnvoll. In der Tat gibt es eine solche Modellierungsmöglichkeit: Es lässt sich zeigen [Wooldridge, 2002], dass die gemeinsame auf

$$x_i, c_i \text{ und } m_i \equiv \sum_{t=1}^T y_{it} \quad (5.10)$$

$m_i$ : Anzahl der Realisationen innerhalb einer Einheit über alle  $t$

bedingte Verteilung der  $y_i$  nicht mehr von  $c_i$  abhängt. Sodann ist eine ML Schätzung möglich. Ein Haken dabei ist, dass für die bedingte Verteilungsfunktion Beobachtungen nicht informativ sind, für die  $m_i = 0$  oder  $m_i = T$  gilt. Die Schätzung basiert also nicht auf allen Fällen. Im Falle dieser Studie fielen jene heraus, die nie sitzen geblieben sind, sowie jene die jedes Jahr sitzen blieben (letzteres betrifft allerdings niemanden). Ein solches Modell würde dann die Einflüsse auf die Wahrscheinlichkeit des Sitzenbleibens nur bei jenen untersuchen, die mindestens einmal sitzen blieben. Die Fragestellung und die Grundgesamtheit wären andere, als beabsichtigt. Deshalb wird ein solches Modell nicht spezifiziert.

Auf Grund der Überlegung, dass Schulleistung der Variable Sitzenbleiben kausal vorgelegt ist, kann jedoch durch Aufnahme dieser erklärenden Variable, der Einfluss unbeobachteter Variablen (wie Ability), die nur über den Weg der Schulleistung die Sitzenbleib-Wahrscheinlichkeit beeinflussen, ausgeschaltet werden. Diese Argumentation entspricht den Überlegungen der Autoren der aktuellsten Senegalstudie (s.o.) und Überlegungen in der PASEC-Studie. Die Variable Schulleistung fungiert hierbei als so genannte Proxy-Variable.

Somit kann ein random effects Probit Modell spezifiziert werden, mit der abhängigen Variable Sitzenbleiben (ja/nein) und der Variablen Schulleistung als einer der erklärenden Variablen.

Grundannahme eines solchen Modells ist [Wooldridge, 2002]:

$$P(y_{it} = 1|x_i, c_i) = P(y_{it} = 1|x_{it}, c_i) = \Phi(x_{it}\beta + c_i), \quad t = 1, \dots, T \quad (5.11)$$

wobei  $\Phi$  die Dichtefunktion der Standardnormalverteilung darstellt, die wie üblich in Probit Modellen die link-Funktion ist. Die erste Gleichung bedeutet strikte Exogenität von  $x_{it}$  bedingt auf  $c_i$ .

Eine weitere Annahme lautet [Wooldridge, 2002]:

$y_{i1}, \dots, y_{iT}$  sind unabhängig, bedingt auf  $(x_i, c_i)$

Auch in diesem Fall muss für das random effects Modell angenommen werden, dass  $c_i$  und  $x_i$  unkorreliert sind. Zudem muss eine Verteilungsannahme über die  $c_i$ s gemacht werden:

$$c_i|x_i \sim Normal(0, \sigma_c^2) \quad (5.12)$$

Das  $\sigma_c^2$  muss mit den  $\beta$ s geschätzt werden. Zur Schätzung steht ein Conditional Maximum Likelihood Ansatz zur Verfügung, der u.a. in STATA implementiert ist und bei dem die  $c_i$ s heraus integriert werden. Die Schätzer sind asymptotisch normalverteilt [Wooldridge, 2002].

Zum Vergleich werden ein gepooltes Probit-Modell und ein gepooltes Logit-Modell geschätzt. Da für die Wiederhol-Wahrscheinlichkeits Modelle auch Schüler- und Schulcharakteristika als erklärende Variablen zur Verfügung stehen, die nicht über die Zeit variieren, kann es sein, dass ein random effects Modell keine bessere Anpassung an die Daten liefert, als die gepoolten Modelle.

## Modelldiagnose

Zur Modelldiagnose stehen zum einen Pseudo- $R^2$ -Werte zur Verfügung. Zum anderen kann die Vorhersagefähigkeit des Modells untersucht werden anhand richtig klassifizierter Fälle, wenn für Fälle mit vorhergesagter Realisationswahrscheinlichkeit größer als 0,5 eine Realisation vorhergesagt wird und für die übrigen keine Realisation.

Sollen Modelle mit einander verglichen werden, kann dies anhand der BIC-Werte geschehen. Genestete Modelle können auch gegeneinander getestet werden.

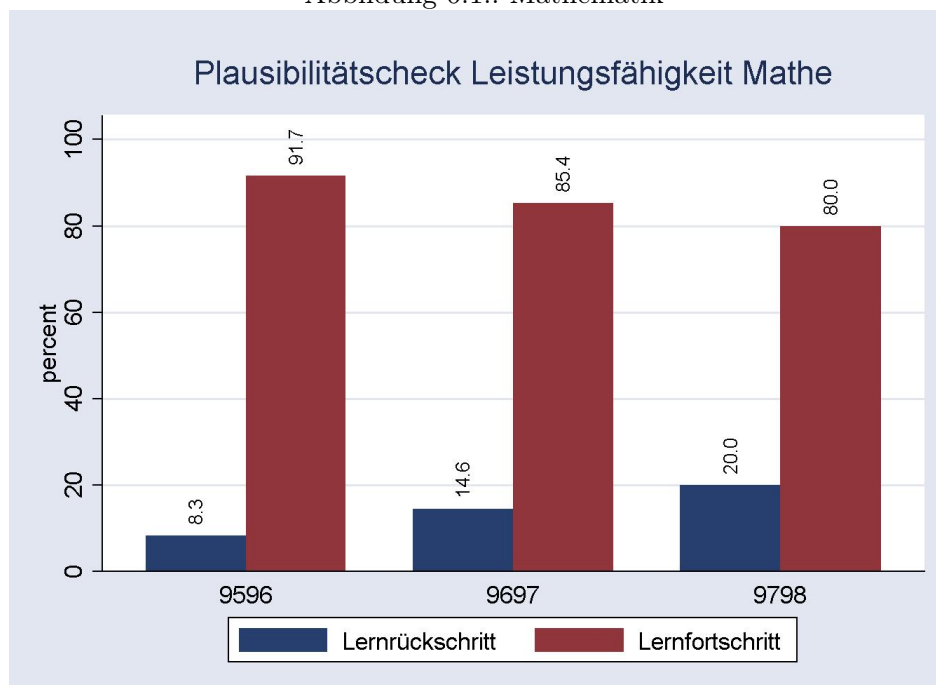
Zur Identifizierung möglicher Ausreißer stehen einige Diagnoseplots zur Verfügung, die auf Hosmer und Lemeshow zurückgehen [Hosmer and Lemeshow, 1989]. Da die selben Variablen zur Verfügung stehen, wie bei den Analysen zum Lernfortschritt erfolgt keine erneute Überprüfung von Multikollinearität. Es werden zudem von Beginn an robuste Standardfehler-Schätzer verwendet.

Teil III.  
Analyse

## 6. Datenüberprüfung

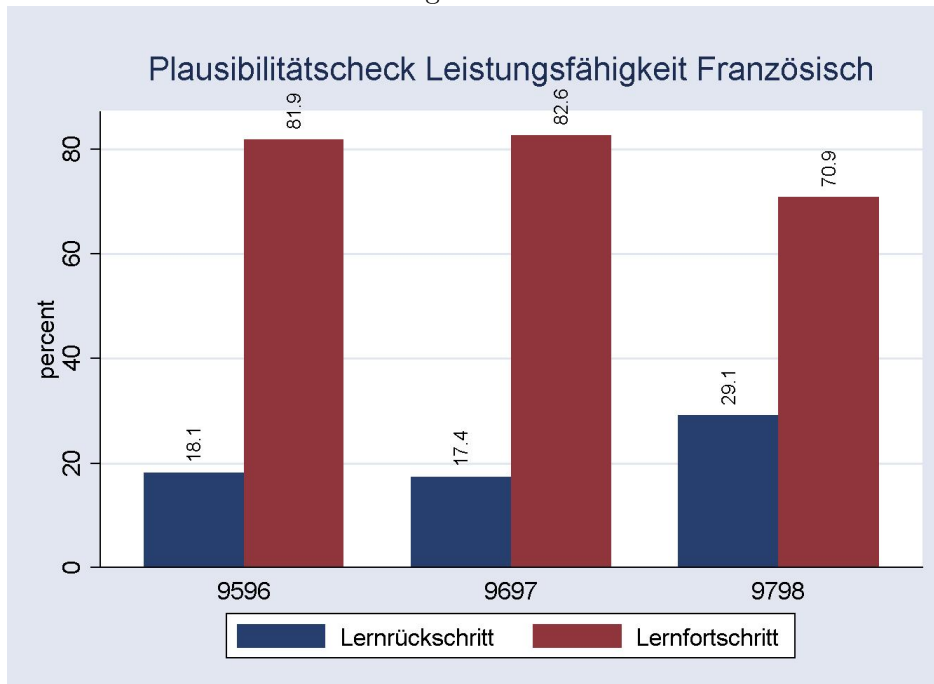
Auf Grund der Unsicherheit über die Validität der Daten, insbesondere der Leistungsvariablen, die daraus resultiert, dass die Daten in dieser neuen Zusammensetzung aus den Rohdaten und den ermittelten Leistungsvariablen hier zum ersten Mal benutzt werden, erfolgt zunächst eine Überprüfung der Leistungsvariablen. Wie im Abschnitt zur Item Response Theorie dargestellt, sollte die Leistung der Schüler von Jahr zu Jahr wachsen und höchstens in Ausnahmefällen zurück gehen. Hierzu werfen wir also zunächst einen Blick auf die Variablen Lernfortschritt in Mathematik und Französisch (Abb.6.1 und 6.2).

Abbildung 6.1.: Mathematik



Offensichtlich gibt es in beiden Fällen und in allen Jahren einen erheblichen Teil von Schülern mit negativem Lernfortschritt. Zu vermuten ist, dass die Kalibrierung der verschiedenen Testschwierigkeiten mittels der Anchor-Items nicht richtig funktioniert hat. Da es bei den Mathe-Tests offensichtlich besser funktioniert hat, als bei den Französisch-Tests, verwende ich für die ersten Modelle die Variable Lernfortschritt in Mathematik als abhängige Variable. Sollte das Problem tatsächlich nur darin bestehen, dass die Testniveaus nicht richtig eingeschätzt wurden, so behalten die Differenzen von zwei Leistungsangaben möglicherweise trotzdem ihre Bedeutung, sie wären nur etwas verschoben.

Abbildung 6.2.: Französisch



So hätte ein Schüler mit Null Veränderung einen besseren Fortschritt erzielt, als einer mit negativer Veränderung. trotzdem bleibt eine Unsicherheit über die Gültigkeit der Variablenwerte, da ja nicht nur 2 Tests mit einander verglichen werden, sondern ohne Wiederholer 3 und mit Wiederholern sogar 5.

