Kapitel 11

Paneldaten

Vorbemerkungen

Wiederholte Beobachtungen der selben Querschnitts-Einheiten (*units*, z.B. Personen, Firmen, Länder, ...) über die Zeit.

Beispiel: 'marriage premium, marriage penalty'

$$\log(\text{wage})_i = \beta_1 + \beta_2 \text{married}_i + \varepsilon_i$$

Problem: Verheiratete haben andere unbeobachtete Charakteristika (OVB), z.B. 'social skills', Aussehen, ... $\Rightarrow$ $\text{cov}(\text{married}, \varepsilon) \neq 0$ $\Rightarrow$ OLS liefert verzerrte Ergebnisse!

Notation:

y_i	log(wage)
$y_i(0)$, $y_i(1)$	'Potential Outcome'
m_i	'married', $m_i \in \{0, 1\}$
x_i	beobachtete Regressor(en)
a_i	unbeobachtete Einflussfaktoren (z.B. 'ability', Aussehen, ...)

Conditional independence:

$$\text{E}(y_i(0)|a_i, x_i.m_i) = \text{E}(y_i(0)|a_i, x_i)$$

$\Rightarrow$ m_i ist *'as good as randomly assigned conditional on a_i and x_i'*

Wenn wir a_i beobachten könnten, könnten wir den Einfluss von 'married' auf log(wage) konsistent messen.

Falls a_i nicht beobachtet werden kann führt dies zu Endogenität (OVB).

Paneldaten: angenommen, wir haben in der Zeit t wiederholte Beobachtungen

y_{it}	log(wage) von Person i im Zeitpunkt t
$y_{it}(0)$, $y_{it}(1)$	'Potential Outcome'
m_{it}	Status von Person i im Zeitpunkt t, $m_{it} \in \{0, 1\}$
x_{it}	beobachtete Regressor(en) von Person i im Zeitpunkt t
a_i	unbeobachtete Einflussfaktoren; Annahme: *zeitinvariant*!
λ_t	Zeiteffekte

Annahmen:

1. a_i ist ein unbeobachteter *fixer* (d.h. zeitinvarianter) Einflussfaktor (Regressor)
2. Conditional independence: $\mathrm{E}(y_{it}(0)|a_i, x_{it}.m_{it}, t) = \mathrm{E}(y_{it}(0)|a_{it}, x_{it}, t)$
3. Linearität:
$$\mathrm{E}(y_{it}(0)|m_{it}, a_{it}, x_{it}, t) = \mu + \gamma a_i + \lambda_t + \beta_1 m_{it} + \beta_2 x_{it}$$

Also
$$y_{it} = \mu + \gamma a_i + \lambda_t + \beta_1 m_{it} + \beta_2 x_{it} + \varepsilon_{it}$$
Wir definieren
$$\mu_i := \mu + \gamma a_i$$
und erhalten
$$y_{it} = \mu_i + \lambda_t + \beta_1 m_{it} + \beta_2 x_{it} + \varepsilon_{it}$$
Im *fixed effects* Modell sind die Individuen-fixen Effekte μ_i und die Zeit-fixen Effekte λ_t Parameter betrachtet, die aus der Stichprobe geschätzt werden.
Die Schätzung kann auf 3 Arten erfolgen:

1. LSDV: Dummy Variablen für jedes Individuum und für jede Zeitperiode (abzüglich jeweils einer Referenzkategorie, um Dummyvariablenfalle zu vermeiden)
2. Fixed Effects (FE): individuenspezifische Mittelwertbereinigung; siehe Frisch-Waugh-Lovell Theorem.
3. First Differences (FD):
$$\Delta y_{it} = \Delta\lambda_t + \beta_1 \Delta m_{it} + \beta_2 \Delta x_{it} + \Delta\varepsilon_{it}$$

Alle drei Methoden ‘killen’ den unbeobachteten individuenspezifischen fixen Effekt!
Für zwei Zeitperioden führen alle drei Methoden zum gleichen Ergebnis, für mehr Zeitperioden nicht.
Mittelwertbereinigung: Die Mittelwerte über die Individuen sind
$$\bar{y}_i = \mu_i + \bar{\lambda} + \beta_1 \bar{m}_i + \beta_2 \bar{x}_i + \bar{\varepsilon}_i$$
und das mittelwerttransformierte Modell
$$\begin{aligned}(y_{it} - \bar{y}_i) \;\; &= \;\; \underbrace{(\mu_i - \mu_i)}_{=0} + (\lambda_t - \bar{\lambda}) + \\ & \quad\;\; \beta_1(m_{it} - \bar{m}_i) + \beta_2(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i\end{aligned}$$

