

Analysen politikwissenschaftlicher Datensätze mit Stata

JOHANNES
GUTENBERG
UNIVERSITÄT
MAINZ

Sitzung 8: „Ethnicity, Insurgency, and
Civil War“ (Fearon/Latin 2003)

Vorbereitung

- Bitte laden Sie den Datensatz
z: \daten\fl\repdata.dta

Kontext

- Seit den 1950er Jahren ist die Zahl der Bürgerkriege weltweit erheblich angestiegen
- (Spätestens) in den letzten beiden Dekaden hat die Zahl ethnischer Konflikte zugenommen

Fragestellung

- Hatte das Ende des Kalten Krieges einen relevanten Einfluß auf die Zahl der Bürgerkriege?
- Können die Bürgerkriege seit dem Zweiten Weltkrieg wirklich durch
 - ethnische Heterogenität per se
 - bzw. durch die Diskriminierung von ethnischen Gruppen und die Politisierung ethnischer Spannungslinien erklärt werden?
- Oder läßt sich der Ausbruch von Bürgerkriegen eher durch die Schwäche der Zentralregierungen erklären?

Design

- Untersuchungseinheit sind Länder bzw. Kolonialreiche → Reines Makro-Modell!
- Für jedes Land wird jedes Jahr kodiert (0/1), ob ein Bürgerkrieg *beginnt*
 - Auseinandersetzung zwischen Regierungstruppen (im weiteren Sinne) und organisierten Gegnern
 - mit wenigstens 1000 Toten insgesamt
 - wenigstens 100 Toten pro Jahr
 - wenigstens 100 Toten auf beiden Seiten

Design

- Die jährliche Beobachtung bildet jeweils einen „Fall“
- rund 6000 „Fälle“ (1945-1999) insgesamt
- Zahl der Länder nimmt von 63 auf 156 zu
- tab year
- tab country
- „Panel“ von Ländern, die jährlich untersucht werden
- Als Time-Series Cross-Sectional Design bekannt (TSCS)
- Beliebt, weil es das „few n, many k“ Problem entschärft

TSCS

	1975	1976	1977	...
Deutschland	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Belgien	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Frankreich	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
...	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Zur Erinnerung...

- „Normales“ Regressionsmodell
 - $y = a + \beta_1 x_1 + \beta_2 x_2 \dots + \varepsilon$ bzw.
 - $y = a + b_1 x_1 + b_2 x_2 \dots + e$
- abhängige Variable geht auf unabhängige Variable (systematischer Teil) und (hoffentlich) zufällige Einflüsse (disturbances, Störungen) zurück
- diese Störungen sind voneinander unabhängig und können über Residuen (Differenz zwischen beobachtetem und erwartetem Wert der abhängigen Variablen) geschätzt werden
- Standardfehler der Koeffizienten zeigt an, wie Parameterschätzungen aufgrund von Störungen um den wahren Wert variieren
- Berechnung basiert auf der Schätzung der zufälligen Einflüsse über die Residuen

Zur Erinnerung...

- Wenn die zufälligen Einflüsse (positiv) miteinander korreliert sind,
- handelt es sich faktisch um einen systematischen Einfluß, der aber nicht modelliert wird
 - Die beobachteten Residuen sind zu klein
 - unterschätzen deshalb den Einfluß der Störungen,
 - die Standardfehler sind zu klein,
 - nicht modellierte Störgrößen haben tatsächlich einen größeren Einfluß auf die abhängige Größe als angenommen

Fehlerprozeß bei TSCS

- störende Einflüsse sind typischerweise
 - von länder- / regionsspezifischer Varianz (innerhalb Afrikas/Nigerias variieren sie stärker als innerhalb Europas/Deutschlands: Heteroskedastizität)
 - in der Zeit korreliert (zu einem bestimmten Zeitpunkt beeinflußt eine im Modell nicht erfaßte Größe die Wahrscheinlichkeit von Bürgerkriegen in einigen oder allen Ländern)
 - über die Zeit korreliert (eine Störung im Januar 1963 wirkt ein oder zwei Jahre später immer noch nach: Autokorrelation)
- dies kann u.a. dazu führen, daß die Standardfehler über-optimistisch sind, d.h. man schenkt seinen Resultaten zuviel Vertrauen