Deshalb verende ich in einem zweiten Schritt eine andere abhängige Variable. Es werden hier nur Nicht-Wiederholer berücksichtigt und die Variable Leistung (arithmetisches Mittel aus Mathe-Leistung und Französisch-Leistung) für jedes Jahr standardisiert. Um zu entscheiden, ob robuste Mittelwert- und Varianzschätzer für die Standardisierung zu verwenden sind sehen wir uns zunächst die Verteilung der Leistung für die drei Jahre an, sowie einige deskriptive Maßzahlen:

Tabelle 6.1.: Vergleich von robusten Schätzern und standardmäßigen

Jahr	arith. Mittel	Median	Stdabw.	IQR/1,35
9596 (Anfang)	-.0661285	-.067129	.7571996	.87011629
9596	.6325821	.7427995	.8456547	.86884922
9697	1.239425	1.405596	.8649254	.85580888
9798	1.305524	1.34873	.9484935	1.1658408

Wie anhand der Histogramme (Abb.6.4 bis 6.6) unschwer zu erkennen ist, ist die Leistungsvariable zu keiner Zeit auch nur annähernd normalverteilt. Wie in Tabelle 6.1 dargestellt unterscheiden sich auch die robusten Schätzer, Median für Mittelwert und

Abbildung 6.3.: Verteilung Leistungsvariable Anfang 95/96

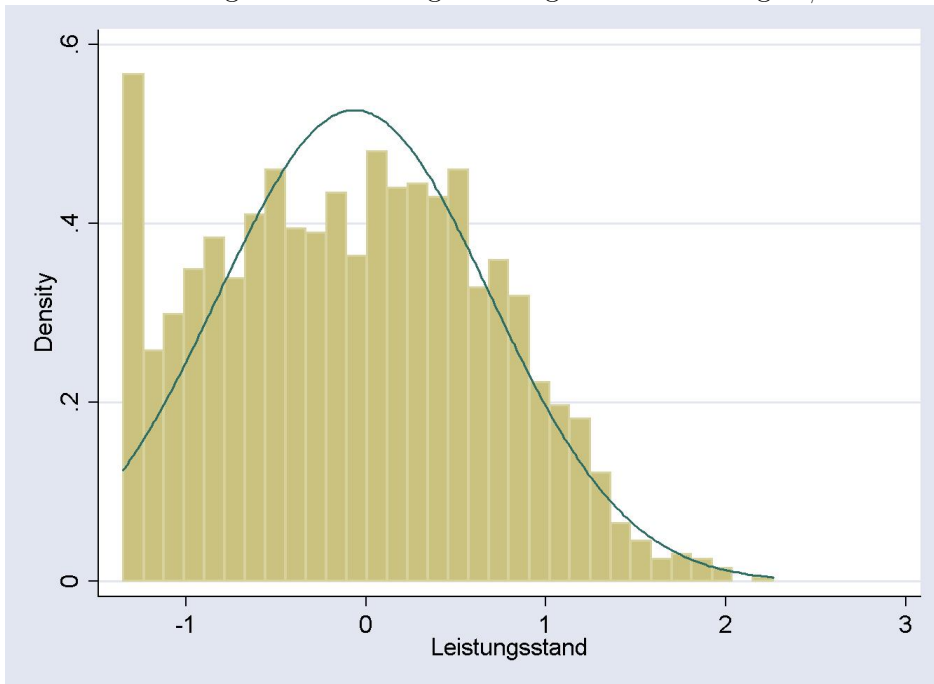


Abbildung 6.4.: Verteilung Leistungsvariable Ende 95/96

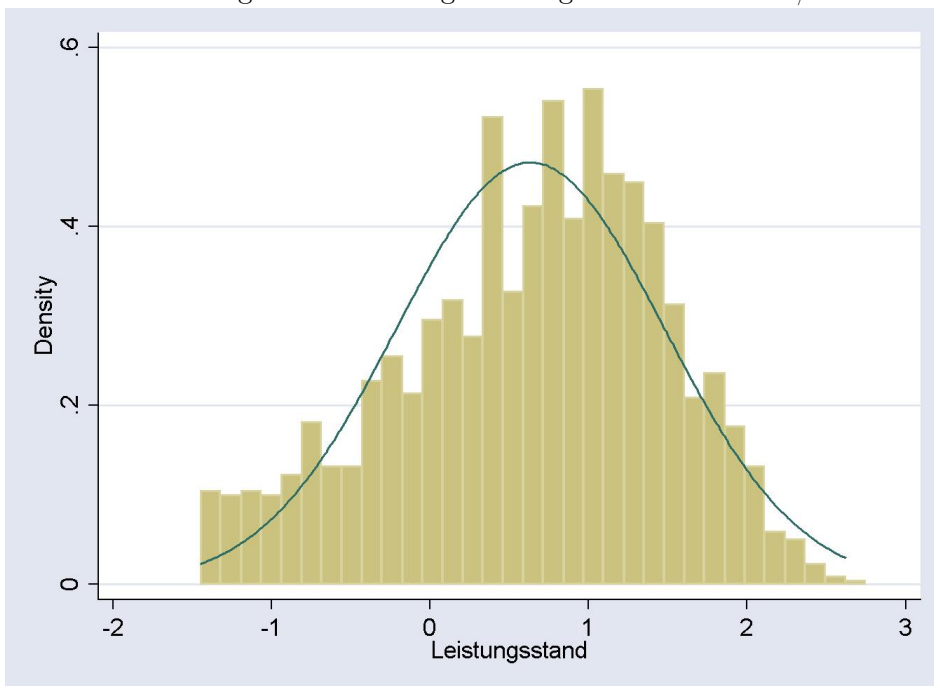


Abbildung 6.5.: Verteilung Leistungsvariable Ende 96/97

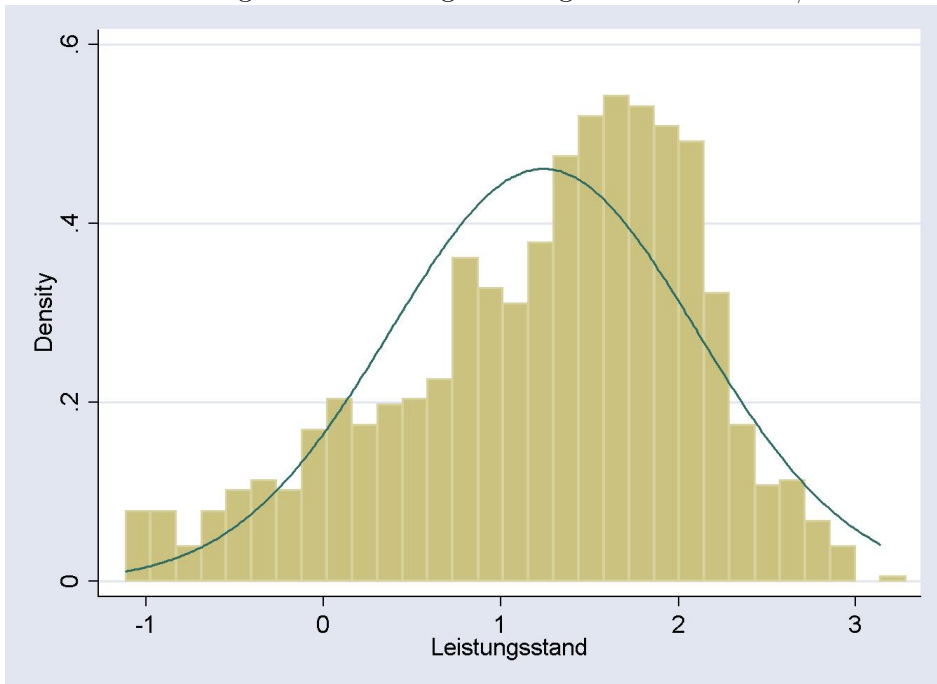
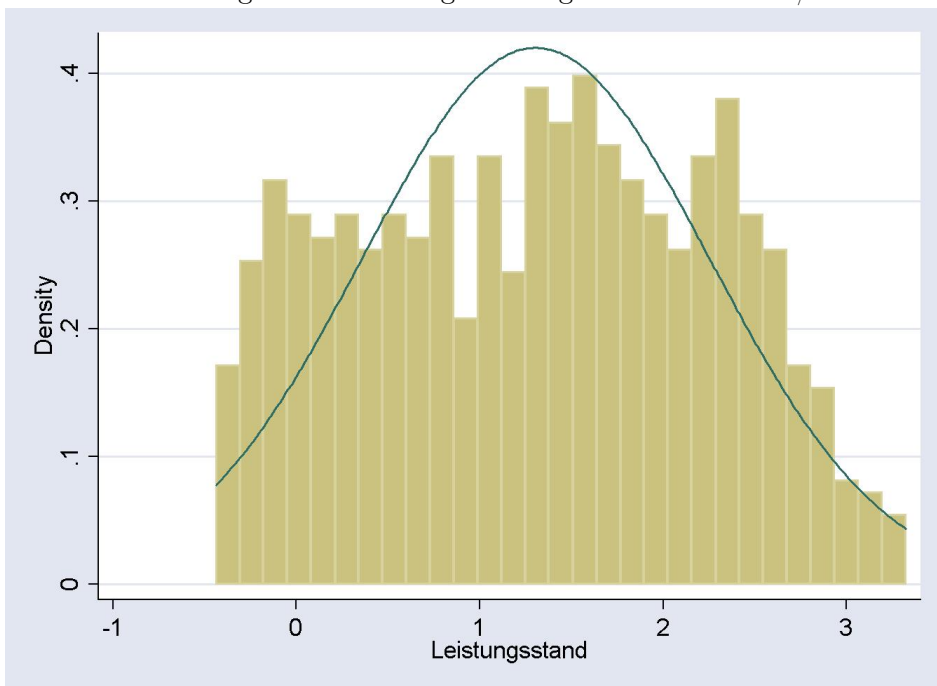


Abbildung 6.6.: Verteilung Leistungsvariable Ende 97/98

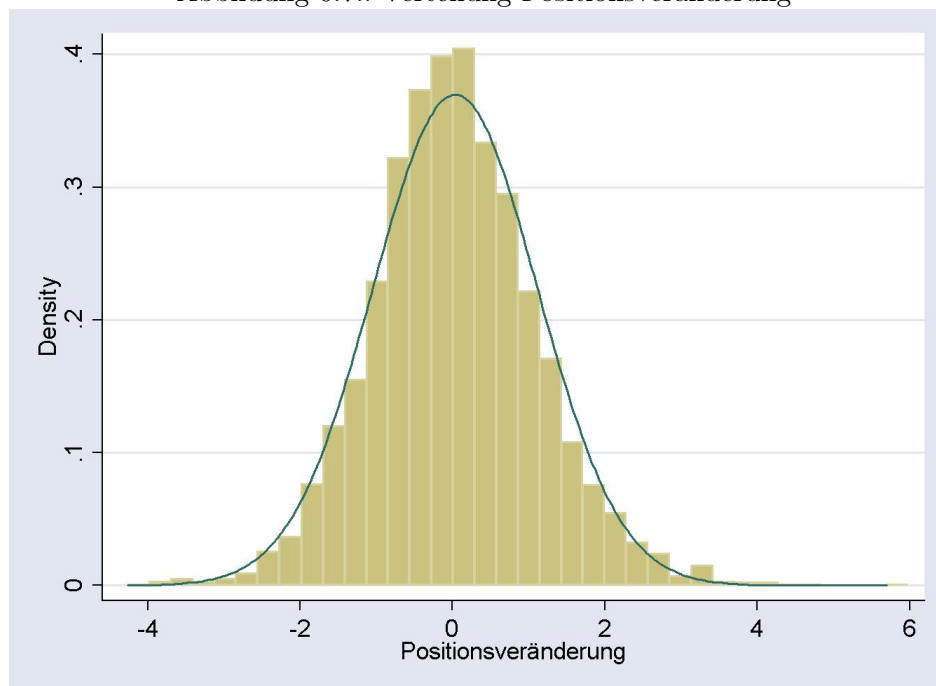


Interquartilsabstand geteilt durch 1,35 (ergibt im Falle der Normalverteilung den selben Wert wie die normale Standardabweichungsschätzung) z.T. deutlich von den optimalen Schätzern bei einer Normalverteilung.