Schätzung des FE Modells: diese Modelle können mit allen üblichen statistischen Programmen geschätzt werden.
R: siehe `https://www.uibk.ac.at/econometrics/einf/r_paneldaten.pdf`
Stata:
Dummy Variable Estimator: `xi: reg Y M X i.id i.year`
Fixed Effects Estimator: `xtreg Y M X, fe`
First Difference Estimator: `reg D.(Y M X)`

Interpretation des FE Modells:

- Alle drei Methoden verwenden nur den *within-unit change* zur Identifikation, d.h., die empirische *within-unit variation* ist für die Schätzung entscheidend (jedes Individuum dient gewissermaßen als seine eigene Kontrollgruppe).
- Nur wenn genügend *'switcher'* vorhanden sind ist der Effekt identifiziert!
- Es wird für jede (auch unbeobachtete) *zeitinvariante* Heterogenität kontrolliert!
- Für 'Omitted Variables', die sich über die Zeit ändern, wird dadurch *nicht* kontrolliert! Solche Heterogenität führt also weiterhin zu Endogenität.

  Zum Beispiel: was verursacht eine Scheidung? Positive und negative Schocks können sich auf den Lohn und die Scheidungswahrscheinlichkeit auswirken (z.B. Alkohol).

11.1 Panelmodelle in Matrixschreibweise

Im Folgenden werden wir die Querschnitts-Einheiten einfachheitshalber als 'Units' oder Individuen bezeichnen.

Notation:

y_{it} = der Wert der abhängigen Variable von Unit i zum Zeitpunkt t mit $i = 1, \dots N$ und $t = 1, \dots T$.
x_{kit} = Wert der k-ten erklärenden Variable für Unit i zum Zeitpunkt t.

Zeitreihendaten: $N = 1$, T groß.
Querschnittsdaten: N groß, $T = 1$

Häufigster Fall bei Paneldaten: großes N, kleines T
(für die Asymptotik wird häufig ein fixes T und $N \to \infty$ angenommen).

Das Modell für ein Individuum i:

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iN} \end{pmatrix}, \qquad \boldsymbol{X}_i = \begin{pmatrix} x_{1i1} & x_{2i1} & \cdots & x_{Ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{Ki1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{KiT} \end{pmatrix} \qquad \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

Balanced vs. Unbalanced Panels:

- *Balanced Panels:* für jede Unit i existieren gleich viele (T) Beobachtungen, d.h. insgesamt NT (N mal T) Beobachtungen.
- *Unbalanced Panels:* Die Anzahl der Beobachtungen ist unterschiedlich, z.B. fehlende Werte. Diese sind vor allem dann gefährlich, wenn Beobachtungen *nicht zufällig* fehlen (*selection bias*).

Die folgenden Ausführungen beschränken sich ausschließlich auf *"balanced panels"*, d.h. für jedes Individuum existieren T Beobachtungen. Die Ergebnisse gelten aber in den meisten Fällen (zmindest für statische Modelle) auch für *'unbalanced panels'*.

Meist werden die Daten nach Individuen geordnet 'übereinander angeordnet' in eine Matrix bzw. Vektor geschrieben, d.h. *'gestackt'*.

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{pmatrix}$$

wobei $\boldsymbol{y}$ und $\boldsymbol{\varepsilon}$ Spaltenvektoren mit der Dimension $NT \times 1$ sind und $\boldsymbol{X}$ die Dimension $NT \times K$ hat.

Das Modell kann in dieser Notation einfach geschrieben werden als

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

mit

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}$$

Grundidee Was dieses Modell von den bisherigen Modellen unterscheidet ist einzig das a priori Wissen, dass diese Daten wiederholte Beobachtungen der gleichen Individuen repräsentieren. Und dieses Wissen gilt es für die Schätzung zu nützen.

11.2 Vorteile von Panelschätzungen

Vorteile von Panels (Baltagi, 2005, Hsiao, 2003):

- Mehr Daten, d.h. genauere (effizientere) Schätzer und exaktere Inferenz.
- Berücksichtigung von Heterogenität und Messfehlern; Reduktion des *"missing variable bias"* durch fixe (zeitinvariante) Uniteffekte.
- Berücksichtigung von Einflüssen, die für alle Units gleich sind (Zeiteffekte).
- Dynamische Anpassung auf der Mikroebene.

11.3 Drei Modelle

Das Modell

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

sieht altvertraut aus, und ist es auch.

Die zentralen Gauss-Markov Annahmen, die für Erwartungstreue und Effizient erforderlich sind, sind

1. $\mathrm{E}(\varepsilon_{it}|\boldsymbol{X}) = 0$ (keine Endogenität)
2. $\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$ (keine Heteroskedastizität & Autokorrelation)

Aus der einführenden Statistik ist bekannt, dass bei Verletzung der ersten Annahme der OLS Schätzer für $\boldsymbol{\beta}$ verzerrt und auch nicht konsistent ist, während bei einer Verletzung der zweiten Annahme der der OLS Schätzer für $\boldsymbol{\beta}$ zwar unverzerrt und konsistent ist, aber nicht mehr effizient (die Standardfehler von $\boldsymbol{\beta}$ sind in diesem Fall auch verzerrt).