Fehlerprozeß bei TSCS

- Abhilfe
 - GLS-ARMA als spezielles Schätzverfahren („klassische“ Lösung)
 - robuste Standardfehler mit Clustering als „kleine Lösung“
 - Diverse Verfahren zur Korrektur der Standardfehler
 - evtl. auch gar kein Problem

Modellierung

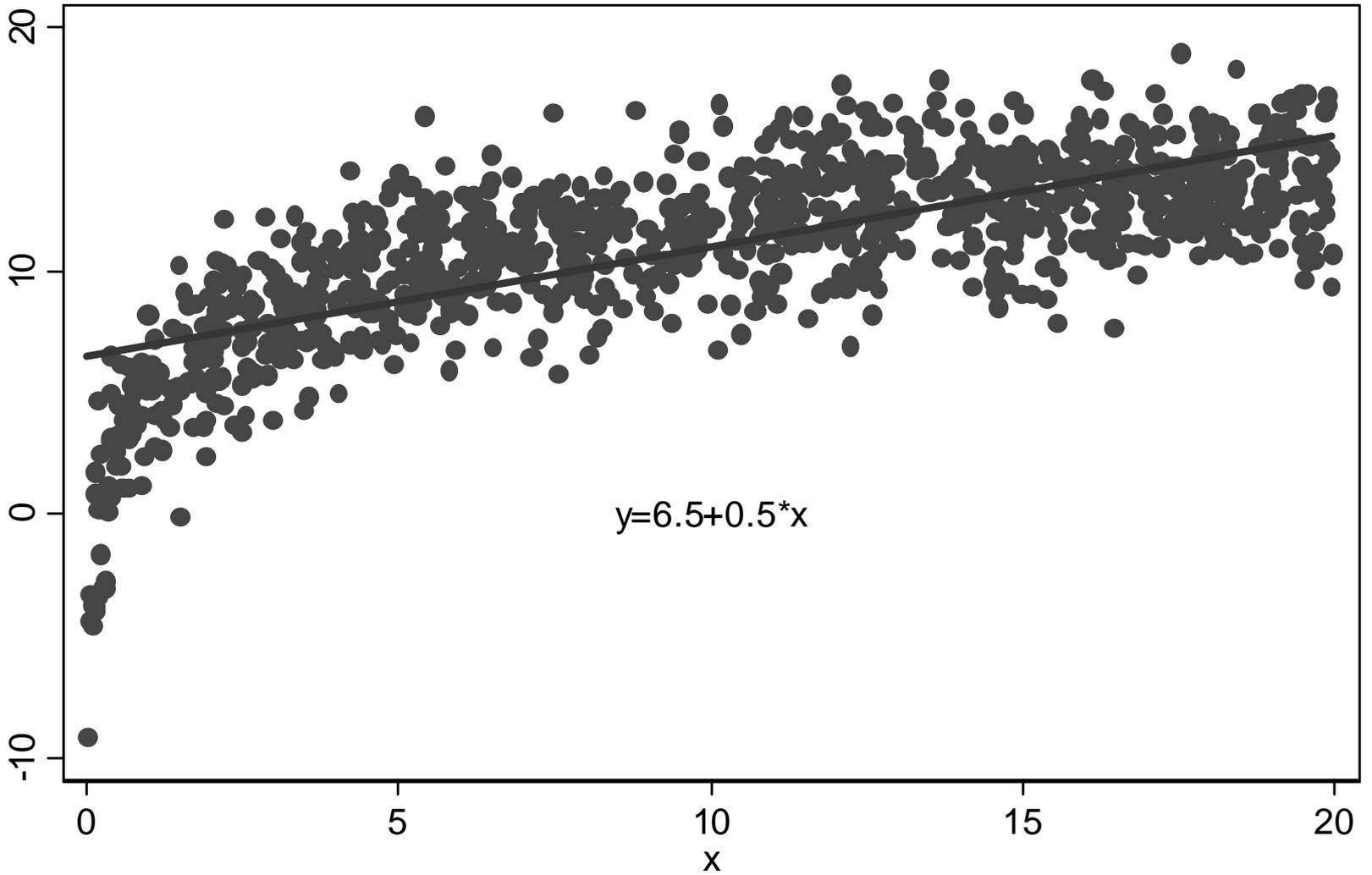
- Abhängige Variable: Ausbruch (1) vs. Nicht-Ausbruch (0) eines Bürgerkrieges wird zurückgeführt auf diverse unabhängige Variablen
- Wie hängt die Wahrscheinlichkeit eines Ausbruchs mit Bürgerkriegen in der Vergangenheit zusammen?
- Mögliche Alternativen:
 - Zahl der Bürgerkriege in dem betreffenden Land in diesem Jahr (Count-Daten)
 - Findet in diesem Land in diesem Jahr ein Bürgerkrieg statt (0/1) (Ereignis-Daten)