Die Variablen werden deshalb mit den robusten Schätzern standardisiert. Hierfür wurden nur die Schüler berücksichtigt, die in allen 3 Wellen am Panel teilgenommen haben. Es wurde von den Leistungsvariablen also jeweils der Median abgezogen und durch  $IQR/1,35$  geteilt. Darauf wurde die Variable Positionsveränderung gebildet als Differenz von jeweils zwei benachbarten Zeitpunkten.

Diese neue Variable drückt demnach nicht mehr Lernfortschritt aus, sondern Positionsveränderung des Schülers gegenüber den anderen Schülern. Sie wird als abhängige Variable in der zweiten Analyse benutzt. Die Reduzierung auf die Schüler, die zu allen 3 Zeitpunkten am Panel teilgenommen haben, erfolgt, um eine Veränderung der Position, die lediglich durch das Ausscheiden schwacher Schüler, die abbrechen, verursacht wird, zu vermeiden. Zudem wurden Wiederholer herausgenommen, da diese andere Tests gemacht haben.

Abbildung 6.7.: Verteilung Positionsveränderung



# 7. Analysen Lernfortschritt

## 7.1. Lernfortschritt Mathematik

Ausgangspunkt bildet ein fixed effects Modell, zu dem ein random effects Vergleichsmodell spezifiziert wird.

### Variablenauswahl

Zunächst wurde der Datensatz mit der within Transformation transformiert (5.6). Danach wurde mit einer backward stepwise Regressionsprozedur aus einer Reihe von möglichen erklärenden Variablen eine Auswahl getroffen.<sup>1</sup> Bei der Prozedur werden zunächst alle Variablen aufgenommen und dann anhand der Signifikanz ihrer t-Werte herausgenommen oder im Modell gelassen, wobei nach der Herausnahme einer Variable immer geprüft wird, ob eine bereits aus dem Modell genommene Variable wieder aufgenommen werden kann. Es ist somit sowohl ein Rausschmiss-Wert als auch ein Wiederaufnahme-Wert anzugeben. Als ersten Wert habe ich eine Signifikanz von 0,2 gewählt, beim zweiten eine von 0,1. Leichte Veränderung der Werte, auf 0,15 und 0,05, ergaben die selbe Variablenauswahl.

### Annahmenüberprüfung

Im nächsten Schritt wurden die Annahmen gemäß 5.8 überprüft. Hierzu mussten alle Individualeffekte in den nicht-transformierten Daten mitgeschätzt werden (Dummy-Modell). Da ich keinen Zugang zu einer erweiterten STATA Version, sondern nur der Standardversion hatte, die nur beschränkte Matrizengrößen zulässt, musste ich eine Zufallsstichprobe von 700 Schülern aus den 1963 ziehen.

Mit dieser Stichprobe wurde dann das Dummy-Regressionsmodell geschätzt mit allen im ersten Schritt ausgewählten Variablen. Die Residuen wurden auf Heteroskedastizität und Autokorrelation über t kontrolliert.

Wie schon an den Boxplots in Abb. 7.1 zu erkennen ist, ist die Varianz der Residuen nicht über alle Jahre hinweg konstant und auch eine negative Autokorrelationsstruktur ist zu erkennen. Auf Heteroskedastizität wurden zwei Tests angewendet, die in einem Fall die Nullhypothese Homoskedastizität gegen in Jahr monoton steigende Heteroskedastizität (Szoeter-Test) und im zweiten Fall gegen multiplikative Heteroskedastizität (Breusch-Pagan Test) prüft. Bei beiden Tests muss die Nullhypothese schon auf dem 1%-Niveau abgelehnt werden.

---

<sup>1</sup>Variablenauflistung und -beschreibung s. Anhang

Abbildung 7.1.: Residuen Boxplots über Jahr

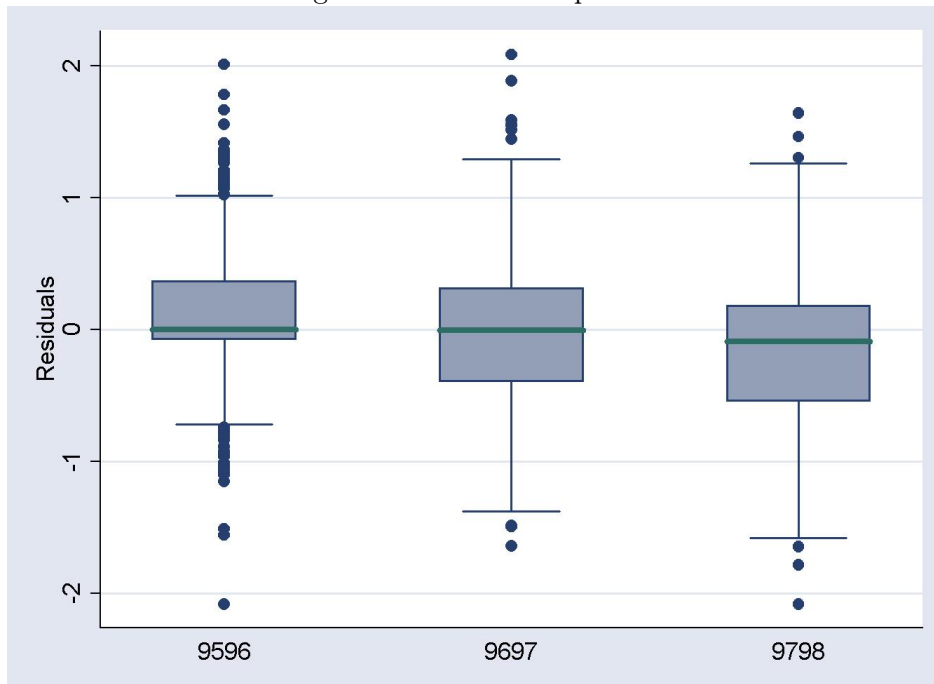


Tabelle 7.1.: Tests auf Heteroskedastizität

Test	Teststatistik $\chi^2(1)$	Signifikanz
Szroeter	11.54	0.0007
Breusch Pagan	11.59	0.0007

Im nächsten Schritt soll auf Autokorrelation getestet werden, wobei ein Durbin-Watson Test für Paneldaten hier nicht zur Verfügung steht. Stattdessen betrachte ich schlicht die Korrelationen zwischen den Residuen der benachbarten Jahre und deren Signifikanz:

Tabelle 7.2.: Korrelation der Residuen (Signifikanz)

	res9596	res9697	res9798
res9596	1.0000		
res9697	-0.6602 (0.0000)	1.0000	
res9798	-0.5600 (0.0000)	-0.5872 (0.0000)	1.0000

Diese Überprüfungen wurden mehrfach mit verschiedenen Zufallsstichproben durchgeführt, mit immer den gleichen Schlüssen. Daher komme ich zu dem Urteil, dass sowohl Heteroskedastizität, als auch Autokorrelation über t vorliegt.

## Modellschätzung und Vergleich mit Alternativmodell

Auf Grund des im Methodenteil dargelegten Nachteils eines fixed effects GLS Schätzers, nämlich dem Verlust einer kompletten Welle, habe ich das within-transformierte Modell stattdessen mit robusten Standardfehlern geschätzt.

Hierzu wurde zunächst erneut eine stepwise Regression durchgeführt, die zu der selben Variablenauswahl führte, wie die zu Beginn. Dann wurde folgendes Modell geschätzt:

Tabelle 7.3.: Schätzergebnisse fixed effects mit robusten Standardfehlern

Variable	Coefficient	(Std. Err.)
JOBEXP	-0.102**	(0.033)
CLASSIZE	0.120**	(0.034)
SEMFREQ	0.051 <sup>†</sup>	(0.029)
EXCHANGE	0.208**	(0.063)
LOCLANG	-0.284**	(0.068)
FOFTEN	0.119*	(0.053)
BACplus	-0.370**	(0.057)
TRAINDUR_1	-0.167**	(0.060)
ACTIV_T	-0.322**	(0.049)
TUITION	0.377**	(0.085)
GUIDEM	-0.099	(0.062)
UNION	-0.163**	(0.055)
REPEAT	0.216**	(0.055)
Intercept	0.487**	(0.096)
N		3373
R <sup>2</sup>		0.07
F (13,3359)		17.6

Significance levels : † : 10% \* : 5% \*\* : 1%

Als Vergleichsmodell wurde nun ein random effects Modell geschätzt, wobei das Baltagi-Wu GLS Verfahren angewendet wurde. Als mögliche erklärende Variablen wurden hier noch eine Reihe von Schüler- und Schuleigenschaften berücksichtigt, die nicht über die Zeit variieren. Die Variablenauswahl erfolgte per Hand, da hier keine stepwise Prozedur zur Verfügung stand. Es wurde sozusagen eine backward selection durchgeführt, bei der neue erklärende Variablen, die nicht in dem fixed effects Modell auftauchen mit einer Signifikanz größer 0,2 herausgenommen wurden. Alle auch im fixed effects Modell vertretenen erklärenden Variablen wurden im Modell gelassen.

Nach der Schätzung des Modells (Tab. 7.5) wurde mit den gespeicherten Schätzergebnissen beider Modelle der Hausman-Test durchgeführt. Hierbei ist das random effects Modell das effizientere, wenn Exogenität der unbeobachteten Variablen bzw. Unkorreliertheit mit den anderen erklärenden Variablen vorliegt. Das fixed effects Modell ist das konsistente Schätzverfahren bei Korrelation mit den anderen erklärenden Variablen. Der Hausman Test kann als Test auf nicht systematische Differenzen in den Koeffizienten der beiden Modelle interpretiert werden. Die Nullhypothese, dass keine systematischen

Differenzen vorliegen musste auf dem 1% Niveau verworfen werden, so dass das fixed effects Modell vorzuziehen ist.

Tabelle 7.4.: Hausman Test

Teststatistik $\chi^2(13)$	Signifikanz
416.29	0.0000

Das geschätzte fixed effects Modell ist also jenes, welches ich auswähle.