Für Paneldaten kann man sich vorstellen, dass der Störterm ε_{it} aus einer individuenspezifischen Komponente μ_i und dem Rest υ_{it} besteht

$$\varepsilon_{it} = \mu_i + \upsilon_{it}$$

Man beachte, dass der individuenspezifische Effekt μ_i keinen Zeitindex hat, in Worten, der individuenspezifische Effekt μ_i wird *zeitinvariant* angenommen. Die Vorstellung dahinter ist, dass sich bestimmte Eigenschaften der Individuen nicht in der Zeit ändern, wie z.B. das Geschlecht von Personen, die Gesellschaftsform von Unternehmen, oder kulturelle Eigenschaften von Regionen. Je nach den Annahmen über den (v.a. individuenspezifischen) Störterm kann man drei Modelle unterscheiden

Pooled Estimator: Im aller einfachsten Fall, wenn $\mathrm{E}(\varepsilon_{it}|\boldsymbol{X}) = 0$ und $\varepsilon_{it} \sim \mathrm{N}(0, \sigma^2)$, kann das 'gestackte' Modell $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ganz normal mit OLS geschätzt werden. Der gepoolte Schätzer ist einfach der OLS Schätzer

$$\hat{\boldsymbol{\beta}}^{\text{Pool}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

Aber wie wir gleich sehen werden sind diese beide Annahmen für Panel Daten ziemlich selten erfüllt.

Fixe Effekte (*'fixed effects model'*) dieses Modell erlaubt konsistente Parameterschätzungen selbst dann, wenn die *zeitinvarianten* Individueneffekte μ_i mit x_{it} korreliert sind! Dies ist einer der Gründe, warum dieses Modell bei Ökonometrikerinnen so beliebt ist. Im Kern kann man sich das einfach so vorstellen, dass für jedes Individuum eine Dummy eingeführt und mitgeschätzt wird, das Modell ist in der Struktur also ziemlich einfach.

Zufällige Effekte (*'random effects model'*) in diesem Modell wird – wie beim 'Pool' Modell – vorausgesetzt, dass die zeitinvarianten Individueneffekte mit den x_{it} *unkorreliert* sind, eine für ökonomische Zusammenhängen oft ziemlich wirklichkeitsfremde Annahme. Aber im Gegensatz zum Pool Modell wird die Panelstruktur der Daten berücksichtigt. Wir wissen nämlich, dass die wiederholte Beobachtung von N Individuen über T Perioden in der Regel nicht gleich viel Information enthält, wie die Beobachtung von NT unabhängigen Individuen! Der Grund ist einfach, die Streuung der Beobachtungen eines Individuums über die Zeit ist vermutlich kleiner als die Streuung zwischen den Individuen; oder in anderen Worten, zwei Beobachtungen, die von dem gleichen Individuum stammen, sind sich *wahrscheinlich* ähnlicher als zwei Beobachtungen, die von unterschiedlichen Individuen stammen. Diese Information gilt es zu nützen, und das *random effects model* nützt sie mit Hilfe der GLS ('generalized least squares') Methode.

Diese Modelle werden später ausführlich dargestellt, dies nur als vorab Information.

11.4 Varianzanalyse als Regression auf Unit-Dummies

Zur Einführung wollen wir einen besonders einfachen Fall untersuchen. Wir werden zeigen, dass eine einfache OLS Regression auf Individuendummies zum gleichen

Ergebnis führt wie eine Varianzanalyse. Das Nette daran ist, dass sich fast alle Ergebnisse daraus auf komplexere Modelle übertragen lassen.

In der einfachen Varianzanalyse wird untersucht, ob und wie sich die Erwartungswerte einer metrischen Zufallsvariable $\boldsymbol{y}$ in verschiedenen Gruppen (in der Panelanalyse 'Units' bzw. Individuen) unterscheiden. Getestet wird, ob die *Varianz zwischen den Gruppen* (Individuen) größer ist als die *Varianz innerhalb der Gruppen* (Individuen). Dadurch kann ermittelt werden, ob die Gruppeneinteilung sinnvoll ist oder nicht bzw. ob sich die Gruppen signifikant unterscheiden oder nicht.

Wir wollen die Mittelwerte einer Variable $\boldsymbol{y}$ von N 'Units' (Gruppen) vergleichen, und für jede Gruppe existieren T Beobachtungen, insgesamt stehen also NT Beobachtungen zur Verfügung.