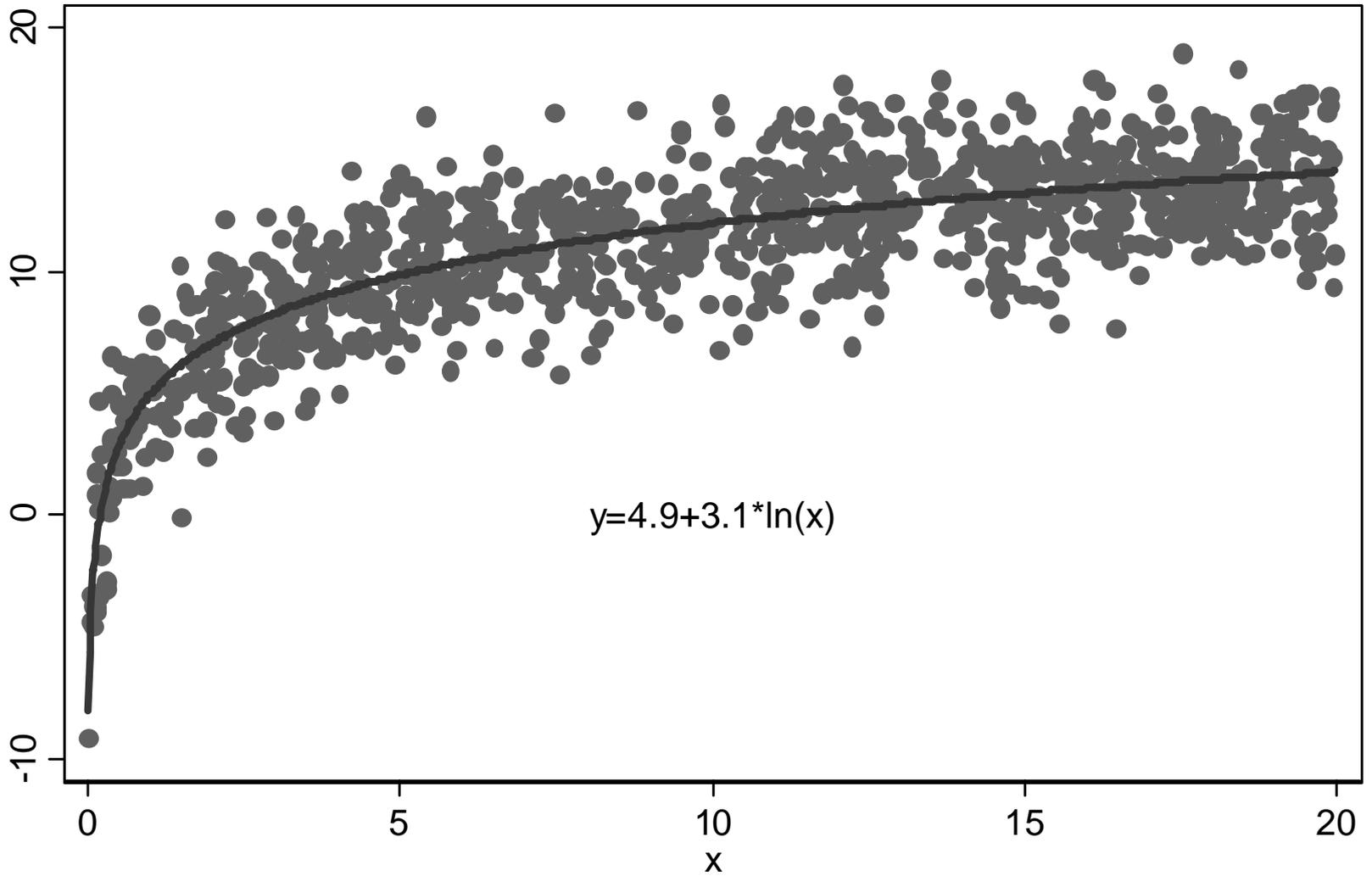
Besonderheiten in der Modellierung

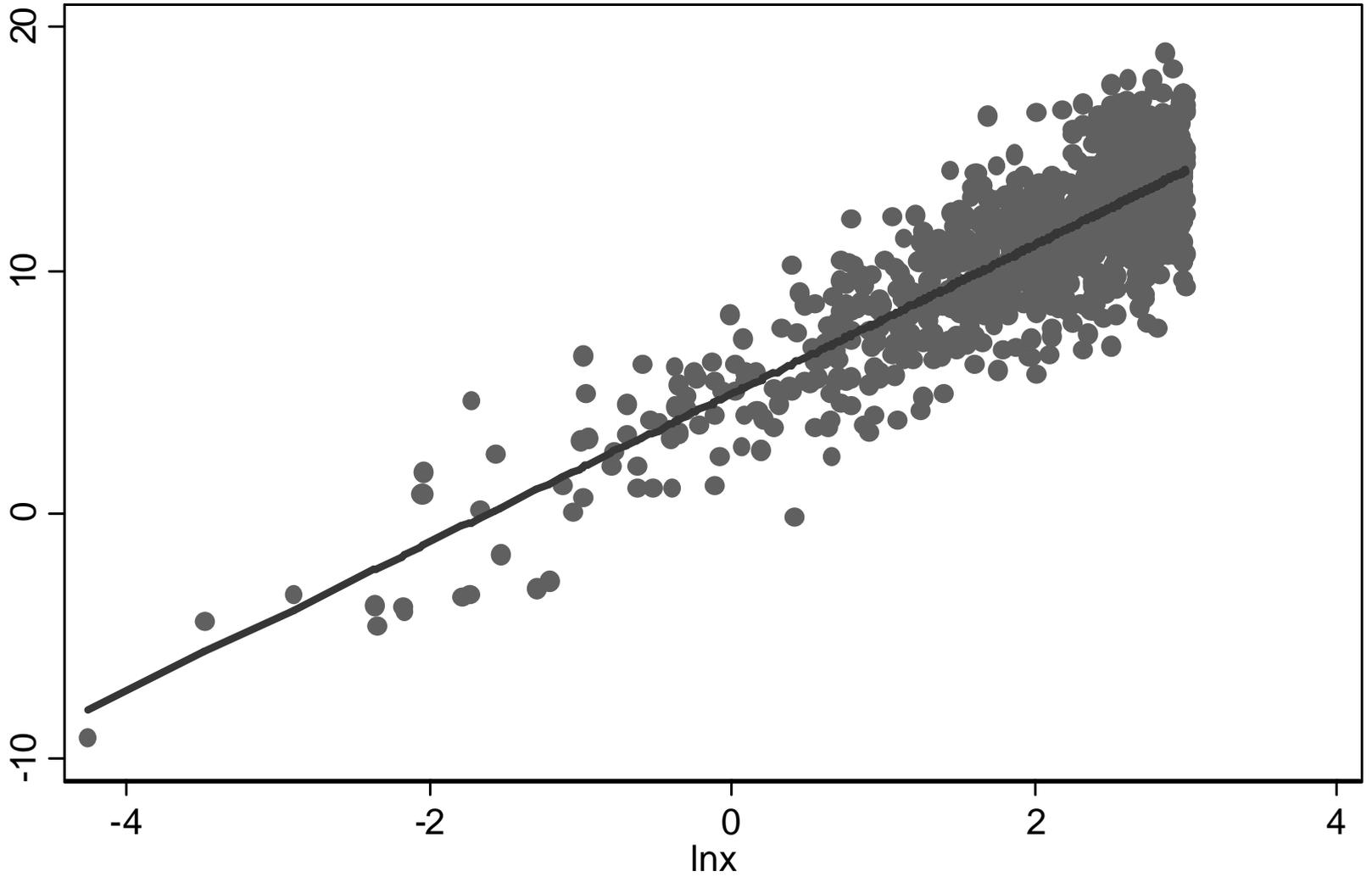
- „Verzögerte“ Variablen (lags)
 - der Wert einer Variablen wird jeweils durch den Wert der Vor-Periode (hier: Vorjahr) ersetzt
 - in Stata mit dem L. Operator
 - sort ccode year
 - gen vpop=L.pop
 - list country year pop vpop
 - Veränderungen einer unabhängigen Variablen brauchen Zeit, um wirksam zu werden
 - Kausalitätsprobleme können entschärft werden (verursacht niedriges GNP Bürgerkriege oder umgekehrt?)
 - Hier verwendet bei GNP, Bevölkerung, Regime-Variablen