Tabelle 7.5.: Schätzergebnisse random effects GLS

Variable	Coefficient	(Std. Err.)
JOBEXP	-0.011**	(0.002)
CLASSIZE	-0.001	(0.001)
MALE_T	0.054†	(0.029)
LOCLANG	-0.081**	(0.030)
FOFTEN	0.046†	(0.028)
VOLUNT	-0.232*	(0.094)
GUIDEM	0.082**	(0.029)
BACplus	0.033	(0.028)
UNION	0.007	(0.026)
REPEAT	-0.086*	(0.036)
LIT1P	0.068*	(0.027)
NCLASS_D3	-0.409**	(0.031)
NCLASS_D4	-0.786**	(0.038)
HELP	-0.083†	(0.044)
HOMEWORK	0.078	(0.061)
SEMFREQ	0.003	(0.014)
EXCHANGE	0.018	(0.030)
INSPECT	0.005	(0.033)
TRAINDUR_1	-0.015	(0.028)
ACTIV_T	-0.009	(0.029)
TUITION	-0.007	(0.047)
Intercept	0.952**	(0.086)
N	3399	
$\chi^2_{(22)}$	694.699	

Significance levels : † : 10% \* : 5% \*\* : 1%

## Modelldiagnose

Zunächst soll das Modell auf Multikollinearität hin überprüft werden. Hierzu berechne ich die Variance Inflation Faktoren (Tab. 7.6). Auf Grund der Tatsache, dass die meisten Variablen Dummies sind, sind die VIFs nicht die optimalen Maße, da auf abhängige binäre

Variablen lineare Regressionen angewendet werden. Trotzdem stellen sie einen Hinweis auf Multikollinearität dar. Für einige Variablen, habe ich im nicht transformierten Datensatz probit Modelle geschätzt, die allesamt niedrige Pseudo- $R^2$  Werte ergaben. Die stärkste Abhängigkeit besteht zwischen TUITION und ACTIV\_T, was aber nicht problematisch, ist, da beide Antwortmöglichkeiten einer Frage mit 3 Antwortmöglichkeiten darstellen.

Tabelle 7.6.: Multikollinearitätsüberprüfung

<b>Variable</b>	<b>VIF</b>	<b>1/VIF</b>
ACTIV_T	1.48	0.673535
TUITION	1.45	0.690290
BACplus	1.33	0.752166
JOBEXP	1.28	0.782971
FOFTEN	1.25	0.796819
EXCHANGE	1.23	0.811794
LOCLANG	1.22	0.816783
GUIDEM	1.17	0.854397
CLASSIZE	1.16	0.865287
TRAINDUR_1	1.13	0.881460
SEMFREQ	1.08	0.929674
UNION	1.05	0.951356
REPEAT	1.02	0.985182

Hieraus folgere ich, dass keine starke Multikollinearität vorliegt. Hieraufhin habe ich das Modell auf Ausreißer hin überprüft. Da STATA Werte wie Cook' Distance, Dfbetas und Leverage nicht bei robusten Standardfehlern ausgibt, wurde das Modell noch einmal ohne Robustheit geschätzt und diese Werte berechnet. Da in diesem Fall die Standardfehler der Koeffizienten unterschätzt werden, fallen mehr Beobachtungen auf, so dass Faustformeln wie in Hamilton [Hamilton, 2004], ab wann eine Beobachtung als Ausreißer zählt, nichts nutzen.

Die Beobachtungen mit den betragsmäßig jeweils größten Werten bei Cook's Distance, Leverage und den Dfbetas für CLASSIZE wurden auf ihre Werte in den Modellvariablen hin überprüft. Dabei konnten keine eindeutigen Eingabefehler oder ähnliches festgestellt werden, weshalb keine Beobachtung herausgenommen wurde.

Die Modellanpassung, bzw. die Erklärung der Varianz der abhängigen Variable innerhalb der Beobachtungen beträgt  $R^2=0,07$ , also werden 7% der within-Varianz erklärt. Schätzt man das Dummyvariablenmodell, erreicht man einen  $R^2$  Wert zwischen 0,44 und 0,45, je nach Stichprobe.

Auf Grund der Unsicherheit mit der abhängigen Variable erfolgt nun die Analyse mit der abhängigen Variable Positionsveränderung (s.o.). Eine Interpretation der Ergebnisse und der Modellanpassung erfolgt im Anschluss.

## 7.2. Positionsveränderung

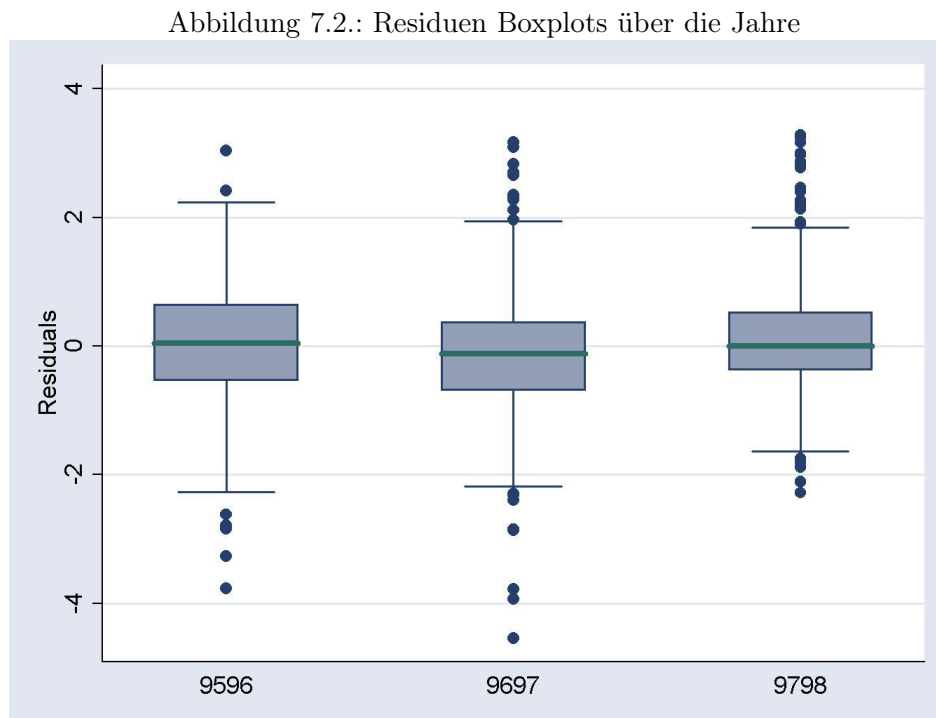
Das Vorgehen ist exakt das selbe wie im vorigen Abschnitt, so dass die Ergebnisse hier etwas knapper präsentiert werden.

### Variablenauswahl

Auch hier wurde aus den selben Variablen wie oben mit einem stepwise Regression Verfahren Variablen ausgewählt. Auch hier kam es zur selben Variablenauswahl, bei leicht veränderten Schwellenwerten.

### Überprüfung der Annahmen

Hierzu wurde das Dummyvariablenmodell bei einer Stichprobe von 700 Schülern geschätzt. Die Residuenverteilung ist in Abb.7.2 dargestellt.



Auch hier ist schon zu erkennen, dass die Varianz der Residuen nicht über alle Jahre hinweg konstant ist.

Bei beiden Tests auf Heteroskedastizität (Tab.7.8) muss die Nullhypothese zumindest auf dem 5%-Niveau abgelehnt werden. Im nächsten Schritt wird wieder auf Autokorrelation getestet, wobei erneut die Korrelationen zwischen den Residuen der benachbarten Jahre und deren Signifikanz betrachtet wird (Tab.7.8).

Tabelle 7.7.: Tests auf Heteroskedastizität

Test	Teststatistik $\chi^2(1)$	Signifikanz
Szroeter	6.57	0.0104
Breusch Pagan	6.54	0.0106

Tabelle 7.8.: Korrelation der Residuen (Signifikanz)

	res9596	res9697	res9798
res9596	1.0000		
res9697	-0.6456 (0.0000)	1.0000	
res9798	-0.5131 (0.0000)	-0.5094 (0.0000)	1.0000

Auch diese Überprüfungen wurden mehrfach mit verschiedenen Zufallsstichproben durchgeführt, mit immer den gleichen Schlüssen. Daher komme ich zu dem Ergebnis, dass auch hier Heteroskedastizität und Autokorrelation über t vorliegt.

## Modellschätzung und Vergleich mit Alternativmodell

Nun wurde das within-transformierte Modell mit robusten Standardfehlern geschätzt. Hierzu wurde zunächst erneut eine stepwise Regression durchgeführt, die zu der selben Variablenauswahl kam, wie die zu Beginn. Dann wurde folgendes Modell geschätzt:

Tabelle 7.9.: Schätzergebnisse fixed effects mit robusten Standardfehlern

Variable	Coefficient	(Std. Err.)
JOBEXP	-0.133**	(0.034)
CLASSIZE	0.109**	(0.033)
JOBSATIS	0.229**	(0.057)
MALE_T	0.217**	(0.082)
FOFTEN	0.123*	(0.058)
TRAINDUR_1	-0.284**	(0.068)
SEMFREQ	-0.142**	(0.029)
TUITION	0.144†	(0.077)
GUIDEM	0.243**	(0.067)
CHANGE	-0.110*	(0.051)
UNION	-0.141*	(0.058)
Intercept	-0.253**	(0.097)
N		2326
R <sup>2</sup>		0.05
F (11,2314)		8.43

Significance levels : † : 10% \* : 5% \*\* : 1%

Als Vergleichsmodell wurde nun wieder ein random effects Modell geschätzt, wobei erneut das Baltagi-Wu GLS Verfahren angewendet wurde.

Nach der Schätzung des Modells (Tab. 7.11) wurde mit den gespeicherten Schätzergebnissen beider Modell der Hausman-Test durchgeführt. Die Nullhypothese, dass keine systematischen Differenzen vorliegen musste auf dem 1% Niveau verworfen werden, so dass das fixed effects Modell vorzuziehen ist.

Tabelle 7.10.: Hausman Test	
Teststatistik $\chi^2(11)$	Signifikanz
135.62	0.0000

Das geschätzte fixed effects Modell ist also auch hier jenes, welches ich auswähle.

Tabelle 7.11.: Schätzergebnisse random effects GLS

<b>Variable</b>	<b>Coefficient</b>	<b>(Std. Err.)</b>
JOBEXP	-0.017**	(0.003)
JOBSATIS	0.058	(0.047)
CLASSIZE	0.004*	(0.002)
EXCHANGE	0.019	(0.054)
INSPECT	0.128*	(0.061)
MALE_T	0.088	(0.056)
FOFTEN	0.034	(0.050)
TUITION	-0.112	(0.077)
TRAINDUR_1	-0.039	(0.051)
SEMFREQ	-0.061*	(0.029)
CHANGE	-0.079†	(0.047)
GUIDEM	0.131*	(0.056)
GUIDEF	0.131†	(0.068)
UNION	-0.050	(0.047)
RURAL	-0.093	(0.063)
BACplus	0.103*	(0.051)
LIT1P	0.101†	(0.061)
HELP	-0.148*	(0.069)
Intercept	-0.171	(0.167)
N	2287	
$\chi^2_{(19)}$	80.194	

Significance levels : † : 10% \* : 5% \*\* : 1%

## Modelldiagnose

Als erstes erfolgt eine Überprüfung auf Multikollinearität, zu der wieder die VIFs angegeben werden.

Tabelle 7.12.: Multikollinearitätsüberprüfung

Variable	VIF	1/VIF
MALE_T	1.22	0.821244
JOBEXP	1.16	0.860896
TUITION	1.16	0.861009
FOFTEN	1.15	0.869977
GUIDEM	1.13	0.887429
CHANGE	1.11	0.900315
TRAINEDUR_1	1.10	0.910863
SEMFREQ	1.09	0.915562
UNION	1.08	0.925341
CLASSIZE	1.07	0.931407
JOBSATIS	1.07	0.935459

Hieraus folgere ich auch hier, dass keine starke Multikollinearität vorliegt. Auch die Überprüfung von Ausreißern führte zu keinem Ausschluss von Fällen.