Das Modell ist

$$y_{it} = \mu_i + \varepsilon_{it}$$

mit

$$\varepsilon_{it} \sim \text{i.i.d.}(0, \sigma^2)$$

Falls die üblichen Annahmen erfüllt sind, sind die Erwartungswerte von $\boldsymbol{y}$ die Gruppenmittelwerte

$$\begin{aligned} \mathrm{E}(y_{1t}) &= \mu_1 \\ \mathrm{E}(y_{2t}) &= \mu_2 \\ \vdots \quad & \quad \vdots \\ \mathrm{E}(y_{Nt}) &= \mu_N \end{aligned}$$

Alternativ dazu kann man auch ein Interzept einführen, aber um nicht in die Dummy-Variablenfalle zu tappen muss in diesem Fall eine Dummy Variable weggelassen werden. Das Modell mit Interzept α ist

$$y_{it} = \alpha + \mu_i + \varepsilon_{it}, \qquad i = 1, \ldots, N-1; \quad t = 1, \ldots, T$$

In diesem Fall ändert sich die Interpretation der Parameter, die μ_i messen den Unterschied zum weggelassenen Individuum. Wenn z.B. die Dummy für Unit N weggelassen wird ist

$$\begin{aligned} \mathrm{E}(y_{1t}) &= \alpha + \mu_1 \\ \mathrm{E}(y_{2t}) &= \alpha + \mu_2 \\ \vdots \quad & \quad \vdots \\ \mathrm{E}(y_{Nt}) &= \alpha \end{aligned}$$

deshalb ist $\mu_1 = \mathrm{E}(y_{1t}) - \mathrm{E}(y_{1N})$, $\mu_2 = \mathrm{E}(y_{2t}) - \mathrm{E}(y_{1N})$, usw.

Welche der beiden Versionen man wählt spielt in diesem Zusammenhang keine große Rolle, deshalb wählen wird die einfachere Methode ohne Interzept.

Wir werden nun anhand eines sehr einfachen Beispiels zeigen, dass mit Hilfe einer einfachen OLS Regression auf Gruppen- bzw. Individuendummies eine Zerlegung der Varianz in *'within'* und *'between'* Effekte erfolgen kann. Dies geschieht mit Hilfe der Projektionsmatrix $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ und der residuenerzeugenden Matrix $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$, die wir bereits im Kapitel zur Matrixalgebra eingeführt haben.

Wie gesagt beschränken wir uns nun auf den allereinfachsten Fall mit drei Individuen, die zu zwei Zeitpunkten beobachtet werden, d.h. $N = 3$ und $T = 2$, d.h. $NT = 6$. Die Designmatrix $\boldsymbol{X}$ enthält die 'gestackten' Individuendummies.

$$\boldsymbol{y} = \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix}, \qquad \boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

Hier haben wir für jedes Individuum eine Dummy Variable, aber dafür kein Interzept um die Dummy-Variablen Falle zu vermeiden. Alternativ dazu könnte die $\boldsymbol{X}$ Matrix ein Interzept *anstelle* einer Dummy Variablen enthalten, dies hat keine inhaltlichen Auswirkungen auf die folgenden Ausführungen.

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Für T Beochactungen pro Individuum hätte diese Matrix die Dimension $T \times T$, und auf der Hauptdiagonale der Inversen (unten) würde entsprechend auf der Hauptdiagonale $1/T$ stehen.

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \qquad \boldsymbol{X}' = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

Der OLS Schätzer für $\boldsymbol{\mu}$ ist

$$\begin{aligned} \hat{\boldsymbol{\mu}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} &= \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} \\ &= \begin{pmatrix} 0.5(y_{11} + y_{12}) \\ 0.5(y_{21} + y_{22}) \\ 0.5(y_{31} + y_{32}) \end{pmatrix} = \begin{pmatrix} \bar{y}_{1.} \\ \bar{y}_{2.} \\ \bar{y}_{3.} \end{pmatrix} \end{aligned}$$

wobei

$$\bar{y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

Wir erinnern uns, dass eine OLS Regression auf das Interzept den Mittelwert lieferte. Wenn man Paneldaten auf eine Matrix von Individuendummies regressiert erhält man analog dazu die Mittelwerte über die Zeit für die einzelnen Individuen.

Projektionsmatrix $\boldsymbol{P}$: wie wir schon im Kapitel zur Matrixalgebta gezeigt haben liefert die Vormultiplikation eines Vektors $\boldsymbol{y}$ mit der Projektionsmatrix die gefitteten Werte $\widehat{\boldsymbol{y}}$. Dies funktioniert natürlich auch hier, wie man einfach sehen kann

$$\begin{aligned}
\boldsymbol{P} &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix} \\
&= \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{P}\boldsymbol{y} &= \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} \\
&= \begin{pmatrix} 0.5(y_{11}+y_{12}) \\ 0.5(y_{21}+y_{22}) \\ 0.5(y_{31}+y_{32}) \end{pmatrix} = \begin{pmatrix} \bar{y}_{1.} \\ \bar{y}_{1.} \\ \bar{y}_{2.} \\ \bar{y}_{2.} \\ \bar{y}_{3.} \\ \bar{y}_{3.} \end{pmatrix}
\end{aligned}$$

Vormultiplikation mit dieser Projektionsmatrix $\boldsymbol{P}$ liefert also tatsächlich die Unit-spezifischen Mittelwerte über die Zeit.