Besonderheiten in der Modellierung

- Logarithmierung der unabhängigen Variablen
 - kann evtl. non-lineare Beziehungen linearisieren und Probleme mit Heteroskedastizität reduzieren
 - Inhaltliche Bedeutung:
 - die Veränderung der abhängigen Variablen
 - hängt nicht von absoluten,
 - sondern von proportionalen (prozentualen) Veränderungen der unabhängigen Variablen ab
 - D.h. $x(2 \rightarrow 4)$ und $x(4 \rightarrow 8)$ haben den gleichen Effekt
 - Prinzipiell auch bei Logit-Modellen möglich







Replikation Abbildung 1

- Wir wollen temporär einen Datensatz haben, der
 - jeweils ein Jahr als Fall enthält
 - für dieses Jahr die Gesamtzahl der Bürgerkriege (Summe über alle Länder) enthält
 - den Anteil der Länder mit Bürgerkrieg an der Zahl aller Länder enthält
- Variablen im ursprünglichen DS
 - war: Krieg ja nein
 - wars: Zahl der Bürgerkriege im Land

Zum Ausprobieren

- preserve
- use z: \daten\logreg
- reg y x
- predict lin
- reg y lnx
- predict log
- graph twoway scatter y lin log x
- restore

Replikation Abbildung 1

- preserve
- Aggregation:
 - collapse (mean) war (sum) wars,by(year)
 - Jährlicher Mittelwert der Variable Bürgerkrieg ja/nein ~ Anteil der Länder mit Bürgerkrieg
 - replace war=war*100 (um auf Prozente zu kommen)
 - Summe über die Zahl der Konflikte pro Land = Gesamtzahl
- Grafik hat zwei unterschiedlich skalierte y-Achsen (vermeiden!)
- doedit z: \daten\fl\grafik1.do
- Falls restore fehlschlägt: use z: \daten\fl\repdata,replace

Hypothesen

- Klassisch:
 - Ethnische Differenzen und „Ungerechtigkeiten“
 - (im Zusammenhang mit Modernisierungsprozessen)
 - lassen das Risiko von Bürgerkriegen steigen
 - Vorsicht: Modelliert wird der *Beginn* von Bürgerkriegen!
- Revisionistisch:
 - Bedingungen, die (kleinere) „Aufstände“ (insurgency) und die Anwendung von Guerilla-Methoden erleichtern, machen Bürgerkriege wahrscheinlicher
 - Nochmal Vorsicht: Modelliert wird der *Beginn* von Bürgerkriegen!

Hypothesen Klassisch

- Höhere Diversität → Höheres Risiko für Bürgerkrieg
- Besonders dann, wenn höheres Pro-Kopf-Einkommen (Indikator für Modernisierung)
 - (Glauben wir das so?)
- Höheres Risiko dort, wo es eine ethnische Mehrheit und eine signifikante Minderheit gibt
- Geringeres Risiko in Staaten mit hohen Werten für Demokratie / bürgerliche Freiheiten
 - (Problem?)
- Höheres Risiko dort, wo Minderheiten diskriminiert werden
- Höheres Risiko bei großen Ungleichheiten in der Verteilung der Einkommen
- Höheres Risiko für einen ethnischen Bürgerkrieg, wenn eine ethnische Minderheit mindestens fünf Prozent der Bevölkerung stellt

Hypothesen revisionistisch

- (Wenn Bürgerkriege hauptsächlich auf insurgency zurückgehen)
 - Dann sollten unwegsames Gelände, Rückzugsgebiete in anderen Ländern und eine lokale Bevölkerung, die dazu gebracht werden kann, die Rebellen zu unterstützen sowie eine Basis in ländlichen Gebieten
 - die Wahrscheinlichkeit eines Bürgerkrieges erhöhen
- Hilfshypothese: Pro-Kopf-Einkommen ist ein Proxy für
 - die Stärke des Regierungsapparates
 - den Ausbau der Infrastruktur
- Außerdem soll niedriges Pro-Kopf-Einkommen den erwarteten Nutzen einer Guerilla-Karriere erhöhen
 - (Mikro-Argument)
 - (Sind das die einzigen Interpretationen für einen Effekt des Pro-Kopf-Einkommens?)