Die Modellanpassung, bzw. die Erklärung der Varianz der abhängigen Variable innerhalb der Beobachtungen beträgt  $R^2=0,05$ . Also werden 5% der within-Varianz erklärt. Schätzt man das Dummyvariablenmodell, erreicht man einen  $R^2$  Wert zwischen 0,28 und 0,3 je nach Stichprobe.

### 7.3. Interpretation der Ergebnisse

Da die Modelle in den beiden vorangegangenen Abschnitten in etwa dasselbe messen, erfolgt zunächst eine Interpretation der Faktoren, die in beiden letztendlich ausgewählten Modellen auftauchen.

Dies betrifft insgesamt 8, von denen zumindest 6 das gleiche Vorzeichen haben. Für die Koeffizienten der erklärenden Variablen SEMFREQ (Häufigkeit der Seminarbesuche in den letzten 5 Jahren), sowie von GUIDEM (Lehrer verfügt über Anleitung zum Mathematikunterricht) trifft dies nicht zu, wobei in dem Modell zum Lernfortschritt SEMFREQ erst zum 10% Niveau signifikant von Null verschieden ist und GUIDEM sogar erst zum 20% Niveau.

In beiden Modellen zeigt sich zudem ein positiver Einfluss der Variablen Klassengröße, der schwer zu interpretieren ist. Das Problem den Effekt von Klassengröße, die von vielen Faktoren abhängt, zu bestimmen ist in der Bildungsforschung bekannt. So schätzt auch Michaelowa in einem Modell einen signifikant positiven Einfluss der Klassengröße und weist auf das Problem hin [Michaelowa, 2001]. In beiden Modellen wurden die abhängigen und metrischen erklärenden Variablen z-standardisiert, so dass der Koeffizient

von CLASSIZE (in beiden Modellen etwa gleich groß und zwar: 0,1) eine Verbesserung des Lernfortschritts bzw. der relativen Position um eine Zehntel Standardabweichung anzeigt, wenn die Klassengröße um eine Standardabweichung steigt.

Schüler mit Lehrern, die schon länger im Dienst sind und also einen höheren Wert bei der Variable JOBEXP (Anzahl der Dienstjahre) vorweisen, erzielen in beiden Modellen schlechtere Werte. In beiden Fällen wirkt sich eine kurze Ausbildungszeit von unter einem Jahr gegenüber einer längeren Ausbildungszeit negativ aus (Koeffizient von TRAINDUR\_1). Zudem wirkt sich positiv aus, wenn der Lehrer häufig auch privat Französisch spricht (FOFTEN) und wenn er privat Nachhilfe-Unterricht anbietet (TUITION). Wichtiger als eine gute Grundausbildung, die mit der Variable BACplus (erreichter Abschluss des Lehrers Abitur oder höher) im Lernfortschrittsmodell sogar einen deutlich negativen Einfluss hat, scheint also die Berufsausbildung zu sein.

Alle im letzten Absatz aufgeführten Effekte geben ein konsistentes Bild ab. Gut ausgebildete Lehrer, deren Ausbildung noch nicht lange zurück liegt und die den Schülern Nachhilfe anbieten, haben einen positiven Einfluss auf die Ergebnisse ihrer Schüler. Einen negativen Einfluss hat offenbar die Mitgliedschaft in einer Gewerkschaft (UNION).

Im Modell zum Lernfortschritt zeigten sich zudem der häufige Austausch des Lehrers mit Kollegen (EXCHANGE), sowie die Tatsache Wiederholer zu sein (REPEAT) als positive Einflüsse auf den Lernfortschritt. Letzteres überrascht, widerspricht es doch den Ergebnissen der PASEC-Studie. Allerdings ist gerade dieses Ergebnis leider sehr fraglich, da die Wiederholer einen anderen Test bekommen haben, als die übrigen Schüler und somit auf Grund der Unsicherheit über die Einschätzung der Schwierigkeitsgrade anhand der IRT-Methode Vorsicht geboten ist. Im Modell zur Positionsveränderung wurden die Wiederholer aus diesem Grund draussen gelassen.

Im ersten Modell muss die Variable TUITION im Gegensatz zu 'keine andere Beschäftigung' interpretiert werden, da die Variable ACTIV\_T (Lehrer hat Nebenbeschäftigung die nicht Nachhilfe ist) auch im Modell ist. Diese hat einen negativen Einfluss. Im Modell zur Positionsveränderung befindet sich nur die Variable TUITION (Nachhilfe), die also im Gegensatz zu keiner oder einer anderen Beschäftigung zu interpretieren ist. In beiden Fällen zeigt sich ein positiver Einfluss auf Lernfortschritt bzw. Positionsveränderung. Die Variable LOCLANG (Kenntnis der lokalen Stammessprache) hat einen negativen Einfluss, der nur insofern sinnvoll zu interpretieren ist, als dass Lehrer die aus anderen Regionen stammen mobiler und auch besser sind, möglicherweise auf Grund einer höheren Motivation sich in der Fremde durch viel Engagement Respekt zu verschaffen.

Im Modell zur Positionsveränderung zeigte sich noch ein positiver Einfluss der Variable JOBSATIS (Lehrer würde selben Beruf erneut wählen) und ein negativer Einfluss der Variable CHANGE (Lehrer beabsichtigt Bewerbung auf andere Stellen).

Zusammenfassend lässt sich sagen, dass der Faktor Klassengröße schwer einzuschätzen ist, da es schwer fällt, vorzuschlagen größere Klassen zu bilden, um den Kindern etwas Gutes zu tun. Zumindest geben diese Ergebnisse keine Berechtigung zu einer Forderung nach Reduzierung der Klassengrößen, was außerdem eine sehr teure Reform wäre.

Wesentlich vielversprechender ist eine Reform der Lehrerausbildung, die offenbar in vielen Fällen zu kurz ist. Auch wenn die Variable 'Anzahl der besuchten Seminare' kein einheitliches Bild abgibt, ist es auf Grund des negativen Einflusses der Anzahl der Dienst-

jahre sinnvoll Lehrer nach einiger Zeit zu Fortbildungen zu verpflichten. Zudem wäre es sehr nützlich Nachhilfe, z.B. in Form von Förderunterricht anzubieten.

## 8. Analysen Wiederholwahrscheinlichkeit

Nun folgt die Analyse zur Wiederholwahrscheinlichkeit mit der abhängigen Variable 'sitzen geblieben (ja/nein)' und verschiedenen erklärenden Variablen, von denen die wichtigste die Leistung ist, wobei hier die standardisierte Leistungsvariable verwendet wird.

### 8.1. Zeitraum 95/96 - 97/98

#### Variablenauswahl

In einem ersten Schritt habe ich mit einem stepwise Regressionsverfahren mit robusten Standardfehlern im gepoolten Probit Modell eine grobe Variablenvorauswahl getroffen. Bei den einzelnen im folgenden spezifizierten Modellen erfolgte dann eine weitere Auswahl anhand der Signifikanzwerten.

Tabelle 8.1.: Schätzergebnis r.e. Probit Modell

Variable	Coefficient	(Std. Err.)
Equation 1 : REPINY		
theta	-0.543**	(0.037)
CLASSIZE	0.009**	(0.002)
EXCHANGE	0.228**	(0.081)
LIT1P	0.295**	(0.073)
TRAINDUR_1	-0.108	(0.074)
ACTIV_T	0.131 <sup>†</sup>	(0.074)
TUITION	0.258*	(0.107)
NCLASS_D3	0.292**	(0.078)
NCLASS_D4	0.373**	(0.092)
WORK1	-0.022	(0.016)
UNION	-0.087	(0.068)
TVRADIO	-0.111	(0.068)
PILOTPRG	0.295**	(0.065)
HELP	-0.289**	(0.088)
Intercept	-2.287**	(0.186)
N		3372
Log-likelihood		-964.743
$\chi^2_{(14)}$		273.087

Significance levels : † : 10% \* : 5% \*\* : 1%

## Schätzung und Vergleich verschiedener Modelle

Wie im Methodenteil erläutert kommt ein fixed effects Logit Modell für mich nicht in Frage, da alle Fälle ohne Variation in der abhängigen Variable ausgeschlossen würden. In Frage kommt daher zunächst ein random effects Probit Modell, welches mit einem gepoolten Probitmodell mit denselben Variablen verglichen wird (Tab. 8.1 und 8.2).

Tabelle 8.2.: Schätzergebnisse gepooltes Probit Modell

Variable	Coefficient	(Std. Err.)
theta	-0.543**	(0.035)
CLASSIZE	0.009**	(0.002)
EXCHANGE	0.228**	(0.079)
LIT1P	0.295**	(0.073)
TRAINDUR_1	-0.108	(0.072)
ACTIV_T	0.131†	(0.073)
TUITION	0.258*	(0.108)
NCLASS_D3	0.292**	(0.076)
NCLASS_D4	0.373**	(0.092)
WORK1	-0.022	(0.016)
UNION	-0.087	(0.068)
TVRADIO	-0.111	(0.068)
PILOTPRG	0.295**	(0.064)
HELP	-0.289**	(0.089)
Intercept	-2.287**	(0.187)
N		3372
$R^2$ (McFadden)		0.143
Log-likelihood		-964.743
BIC		-25340.303
$\chi^2_{(14)}$		318.944

Significance levels : † : 10% \* : 5% \*\* : 1%

Gemessen an dem Log-Likelihood Wert passt sich das random effects Modell den Daten nicht besser an, als das gepoolte Probit-Modell. Anhand des von STATA automatisch mit ausgeführte LR-Test, der die Nullhypothese testet, dass der auf die Variation über die Zeit zurückzuführende Varianzanteil Null ist, zeigt an, dass die Nullhypothese bei einer Irrtumswahrscheinlichkeit von  $\alpha=0,01$  nicht abgelehnt werden kann. Ursache hierfür ist zum einen, dass zum einen einige erklärenden Variablen gar nicht über die Zeit variieren und zum anderen nur 3 Wellen vorliegen.

Eine weitere Möglichkeit stellt die Spezifizierung eines gepoolten Logit-Modells dar. Eine Schätzung eines solchen Modells mit den selben Variablen wie in den Probit Modellen (Tab. 8.3) ergibt eine leicht schlechtere Modellanpassung, gemessen am Log-Likelihood Wert, dem Pseudo- $R^2$ -Wert und den BIC-Werten der beiden gepoolten Modelle.

Da zur Modelldiagnose und Interpretation für das Logit-Modell mehr Verfahren unter

STATA zur Verfügung stehen, wird im folgenden trotzdem dieses Modell weiter verwendet.