Die residuenerzeugende Matrix $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{P})$ Wie der Name schon sagt erzeugt die Vormultiplikation von $\boldsymbol{y}$ mit der Matrix $\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{P})$ die Residuen, d.h. in diesem Fall, die Abweichungen von den Unit-spezifischen Mittelwerten über die Zeit.

$$
\begin{aligned}
\boldsymbol{M} &= \boldsymbol{I} - \boldsymbol{P} \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix} \\
&= \begin{pmatrix} 0.5 & -0.5 & 0 & 0 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & 0 & 0 & -0.5 & 0.5 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{My} &= \begin{pmatrix} 0.5 & -0.5 & 0 & 0 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & -0.5 & 0 & 0 \\ 0 & 0 & -0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & 0 & 0 & -0.5 & 0.5 \end{pmatrix} \begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} \\
&= \begin{pmatrix} 0.5(y_{11} - y_{12}) \\ 0.5(y_{11} - y_{12}) \\ 0.5(y_{21} - y_{22}) \\ 0.5(y_{21} - y_{22}) \\ 0.5(y_{31} - y_{32}) \\ 0.5(y_{31} - y_{32}) \end{pmatrix} = \begin{pmatrix} y_{11} - \bar{y}_{1.} \\ y_{12} - \bar{y}_{1.} \\ y_{21} - \bar{y}_{2.} \\ y_{22} - \bar{y}_{2.} \\ y_{31} - \bar{y}_{3.} \\ y_{32} - \bar{y}_{3.} \end{pmatrix}
\end{aligned}
$$

Der letzte Schritt folgt aus der Definition von $\bar{y}_{i.} = 1/T(y_{i1} + y_{i2})$, mit $T = 2$, denn daraus folgt $y_{i2} = 2\bar{y}_{i.} - y_{i1}$. Einsetzen in $0.5(y_{i1} - y_{i2})$ gibt $0.5[y_{i1} - (2\bar{y}_{i.} - y_{i1})] = y_{i1} - \bar{y}_{i.}$.

Die Vormultiplikation von $\boldsymbol{y}$ mit der residuenerzeugenden Matrix $\boldsymbol{M}$ liefert also die *Abweichungen vom individuenspezifischen Mittelwert*, $y_{it} - \bar{y}_{i.}$.

Damit können wir die Streuung innerhalb (*'within'*) der Individuen (bzw. Gruppen) berechnen.

Schließlich wollen wir noch die Gesamtmittelwerte und die Abweichungen davon berechnen. Dies kann wiederum mit Hilfe einer geeigenten Projektions- bzw. residuenerzeugenden Matrix erfolgen.

11.4.1 Projektionsmatrix für das Gesamtmittel

Sei $\boldsymbol{J}$ eine $NT \times NT$ Einsenmatrix

$$\boldsymbol{J} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

man beachte, dass diese Einsenmatrix auch als das äußeres Produkt eines Einsenvektores dargestellt werden kann. Wenn $\boldsymbol{\iota}_{(NT)}$ einen $NT \times 1$ Einsenvektor (ein Spaltenvektor mit NT Zeilen, gefüllt mit lauter 1) bezeichnet, dann kann man für $\boldsymbol{J}$ auch $\boldsymbol{\iota}_{(NT)}\boldsymbol{\iota}'_{(NT)}$ (oder einfacher $\boldsymbol{\iota}\boldsymbol{\iota}'$) schreiben.

Die Matrix

$$\frac{1}{NT}\boldsymbol{J} = \frac{1}{NT}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

ist eine Projektionsmatrix für das Gesamtmittel, denn

$$\frac{1}{NT}\boldsymbol{J}\boldsymbol{y} = \frac{1}{NT}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix} = \begin{pmatrix} \bar{\bar{y}}_{..} \\ \bar{\bar{y}}_{..} \\ \bar{\bar{y}}_{..} \\ \bar{\bar{y}}_{..} \\ \bar{\bar{y}}_{..} \\ \bar{\bar{y}}_{..} \end{pmatrix}$$

wobei wir mit $\bar{\bar{y}}_{..}$ den Gesamtmittelwert (d.h. über Zeit und Individuen) aller Beobachtungen bezeichnen

$$\bar{\bar{y}}_{..} := \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$$

Die Abweichungen vom Gesamtmittel sind $y_{it} - \bar{\bar{y}}_{..}$, oder wieder in Matrixschreibweise

$$\boldsymbol{y} - \bar{\bar{\boldsymbol{y}}}_{..} = \boldsymbol{y} - \frac{1}{NT}\boldsymbol{J}\boldsymbol{y} = \left(\boldsymbol{I} - \frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}$$

wobei die residuenerzeugende Matrix $(\boldsymbol{I} - (1/NT)\,\boldsymbol{J})$ natürlich wieder idempotent ist.