Hypothesen revisionistisch

- Aufstände (und damit Bürgerkriege) sind wahrscheinlicher
 - kurz nach der Unabhängigkeit
 - in instabilen Staaten
 - in semi-autokratischen Staaten
 - in großen Staaten (Bevölkerung)
 - in nicht-zusammenhängenden Staaten
 - dort wo andere Staaten Aufstände unterstützen
 - dort wo mit kleinen, leicht zu transportierenden Gütern (Diamanten, Drogen) ein Krieg finanziert werden kann
 - in Staaten, die im großen Umfang Öl exportieren
- *Wenn das Pro-Kopf-Einkommen kontrolliert wird, haben ethnische Heterogenität, Freiheiten, Einkommensungleichheiten usw. keinen Effekt mehr*

Probleme?

- Indikatoren
- Hilfsannahmen
 - Bürgerkriege in erster Linie als „Aufstände“ zu betrachten
 - Interpretation von GDP/per capita unklar

Abbildung 2

- Beruht nicht auf Modellschätzung – keine Kontrolle anderer Faktoren
- Contour-Plot (mit Stata nicht möglich, nicht immer sehr anschaulich)
- Zeigt den gemeinsamen Einfluß von GDP/pro Kopf und ethnischer Heterogenität auf das Risiko eines Bürgerkrieges
- Botschaft in etwa:
 - Für alle Staaten nimmt das Risiko neuer Bürgerkriege mit einem höheren GDP ab
 - Bei den reichen Staaten nimmt das geringe Risiko eines (neuen) Bürgerkrieges mit zunehmender Homogenität der Bevölkerung weiter ab
 - bzw. homogeneren Staaten reicht ein niedrigeres GDP für ein gleich niedriges Bürgerkriegsrisiko
 - Bei sehr armen Staaten nimmt das hohe Risiko eines neuen Bürgerkriegs mit zunehmender Homogenität weiter zu
- Ethnische Heterogenität scheint per se keine wichtige Determinante von Bürgerkriegen zu sein

Replikation Tabelle 1

- logit onset warl gdpenl lpopl lmtnest
ncontig Oil nwstate instab polity2l
ethfrac relfrac
- „note: 7 failures and 0 successes
completely determined“
 - für einige Kombinationen der
unabhängigen Variablen variiert die
abhängige Variable nicht
 - macht in extremeren Fällen Probleme,
da Effekte dann nicht schätzbar

Replikation Tabelle 1

- Welche Effekte sind signifikant?
 - warl gdpenl lpopl1 lmtnest Oil nwstate instab
 - Wenn wir dem Signifikanztest glauben, können alle anderen Effekte ignoriert werden
- Welches sind die stärksten dieser sieben Effekte?
 - Teil oder vollstandardisierte Effekte? – listcoef, std
 - Resultat, Interpretation?
 - Alternativen
 - (maximaler Effekt)
 - Quartilsabstand
 - summ warl gdpenl lpopl1 lmtnest Oil nwstate instab,det
 - Unabhängig von der Betrachtungsweise hat das logarithmierte Pro-Kopf-Einkommen den stärksten Einfluß

Replikation Tabelle 1

- Was bedeutet das substantiell?
 - prgen gdpenl, gen(p1) from(0.048) to(66.735) n(400) rest(median)
 - twoway line p1p1 p1x, xlabel(,grid) ylabel(,grid)
 - summ gdpenl, det
 - twoway line p1p1 p1x if p1x < 1, xlabel(,grid) ylabel(,grid)
 - tab country if gdpenl < 1

Replikation Tabelle 1

- Wie wirken sich andere wichtige Variablen aus?
 - Vergleich „glücklicher“: log der Bevölkerung = 8, log Berge = 1 (jeweils 25% Perzentil), nicht instabil, nicht neu
 - und „unglücklicher“ Staaten (instabil, neu, log Bevölkerung 9.9, log Berge 3.3 (75% Perzentil))
 - Rest jeweils wieder median
 - prgen gdpenl,gen(pg) from(0.048) to(66.735) n(400)
rest(median) x(lpopl1 8 lmtnest 1 instab 0 nwstate 0)
 - prgen gdpenl,gen(pu) from(0.048) to(66.735) n(400)
rest(median) x(lpopl1 9.9 lmtnest 3.3 instab 1 nwstate 1)
 - graph twoway (line pgp1 pgx) (line pup1 pux),xlabel(,grid)
ylabel(,grid)
 - graph twoway (line pgp1 pgx) (line pup1 pux) if
pux<10,xlabel(,grid) ylabel(,grid)
 - graph twoway (line pgp1 pgx) (line pup1 pux) if
pux<1,xlabel(,grid) ylabel(,grid)