Tabelle 8.3.: Schätzergebnisse gepooltes Logit-Modell

Variable	Coefficient	(Std. Err.)
theta	-1.001**	(0.065)
CLASSIZE	0.017**	(0.004)
EXCHANGE	0.394**	(0.151)
LIT1P	0.558**	(0.142)
TRAINDUR_1	-0.177	(0.139)
ACTIV_T	0.251†	(0.139)
TUITION	0.468*	(0.200)
NCLASS_D3	0.577**	(0.145)
NCLASS_D4	0.682**	(0.175)
WORK1	-0.038	(0.030)
UNION	-0.159	(0.129)
TVRADIO	-0.228†	(0.130)
PILOTPRG	0.526**	(0.123)
HELP	-0.498**	(0.167)
Intercept	-4.145**	(0.363)
N		3372
$R^2$ (McFadden)		0.140
Log-likelihood		-968.558
BIC		-25332.673
$\chi^2_{(14)}$		318.734

Significance levels : † : 10% \* : 5% \*\* : 1%

## Modelanpassung

Anhand des McFadden  $R^2$ -Wertes<sup>1</sup> ist die Modelanpassung nicht besonders gut. Dies könnte zum einen daran liegen, dass es wie im Theorie-Teil ausgeführt, keinen einheitlichen Maßstab für das Sitzenbleiben gibt und somit stärker die Position des Schülers relativ zu den anderen Schülern in der Klasse eine Rolle spielt, als der Vergleich zu anderen Schülern im Land. Dies wird bestätigt durch die bessere Anpassung der Modelle im nächsten Teil, in dem nur das Jahr 95/96 berücksichtigt wird und als erklärende Variable die Position in der Klasse zur Verfügung steht. Ein weiterer Grund ist die nicht erhobene Einstellung der Lehrer zum Thema Sitzenbleiben, die sicherlich auch einen wichtigen Erklärungsbeitrag liefern würde. Insofern fehlen möglicherweise wichtige erklärende Variablen. Eine Verzerrung der Schätzergebnisse durch Omitted Variable Bias ist nicht auszuschließen.

<sup>1</sup>s. Anhang

Eine Beurteilung der Güte des Modells anhand von Vorhersagen ergibt folgendes Bild, wenn ab einer geschätzten Wahrscheinlichkeit von 0.5 eine Realisierung vorhergesagt wird und darunter keine Realisierung:

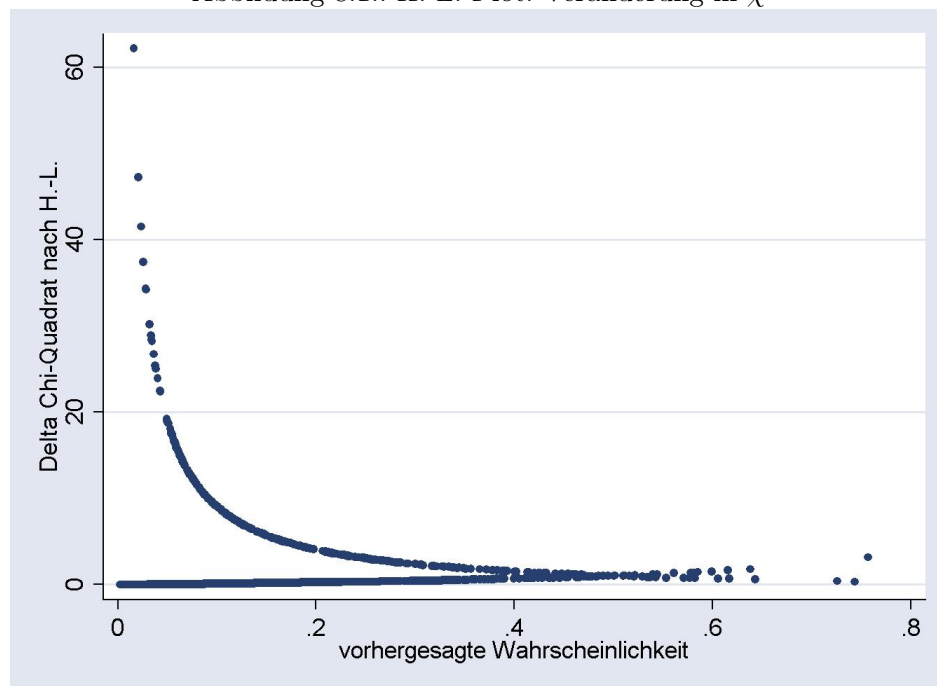
Tabelle 8.4.: Vorhersagen anhand des Modells

	Realisation	k. Realisation
Realisation vorhergesagt	18	16
k. Realisation vorhergesagt	333	3005

Es werden also nur wenige Sitzenbleiber vorhergesagt. Da mein Ziel jedoch nicht ein Modell für gute Voraussagen zu finden, sondern lediglich Faktoren, die einen Einfluss auf die Sitzenbleib-Wahrscheinlichkeit, ist dies noch kein Grund das Modell zu verwerfen.

Nun folgt eine graphische Modelldiagnose mit zwei Graphiken, die Hosmer und Lemeshow vorschlagen [Hosmer and Lemeshow, 1989].

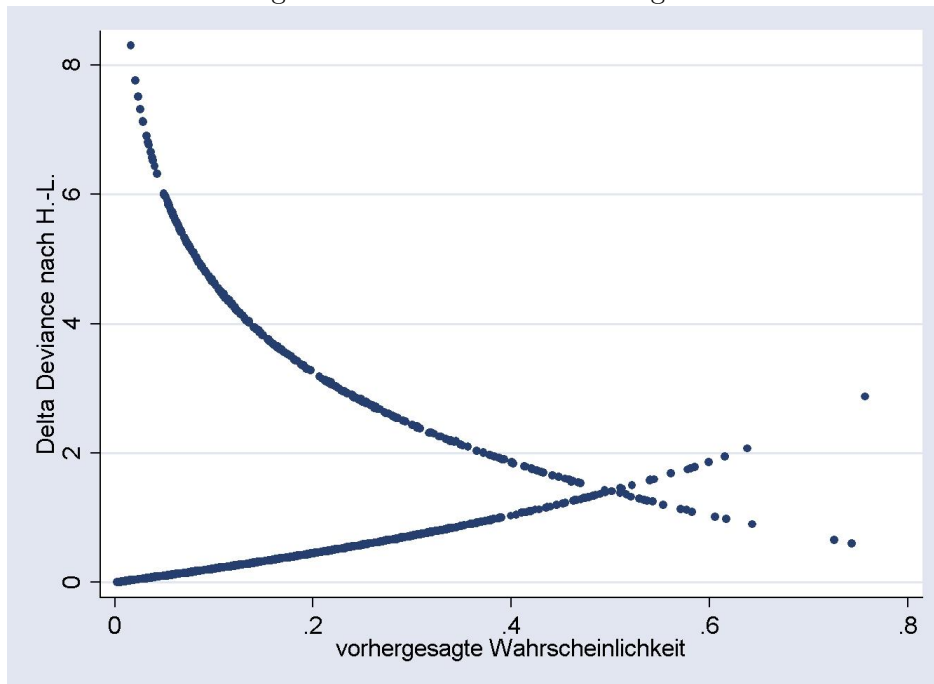
Abbildung 8.1.: H.-L. Plot: Veränderung in  $\chi^2$



Ein Punkt in Abb. 8.1 zeigt die Veränderung in Pearson's  $\chi^2$ -Wert bei Herausnahme des dazugehörigen Falles. Dazu wird die vorhergesagte Wahrscheinlichkeit einer Realisation auf der x-Achse abgetragen. In Abb. 8.2 zeigt ein Punkt die Veränderung in der Devianz bei Herausnahme des entsprechenden Falles an.

Die beiden in den Plots erkennbaren Kurven, gehören zum einen zu Fällen mit Realisation und zum anderen ohne Realisation. Die Kurven, die jeweils von oben links nach unten rechts verlaufen gehören zu den Fällen mit Realisation. Wie nach den schlechten

Abbildung 8.2.: H.-L. Plot: Veränderung in Devianz



Vorhersagen für Realisationen und den guten für Nicht-Realisationen zu vermuten war, erscheinen die meisten Ausreißverdächtigen auf dieser Kurve. Nach Hosmer/Lemeshow sollten die Punkte mit Werten größer 4 näher unter die Lupe genommen werden, da dieser Wert eine grobe Approximation des oberen 95% Perzentils der approximativen  $\chi^2$ -Verteilung der beiden Maße darstellt.

Dies betrifft im Fall 'Veränderung in Pearsons  $\chi^2$ ' 210 Beobachtungen, im Falle der Devianzveränderung 155 Beobachtungen. Diese Beobachtungen habe ich auf ihre Ausprägungen in den Modellvariablen hin überprüft und keine Auffälligkeiten, die auf Dateneingabefehler schließen lassen würden, festgestellt. Insofern bleibt nur der mäßige Modellfit festzuhalten.

## 8.2. Querschnitt 95/96

In diesem Teil wird nur die erste Welle berücksichtigt, um sinnvoll Informationen zu der Position der Schüler in der Klasse nutzen zu können. Dies ist nur im ersten Jahr möglich, da hier in jeder Klasse 20 Schüler befragt wurden, die zufällig ausgewählt wurden. In den folgenden Jahren, werden nur diese Schüler weiterverfolgt, so dass sich für diese Zeitpunkte nichts über die Position des Schülers in der Klasse sagen lässt, da neue Schüler (durch Wiederholen) hinzu gekommen sind.

Es wurde für die Schüler in 95/96 eine Variable gebildet, die anzeigt zu welchem Quartil der Schüler in der Klasse gehört. Hierzu wurde die gemittelte Leistung in Mathe und

Französisch genommen. Die Quartils-Variable wurde in 4 Dummy-Variablen zerlegt, von denen 3 in die Modelle aufgenommen wurden.

### Modellspezifizierung

Es wurden für diesen Teil nur Logit-Modelle verwendet. Die Variablenauswahl erfolgte erneut durch eine stepwise Regressionsprozedur. Zunächst wurde das Logit-Modell in Tab.8.5 geschätzt.

Tabelle 8.5.: Schätzergebnisse Logit Modell

Variable	Coefficient	(Std. Err.)
theta	-0.544**	(0.160)
lq	2.806**	(0.674)
lmq	2.154**	(0.634)
umq	0.979	(0.670)
RURAL	-0.565*	(0.248)
TRAINDUR_1	-0.449†	(0.231)
PILOTPRG	0.709**	(0.203)
CHANGE	-0.310	(0.201)
TVRADIO	-0.522*	(0.219)
LIT1P	0.498*	(0.205)
CLASSIZE	0.013†	(0.007)
Intercept	-5.241**	(0.746)
N		1473
Log-likelihood		-361.953
$R^2$ (McFadden)		0.214
$\chi^2_{(11)}$		196.785

Significance levels : † : 10% \* : 5% \*\* : 1%

Ein Test auf die gemeinsame Signifikanz der Quartils-Dummies ergab einen auf dem 1% Niveau signifikanten Unterschied zu Null.

Im folgenden wurden Vergleichsmodelle mit entweder nur den Quartilsdummies (Tab.8.7) oder nur der Leistungsvariable (Tab.8.6) als genestete Modelle spezifiziert und gegen das 'full model' getestet. In beiden Fällen ergab sich, dass das volle Modell eine signifikant bessere Anpassung liefert, als das reduzierte, was sich auch an den niedrigeren Pseudo- $R^2$  Werten erkennen lässt. Das Modell nur mit den Quartils-Dummies erreicht jedoch eine bessere Modellanpassung, als das Modell nur mit der Leistungsvariable.

Was im Vergleich zu den Modellen im vorigen Teil auffällt, ist, dass sich einige Variablen nicht mehr signifikant zeigen. Dies mag zum einen an der Aufnahme der Quartilsvariable liegen, zum anderen daran, dass hier nur die zweite Klasse betrachtet wird, in den vorigen Modellen zusätzlich die 3. und 4. Klasse.