Schließlich können wir die Abweichungen der Individuen-spezifischen Mittelwerte von dem Gesamtmittelwert berechnen. Wie oben gezeigt erhält man die Individuen-spezifischen Mittelwerte aus $\boldsymbol{P}\boldsymbol{y}$, und den Gesamtmittelwert aus $1/NT\,\boldsymbol{J}\boldsymbol{y}$, deshalb ist

$$\left(\boldsymbol{P} - \frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y} = \begin{pmatrix} \bar{y}_{1.} - \bar{\bar{y}}_{..} \\ \bar{y}_{1.} - \bar{\bar{y}}_{..} \\ \bar{y}_{2.} - \bar{\bar{y}}_{..} \\ \bar{y}_{2.} - \bar{\bar{y}}_{..} \\ \bar{y}_{3.} - \bar{\bar{y}}_{..} \\ \bar{y}_{3.} - \bar{\bar{y}}_{..} \end{pmatrix}$$

11.4.2 Allgemeine Eigenschaften von Projektions- und residuenerzeugenden Matrizen

(diese Eigenschaften gelten allgemein, nicht nur für dieses Beispiel)

1. Projektionsmatrizen und residuenerzeugende Matrizen sind symmetrisch und idempotent, d.h. $\boldsymbol{P} = \boldsymbol{P}'$, $\boldsymbol{PP} = \boldsymbol{P}$, $\boldsymbol{M} = \boldsymbol{M}'$ und $\boldsymbol{MM} = \boldsymbol{M}$.

   Zum Beispiel

$$\frac{1}{NT}\boldsymbol{J}\frac{1}{NT}\boldsymbol{J} = \frac{1}{NT^2}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

$$= \frac{1}{NT^2}\begin{pmatrix} NT & \cdots & NT \\ \vdots & \ddots & \vdots \\ NT & \cdots & NT \end{pmatrix} = \frac{1}{NT}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} = \frac{1}{NT}\boldsymbol{J}$$

   oder

$$\boldsymbol{PP} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$$

   oder

$$\boldsymbol{MM} = (\boldsymbol{I} - \boldsymbol{P})(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{I} - \boldsymbol{P} - \boldsymbol{P} + \boldsymbol{PP} = \boldsymbol{I} - \boldsymbol{P} = \boldsymbol{M}$$

2. Der Rang symmetrischer idempotenter Matrizen ist gleich ihrer Spur.

3. $\boldsymbol{P}$ und $\boldsymbol{M}$ sind orthogonal: $\boldsymbol{PM} = \boldsymbol{P}(\boldsymbol{I} - \boldsymbol{P}) = \boldsymbol{PI} - \boldsymbol{PP} = \boldsymbol{P} - \boldsymbol{P} = \boldsymbol{0}$.

4. $\boldsymbol{P} + \boldsymbol{M} = \boldsymbol{I}$, da $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$.

Im nächsten Schritt werden wir die Quadratsummen der einzelnen Abweichungen bilden. Dazu benötigen wir noch ein Resultat, nämlich die Idempotenz von

$$\left(\boldsymbol{P} - \frac{1}{NT}\boldsymbol{J}\right)\left(\boldsymbol{P} - \frac{1}{NT}\boldsymbol{J}\right) = \left(\boldsymbol{P} - \frac{1}{NT}\boldsymbol{J}\right)$$

Für unser Beispiel

$$\frac{1}{NT}\boldsymbol{PJ} = \frac{1}{NT}\begin{pmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$= \frac{1}{NT}\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \frac{1}{NT}\boldsymbol{J}$$

11.4.3 Varianzzerlegung

Mit Hilfe dieser Resultate können wir nun die gesamte Streuung in eine *'between'* und *'within'* Komponente zerlegen:

Gesamtstreuung SST Die Quadratsumme der Abweichungen vom Gesamtmittel sind $\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{\bar{y}}_{..})^2$, oder in Matrixschreibweise

$$\boldsymbol{y}'\left(\boldsymbol{I}-\frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}$$

***'Between'* Streuung SSB:** Die Quadratsumme der Abweichungen der Unit-Mittelwerte von dem Gesamtmittelwert sind $\sum_{i=1}^{N}(\bar{y}_{i.}-\bar{\bar{y}}_{..})^2$

$$\boldsymbol{y}'\left(\boldsymbol{P}-\frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}$$