Replikation Tabelle 1

- Teilweise höchst hypothetisch
- Multikollinearität?
 - Logit-Befehl zurückholen
 - reg onset war1 gdpent1 lpopl lmtnest ncontig Oil
nwstate instab polity2l ethfrac relfrac
 - vif
- Ausbruch von Bürgerkriege in reichen Ländern selbst unter ungünstigen Bedingungen fast ausgeschlossen
- (Probleme mit Staaten, die noch nicht lange existieren möglich)

Replikation Tabelle 1

- Wirklich kein Effekt ethnischer Fragmentation, auch nicht in Abhängigkeit vom Pro-Kopf-Einkommen?
 - summ ethfrac
 - gen zethfrac=ethfrac-r(mean)
 - summ gdpenl
 - gen zgdpenl=gdpenl-r(mean)
 - gen ethXgdp=zethfrac*zgdp
- logit onset warl zgdpenl lpopl lmtnest ncontig Oil nwstate instab polity2l zethfrac relfrac
 - Zentrieren hat keinen Einfluß auf den Effekt der Variablen selbst!
 - est store orig
 - fitstat,save
- logit onset warl zgdpenl lpopl lmtnest ncontig Oil nwstate instab polity2l zethfrac relfrac ethXgdp
 - lrtest orig
 - fitstat,dif

Replikation Tabelle 1

- Wird die Signifikanz der Effekte möglicherweise überschätzt, weil länderspezifische Störvarianzen ignoriert werden?
 - `logit onset warl zgdpenl lpopl lmtnest ncontig Oil nwstate instab polity2l zethfrac relfrac`
 - Option `cluster` ermöglicht gibt Annahme identischer und unabhängiger Störvarianzen auf
 - `logit onset warl zgdpenl lpopl lmtnest ncontig Oil nwstate instab polity2l zethfrac relfrac,cluster(ccode)`
- Aber: Kann hier überhaupt von Signifikanz etc. gesprochen werden?

Replikation Tabelle 1

- Geht der Einfluß des GDP/Kopf in Wirklichkeit auf einen Trend zurück
 - was wäre der Mechanismus?
 - wäre ein solcher Effekt überhaupt möglich?
 - corr year gdpenl
 - logit onset warl zgdpenl lpopl lmtnest ncontig Oil nwstate instab polity2l zethfrac relfrac year
- Welchen Einfluß haben die Faktoren darauf, ob zu einem bestimmten Zeitpunkt Bürgerkrieg *herrscht*?
 - logit war warl gdpenl lpopl lmtnest ncontig Oil nwstate instab polity2l ethfrac relfrac

Ergebnis

- Wie man es auch dreht und wendet, ist das (logarithmierte) Pro-Kopf-Einkommen der zentrale Prädiktor für den Ausbruch von Bürgerkriegen
 - in reichen Ländern sind Bürgerkriege extrem unwahrscheinlich
 - begünstigende Faktoren werden nur in ärmeren Ländern wirksam (implizite Interaktion im Logit-Modell)
- Ob dieser Prädiktor tatsächlich als Indikator für die Möglichkeiten eines Aufstands genutzt werden kann, ist fraglich
- Aber: die bivariaten Korrelationen mit Demokratie und ethnisch-religiöser Fragmentation, Instabilität etc sind gering: corr gdpenl polity2 ethfrac relfrac instab

Hausaufgabe

- Schreiben Sie eine Datei relig.do, die
 - den Datensatz öffnet
 - das Modell 2 für „ethnische“ Kriege schätzt (wie Modell 1, aber zweitgrößte ethnische Gruppe (Variable second) >4.9 Prozent)
 - Ein Grafik erzeugt, die für diesen Subsample den Einfluß der religiösen Fragmentation auf die Wahrscheinlichkeit eines neuen Bürgerkriegs plottet, getrennt für arme (gdpenl=0.878), mediane (gdpenl=1.8) und reiche (gdpenl=3.92) Staaten. Alle anderen Variablen werden auf ihren Median gesetzt
 - Bitte bis nächsten Mittwoch an do-files@politik.uni-mainz.de