Tabelle 8.6.: Schätzergebnisse mit Quartil-Dummies

Variable	Coefficient	(Std. Err.)
lq	3.875**	(0.597)
lmq	2.800**	(0.605)
umq	1.291†	(0.664)
RURAL	-0.443†	(0.243)
TRAINDUR_1	-0.442†	(0.228)
PILOTPRG	0.640**	(0.199)
CHANGE	-0.346†	(0.199)
TVRADIO	-0.598**	(0.217)
LIT1P	0.357†	(0.198)
CLASSIZE	0.009	(0.007)
Intercept	-5.257**	(0.734)
N		1473
Log-likelihood		-367.854
$R^2$ (McFadden)		0.201
$\chi^2_{(10)}$		184.984

Significance levels : † : 10% \* : 5% \*\* : 1%

Tabelle 8.7.: Schätzergebnisse mit Leistungsvariable

Variable	Coefficient	(Std. Err.)
theta	-1.186**	(0.113)
RURAL	-0.674**	(0.247)
TRAINDUR_1	-0.460*	(0.231)
PILOTPRG	0.751**	(0.204)
CHANGE	-0.211	(0.198)
TVRADIO	-0.468*	(0.216)
LIT1P	0.514*	(0.205)
CLASSIZE	0.018*	(0.007)
Intercept	-3.939**	(0.516)
N		1473
Log-likelihood		-378.091
$R^2$ (McFadden)		0.179
$\chi^2_{(8)}$		164.509

Significance levels : † : 10% \* : 5% \*\* : 1%

## Modellanpassung

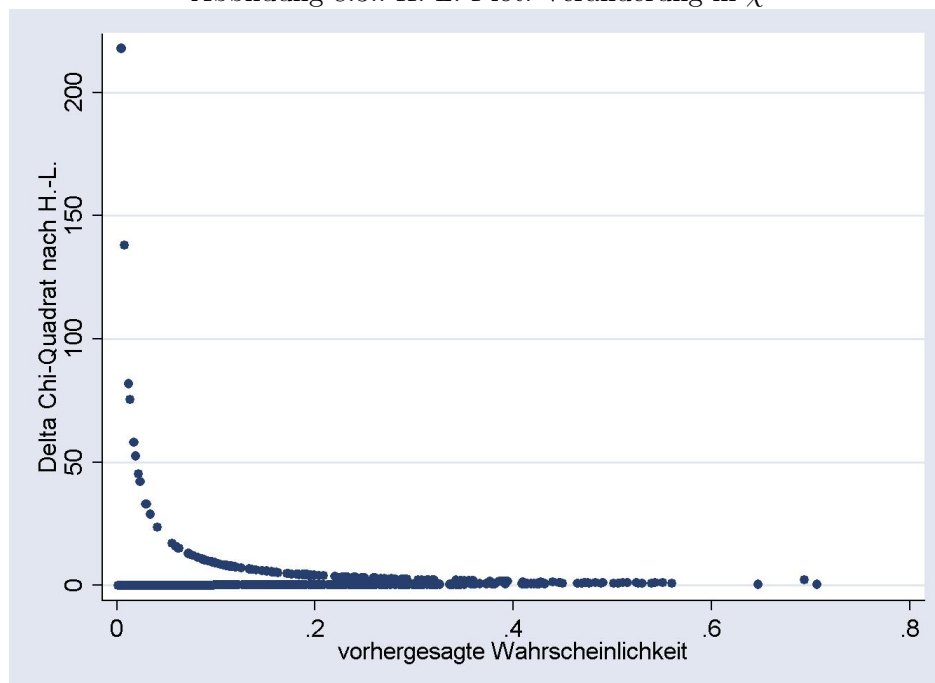
Das volle Modell wird nun noch einigen Begutachtungen zur Modellanpassung unterzogen. Zunächst folgt analog zum vorigen Teil eine Überprüfung der Vorhersageleistung.

Tabelle 8.8.: Vorhersagen anhand des Modells

	Realisation	k. Realisation
Realisation vorhergesagt	8	6
k. Realisation vorhergesagt	131	1328

Hier ergibt sich ein ähnliches Bild wie zuvor. Aber auch hier war das Ziel nicht ein Modell für gute Vorhersagen zu finden, sondern erklärende Faktoren zu ermitteln. Die Abb.8.3 und 8.4 zeigen die oben beschriebenen Diagnosegrafiken von Hosmer/Lemeshow.

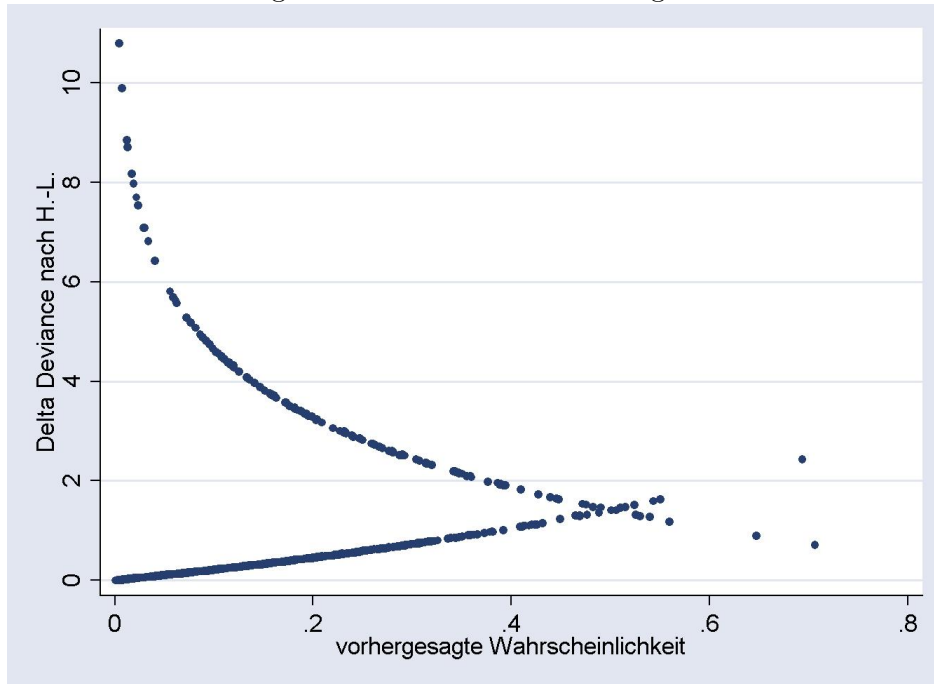
Abbildung 8.3.: H.-L. Plot: Veränderung in  $\chi^2$



Insbesondere die Abb.8.3 zeigt extreme Werte an. Die Grenze 4 überschreiten im Falle der Veränderung in  $\chi^2$  67 Beobachtungen und im anderen Fall 43. Auch hier ergab eine Überprüfung der Ausprägungen in den Modellvariablen keine Hinweise auf Datenfehler.

Der Modellfit ist für eine ganze Reihe von Fällen schlecht.

Abbildung 8.4.: H.-L. Plot: Veränderung in Devianz



### 8.3. Interpretation der Ergebnisse

Die Interpretation der Ergebnisse zur Wiederholwahrscheinlichkeit ist nicht ganz einfach, da beachtet werden muss, dass die Leistung des Schülers als erklärende Variable im Modell ist und somit die Einflüsse der anderen Faktoren unter Kontrolle der Leistung geschätzt wurden. Die Modelle zum Jahr 95/96 ergeben Dank der Quartile in den Klassen deutlich bessere Anpassungen an die Daten, als die Modelle für alle 3 Wellen, weshalb sich die Interpretation hauptsächlich auf die Variablen dieser Modelle stützt. Alle in diesen Modellen vertretenen Variablen (außer den Quartilen), tauchen auch in den Modellen des Teils davor mit gleichem Vorzeichen auf. Einige weitere Variablen zeigten nach Aufnahme der Quartilsdummies keinen signifikanten Erklärungsbeitrag mehr und werden deshalb hier nicht interpretiert.

Mit höherer Leistung des Schülers nimmt die Wahrscheinlichkeit sitzen zu bleiben ab. Ebenso zeigen die Dummies für die Quartile, die in den 95/96-Modellen auftauchen, positive Werte. Es sind Dummies für die drei unteren Quartile, so dass sie zur Referenzkategorie oberstes Quartil zu interpretieren sind. Diese Werte waren so zu erwarten und haben den mit Abstand größten Einfluss auf die Sitzenbleib-Wahrscheinlichkeit. Ein Modell nur mit diesen Variablen (Leistung und Quartile) erzielt immernoch einen Pseudo- $R^2$  Wert von 0,174 mit den Daten von 95/96.

Die Klassengröße hat in allen Modellen einen signifikant positiven Effekt auf die Sitzenbleib-Wahrscheinlichkeit.

In Tabelle 8.9 sind die exponentierten Koeffizienten der Logitschätzung des Modells

Tabelle 8.9.: Schätzergebnisse Logistische Regression

Variable	exp(Coefficient)	(Std. Err.)
lq	36.33825**	(18.98556)
lmq	12.39302**	(6.586575)
umq	2.734631†	(1.634143)
RURAL	.6637496†	(.1598477)
TRAINDUR_1	.6336738*	(.1443151)
PILOTPRG	1.917395**	(.3803597)
CHANGE	.713981†	(.1418083)
TVRADIO	.5483629**	(.1184429)
LIT1P	1.452608†	(.2862541)
CLASSIZE	1.00845	(.0067449)
Intercept	-4.974**	(0.674)
N		1476
Log-likelihood		-372.395
$R^2$ (McFadden)		0.20
$\chi^2_{(10)}$		181.016

Significance levels : † : 10% \* : 5% \*\* : 1%

nur mit den Quartilsdummies (ohne Leistungsvariable) dargestellt, das eine ähnlich gute Anpassung wie das volle Modell ergab. Diese Ergebnisse sind einfacher zu interpretieren. So steigert die Tatsache zum untersten Leistungs-Quartil in der Klasse zu gehören im Gegensatz zum obersten die Chance sitzen zu bleiben gegenüber nicht sitzen zu bleiben (also die odds) um das 36 fache, beim unteren mittleren Quartil um das 12 fache und beim oberen mittleren Quartil um das 2,7 fache.

Darüber hinaus haben Schüler gleichen Niveaus in der Klasse aber auch im Gesamtvergleich (die Vorzeichen der Koeffizienten ändern sich vom Teilmodell zum vollen Modell nicht) eine niedrigere Wahrscheinlichkeit sitzen zu bleiben, ebenso wie Schüler schlecht ausgebildeter Lehrer. In beiden Fällen sinken die Chancen sitzen zu bleiben mit dem Faktor 0,6.

Schüler an Schulen, die mit internationalen Organisationen kooperieren (PILOTPRG) müssen häufiger sitzen bleiben, was überrascht. So sollte man denken, dass diese Organisationen auf niedrige Raten drängen. Andererseits betrifft dies fast die Hälfte der Schulen und die Kooperationen sind zum Großteil wohl nicht so eng. Außerdem könnte es sein, dass sich die Organisationen gerade Schulen in schwache Regionen ausgesucht haben, in denen die Wiederholraten besonders hoch sind.

Eine um einen Schüler größere Klasse birgt ein leicht geringeres Risiko sitzen zu bleiben, obwohl der Koeffizient erst zum 20% Niveau signifikant ist.