***'Within'* Streuung SSW:** Die Quadratsumme der Abweichungen Beobachtungen von den Unit-Mittelwerten: $\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y}_{i.})^2$

$$\boldsymbol{y}'\left(\boldsymbol{I}-\boldsymbol{P}\right)\boldsymbol{y}$$

Aus der Varianzanalyse (ANOVA) ist bekannt, dass

$$\boldsymbol{y}'\left(\boldsymbol{I}-\frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}=\boldsymbol{y}'\left(\boldsymbol{P}-\frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}+\boldsymbol{y}'\left(\boldsymbol{I}-\boldsymbol{P}\right)\boldsymbol{y}$$

Daraus ergibt sich folgende ANOVA Tafel:

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	$N-1$	$\text{SSB}=\boldsymbol{y}'\left(\boldsymbol{I}-\frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}$	$\text{SSB}/(N-1)$
Within	$NT-N$	$\text{SSW}=\boldsymbol{y}'\left(\boldsymbol{P}-\frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}$	$\text{SSW}/(NT-N)$
Total	$NT-1$	$\text{SST}=\boldsymbol{y}'\left(\boldsymbol{I}-\frac{1}{NT}\boldsymbol{J}\right)\boldsymbol{y}$	$\text{SST}/(NT-1)$

Damit kann die gemeinsame Signifikanz der Unit-Effekte (d.h. in diesem Fall der Individuendummies) mit dem üblichen F-Test getestet werden:

$$\text{ANOVA}\quad F\text{-Test}=\frac{\text{SSB}/(N-1)}{\text{SSW}/(NT-N)}\sim F_{N-1,NT-N}$$

11.5 ‘*Between*’ und ‘*Within*’ Schätzer

Siehe Greene (2007, p. 191), Johnston and Dinardo (1996, p. 392)

Kehren wir zurück zur klassischen Panelanalyse. Im vorhergehenden Beispiel haben wir angenommen, dass die Designmatrix $\boldsymbol{X}$ nur Individuendummies enthält. Die Matrix $\boldsymbol{X}$ soll wieder aus den üblichen erklärenden Variablen bestehen.

Das Modell kann auch folgendermaßen geschrieben werden[1]

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

wobei $\boldsymbol{x}_{it}$ ein $K \times 1$ Spaltenvektor ist (der auch ein Interzept beinhalten kann), oder in gestackter Form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Die Fehlerstruktur sein

$$\varepsilon_{it} = \mu_i + \upsilon_{it}$$

wobei μ_i ein fixer oder zufälliger zeitinvarianter individueller Effekt ist. Außerdem sei

- $\mathrm{E}(\upsilon_{it}|\boldsymbol{X}) = 0$, und
- $\upsilon_{it} \sim$ i.i.d.$(0, \sigma^2_\upsilon)$

Der Pool-Schätzer für $\boldsymbol{\beta}$ ist

$$\hat{\boldsymbol{\beta}}^{\text{Pool}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

dieser Schätzer unterstellt aber die weit restriktiveren Annahmen $\mathrm{E}(\varepsilon_{it}|\boldsymbol{X}) = 0$ und $\varepsilon_{it} \sim$ i.i.d.$(0, \sigma^2_\varepsilon)$.

Bevor wir aber nach geeigneteren Schätzern suchen wollen wir den Faden von der Varianzanalyse noch einmal aufgreifen und zeigen, dass die Idee der Varianzzerlegung auch hier anwendbar ist. Zuerst erinnern wir uns, dass die Eigenschaften der Projektions- und residuenerzeugenden Matrizen allgemein gelten. Deshalb können wir diese Matrizen auch hier anwenden, um die gewünschte Datentransformation durchzuführen.

Zuerst könnten wir *nur* die Streuung zwischen den Individuen für die Schätzung zu nützen. Das können wir einfach erreichen, indem wir für jedes Individuum die Mittelwerte über die Zeit berechnen, und eine Regression auf diese N Individuen-Mittelwerte rechnen

$$\bar{y}_{i.} = \bar{\boldsymbol{x}}'_{i.}\boldsymbol{\beta} + \bar{\varepsilon}_{i.}$$

---

[1] Man beachte, dass dies einer strengere Restriktion ist als z.B.

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta}_i + \varepsilon_{it}$$

mit

$$\bar{y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$

Die Berechnung der individuellen Mittelwerte kann auch mit Hilfe einer Projektionsmatrix erfolgen. Dazu bilden wir wieder für jedes Individuum eine Dummy (bzw. $N-1$ Dummies falls die Regression ein Interzept enthält), und diese Matrix wieder wie früher 'gestackt'. Diese $NT \times N$ Matrix mit den Individuendummies nennen wir $\boldsymbol{D}$.