Was den sozialen Hintergrund des Schülers angeht, so zeigt sich, dass Schüler weniger oft sitzen bleiben, die zuhause Zugang zu TV und Radio haben, was ein grober Hinweis auf einen positiven Einfluss sozial besser gestellter Eltern auf die Entscheidung des Lehrers über das eigene Kind darstellen könnte. Dagegen spricht der deutlich positive Effekt

auf die Wiederholwahrscheinlichkeit durch Eltern, von denen mindestens ein Teil lesen und schreiben kann (LIT1P). Somit ergibt sich in diesem Punkt kein einheitliches Bild.

Festzuhalten bleibt zu diesen Analysen, dass es sinnvoll wäre einheitliche Maßstäbe im ganzen Land stärker durchzusetzen. Der Gegensatz Stadt zu Land in den Sitzenbleib-Wahrscheinlichkeiten macht dies deutlich. Zudem sollte es unbedingt Ziel einer Reform sein, in der Ausbildung der Lehrer das Problem des Sitzenbleibens zu thematisieren. Das insbesondere besser ausgebildete Lehrer das Risiko sitzen zu bleiben erhöhen, ist bedenklich. Zum Einfluss der Familie kann hier nicht viel gesagt werden. Hauptsächlich schlägt sich dieser wohl schon in der Leistung der Kinder nieder. Interessant wäre eine Befragung der Lehrer zu ihren Einstellungen zum Sitzenbleiben gewesen. Eine solche ist in der PASEC-Studie (s.o.) im Jahr 98/99 auch durchgeführt worden, allerdings nur mit 30 Lehrern.

## 9. Fazit

Im wesentlichen bestätigen die Ergebnisse die Forschungsergebnisse anderer Studien. Sie haben insofern Bedeutung, als dass sie nicht durch Ability verzerrt geschätzt wurden und Ergebnisse zu anderen Ländern für den Senegal bestätigen. Die Bedeutung der Lehrqualität wird in jüngsten Studien betont [Rivkin et al., 2005].

Die Wichtigkeit der Lehrerausbildung für die Schulqualität wurde auch in dieser Arbeit deutlich. Anders als zur Klassengröße können hier eindeutig Verbesserungen durch Reformen erzielt werden. Ebenso würde eine Bereitstellung von Förderunterricht wahrscheinlich eine Verbesserung des Lernfortschritts bringen.

Einheitliche Lehrpläne und weitere Maßnahmen zur Erreichung gleicher Maßstäbe zum Sitzenbleiben wären ebenfalls wünschenswert, um das System fairer zu gestalten.

Inwiefern der soziale Hintergrund der Schüler eine Rolle spielt für Lernfortschritt und Sitzenbleib-Wahrscheinlichkeit, muss in anderen Studien und wenn möglich auf besserer Datengrundlage untersucht werden. Die Daten der oben dargestellten neuesten Senegalstudie bieten hier sicherlich einige Möglichkeiten.

Am wirkungsvollsten gegen hohe Wiederholraten wird wahrscheinlich die Aufklärung über das Problem in der Lehrerausbildung und in Fortbildungen sein. Wie in der PASEC-Studie gezeigt wurde, sind 77% der befragten Lehrer von der Effektivität des Sitzenbleibens überzeugt [PASEC, 2004].

Die Ergebnisse dieser Arbeit zeigen, dass sich eine längere Ausbildung sogar positiv auf die Wahrscheinlichkeit bei diesem Lehrer sitzen zu bleiben auswirkt. Das spricht sehr dagegen, dass das Thema in der Ausbildung angemessen behandelt wird.

# Anhang

# Variablenbeschreibung

## Kapitel 7

Zunächst erfolgt eine Beschreibung der Variablen, die in den Modellen des 7. Kapitels vorkommen:

Tabelle 9.1.: Variablenbeschreibung

<b>Name</b>	<b>Bedeutung</b>
JOBEXP	Anzahl Dienstjahre
CLASSIZE	Klassengröße
SEMFREQ	Anzahl besuchter Seminare in den letzten 5 Jahren
EXCHANGE	Lehrer tauscht sich oft mit Kollegen aus
LOCLANG	Lehrer spricht lokale Sprache
FOFTEN	Lehrer spricht privat häufig Französisch
BACplus	Lehrer hat einen Bildungsabschluss Abitur oder höher
TRAINDUR_1	Lehrer hat eine Berufsausbildung von unter einem Jahr
ACTIV_T	Lehrer hat Nebentätigkeit
TUITION	Lehrer gibt privaten Nachhilfeunterricht
GUIDEM	Lehrer verfügt über Unterrichtsanleitung für Mathematik
GUIDEF	Lehrer verfügt über Unterrichtsanleitung für Französisch
UNION	Lehrer ist Gewerkschaftsmitglied
REPEAT	Schüler wiederholt Jahr
MALE_T	Geschlecht des Lehrers: Mann
LIT1P	mindestens ein Elternteil kann lesen und schreiben
NCLASS_D3	Schüler ist in Klassenstufe 3
NCLASS_D4	Schüler ist in Klassenstufe 4
HELP	Schüler bekommt zuhause Hilfe bei Schulaufgaben
HOMEWORK	Schüler macht zuhause Hausaufgaben
INSPECT	Inspektor der Behörde kam seit Anfang des Jahres
JOBSATIS	Lehrer würde erneut den Lehrerberuf wählen
CHANGE	Lehrer beabsichtigt Bewerbung auf andere Stelle
RURAL	Gegend ländlich (im Gegensatz zu urban)

Mit Ausnahme der Variablen CLASSIZE, JOBEXP und SEMFREQ sind alle Dummy-Variablen.

## Kapitel 8

Folgende Variablen kamen in den Modellen des 8. Kapitels noch hinzu:

Tabelle 9.2.: Variablenbeschreibung

<b>Name</b>	<b>Bedeutung</b>
theta	Leistung des Schülers (Mittel aus Französisch- und Mathematikleistung)
WORK1	Anzahl der Tätigkeiten, die Schüler zuhause erfüllen muss (Einkaufen, Abwaschen, Babysitten etc.)
TVRADIO	Verfügbarkeit von TV und Radio zuhause
PILOTPRG	Schule nimmt an speziellem Programm teil (z.B. mit NGO oder Austausch mit ausl. Schule)
lq	Schüler gehört zu unterstem Leistungsquartil in seiner Klasse
lmq	...zu unterem mittleren Quartil
umq	...zu oberem mittleren Quartil

Mit Ausnahme der Variable theta sind alle Dummy-Variablen.

## McFadden und Co.

Zur Einschätzung der Modellgüte in Logit und Probit Modellen stehen u.a. eine ganze Reihe von Pseudo- $R^2$  Maßen zur Verfügung. STATA gibt standardmäßig das Maß nach McFadden aus. Es können aber auch noch weitere angefordert werden. Die in SPSS gängigen Maße nach Cox&Schnell oder nach Nagelkerke stehen aber nicht zur Verfügung.

Das Pseudo- $R^2$  Maß nach McFadden ist wie folgt definiert [Long, 1997]:

$$R_{McF}^2 = 1 - \frac{\ln \hat{L}(M_\beta)}{\ln \hat{L}(M_\alpha)}$$

wobei  $L(M_\alpha)$  die Likelihood des Modells ist, das nur den Intercept enthält und  $L(M_\beta)$  die Likelihood des Modells, das auch die Regressoren enthält.

# Literaturverzeichnis

- [Arellano, 2003] Arellano, M. (2003). *Panel Data Econometrics*. Oxford University Press.
- [Barnes, 1999] Barnes, D. (1999). Causes of dropping out. In Randall and Anderson, editors, *Schooling for Success. Preventing Repetition and Dropout in Latin American Primary Schools*. Columbia University Seminar Series.
- [Cortina, 1999] Cortina, R. (1999). The impact of international organisations on educational policy in latin america. In Randall and Anderson, editors, *Schooling for Success. Preventing Repetition and Dropout in Latin American Primary Schools*. Columbia University Seminar Series.
- [Dumas and Lambert, 2005] Dumas, C. and Lambert, S. (2005). Children education in senegal : how does family background influence achievement. *Research Unit Working Papers Laboratoire d'Economie Appliquee (INRA)*, 0503.
- [Forestier, 2004] Forestier, C. (2004). Le redoublement permet-il de résoudre les difficultés rencontrées au cours de la scolarité obligatoire? *AVIS du haute Conseil de l'Evaluation de l'Ecole* (<http://cisad.adc.education.fr/hcee>), (14).
- [Hambleton et al., 1991] Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. SAGE Publications.
- [Hamilton, 2004] Hamilton, L. C. (2004). *Statistics with STATA*. Thomson.
- [Hansen et al., 2004] Hansen, K. T., Heckman, J. J., and Mullen, K. J. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121:39–98.
- [Hosmer and Lemeshow, 1989] Hosmer, D. W. and Lemeshow, S. (1989). *Applied logistic Regression*. Wiley.
- [Kennedy, 2003] Kennedy, P. (2003). *A Guide to Econometrics*. Blackwell Publishing.
- [Long, 1997] Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. SAGE.
- [Michaelowa, 2001] Michaelowa, K. (2001). Primary education quality in francophone sub-saharan africa: Determinants of learning achievement and efficiency considerations. *World Development*, 29(10):1699–1716.

- [Michaelowa, 2002] Michaelowa, K. (2002). Teacher job satisfaction, student achievement, and the cost of primary education in francophone sub-saharan africa. *HWWA Discussion Paper*, 188.
- [Michaelowa and Fröhlich, 2005] Michaelowa, K. and Fröhlich, M. (2005). Peer effects and textbooks in primary education: Evidence from francophone sub-saharan africa. *HWWA Discussion Paper*, 311.
- [PASEC, 2004] PASEC (2004). *Le Redoublement: Pratiques et conséquences dans l'enseignement primaire au Sénégal*. CONFEMEN.
- [Paul and Troncin, 2004] Paul, J.-J. and Troncin, T. (2004). Les apports de la recherche sur l'impact du redoublement comme moyen de traiter les difficultés scolaires au cours de la scolarité obligatoire. *Haut Conseil de l'Evaluation de l'Ecole* (<http://cisad.adc.education.fr/hcee>), (14):1–43.
- [Pino and Koslinski, 1999] Pino, I. R. and Koslinski, M. C. (1999). Government programs to eliminate repetition, school dropout, and exclusion in brazil. In Randall and Anderson, editors, *Schooling for Success. Preventing Repetition and Dropout in Latin American Primary Schools*. Columbia University Seminar Series.
- [Randall et al., 1999] Randall et al. (1999). *Schooling for Success. Preventing Repetition and Dropout in Latin American Primary Schools*. Columbia University Seminar Series.
- [Rivkin et al., 2005] Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73:417–458.
- [Schiefelbein and Schiefelbein, 1999] Schiefelbein, E. and Schiefelbein, P. (1999). Repetition and quality of education in chile. In Randall and Anderson, editors, *Schooling for Success. Preventing Repetition and Dropout in Latin American Primary Schools*. Columbia University Seminar Series.
- [STATA, 2003] STATA (2003). *Cross-Sectional Time-Series*. STATA Press.
- [Wooldridge, 2002] Wooldridge, J. M. (2002). *Econometric Analyses of Cross Section and Panel Data*. MIT Press.