Die Projektionsmatrix mit den Individuendummies ist

$$\boldsymbol{P}_D = \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$$

Vormultiplikation von $\boldsymbol{y}$ und $\boldsymbol{X}$ mit $\boldsymbol{P}_D$ transformiert die Variablen – wie früher – in individuenspezifischen Mittelwerte

$$\boldsymbol{P}_D\boldsymbol{y} = \boldsymbol{P}_D\boldsymbol{X} + \boldsymbol{P}_D\boldsymbol{\varepsilon}$$

Diese Transformation liefert $\bar{y}_{i.} = \bar{\boldsymbol{x}}'_{i.}\boldsymbol{\beta} + \bar{\varepsilon}_{i.}$.

Den *'between'* Schätzer erhält man, wenn man den OLS Schätzer auf dies transformierten Daten anwendet

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{\text{between}} &= [(\boldsymbol{P}_D\boldsymbol{X})'\boldsymbol{P}_D\boldsymbol{X}]^{-1}(\boldsymbol{P}_D\boldsymbol{X})'\boldsymbol{y} \\
&= [\boldsymbol{X}'\boldsymbol{P}'_D\boldsymbol{P}_D\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{P}'_D\boldsymbol{y} \\
&= (\boldsymbol{X}'\boldsymbol{P}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}_D\boldsymbol{y}
\end{aligned}$$

da $\boldsymbol{P}_D$ symmetrisch und idempotent ist. Man beachte, dass diese Schätzung nur auf N Beobachtungen (den Individuenmittelwerten) beruht.

Analog dazu kann man mit Hilfe der residuenerzeugenden Matrix $\boldsymbol{M}_D = \boldsymbol{I} - \boldsymbol{P}_D$ die Abweichungen von den Individuen-spezifischen Mittelwerten berechnen

$$y_{it} - \bar{y}_{i.} = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_{i.})'\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_{i.}$$

und auf Grundlage dieser transformierten Daten den *'within'* Schätzer berechnen.

$$\boldsymbol{M}_D\boldsymbol{y} = \boldsymbol{M}_D\boldsymbol{X} + \boldsymbol{M}_D\boldsymbol{\varepsilon}$$

woraus durch Anwendung des OLS Schätzers der *'within'* Schätzer folgt

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{\text{within}} &= [(\boldsymbol{M}_D\boldsymbol{X})'\boldsymbol{M}_D\boldsymbol{X}]^{-1}(\boldsymbol{M}_D\boldsymbol{X})'\boldsymbol{y} \\
&= [\boldsymbol{X}'\boldsymbol{M}'_D\boldsymbol{M}_D\boldsymbol{X}]^{-1}\boldsymbol{X}'\boldsymbol{M}'_D\boldsymbol{y} \\
&= (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}
\end{aligned}$$

da $\boldsymbol{M}_D$ symmetrisch und idempotent ist.

Dieser *'within'* Schätzer ist übrigens ein ganz normaler *'fixed effects'* Schätzer, den wir gleich ausführlicher darstellen werden.

Alle drei Modelle – das gepoolte, *'between'* und *'within'* Modell – können prinzipiell konsistent, aber nicht notwendigerweise effizient mit OLS geschätzt werden.

Nebenbei bemerkt, der gepoolte OLS Schätzer ist ein gewichteter Mittelwert von *'between'* und *'within'* Schätzer, denn

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}^{\text{Pool}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y} + \boldsymbol{X}'\boldsymbol{P}_D\boldsymbol{y}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{within}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}_D\boldsymbol{X}\hat{\boldsymbol{\beta}}^{\text{between}}
\end{aligned}
$$

Dieser Schätzer ist zwar konsistent, aber OLS berücksichtigt nicht die Information, dass es sich um wiederholte Beobachtungen der gleichen Individuen handelt. Deshalb ist dieser Schätzer in der Regel nicht effizient!

11.6 Das Modell mit zufälligen Effekten *(Random Effects Model)*

Das Problem beim gepoolten OLS Schätzer besteht – wie schon mehrfach erwähnt – darin, dass die spezielle Fehlerstruktur von Paneldaten nicht berücksichtigt wird, denn wahrscheinlich werden sich die Varianzen *'within'* der Individuen von den Varianzen *'between'* den Individuen unterscheiden. Dies legt die Anwendung eines GLS Schätzers nahe, der diese Information nützt, und genau darum handelt es sich beim *'random effects'* Modell.

Das Modell ist

$$
\begin{aligned}
y_{it} &= \boldsymbol{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} \\
\varepsilon_{it} &= \mu_i + \upsilon_{it}
\end{aligned}
$$

Der Störterm ε_{it} enthält wieder zwei Komponenten, einen zeitinvarianten individuenspezifischen Fehler μ_i und den Rest υ_{it}.

Der wichtigste Unterschied zum *'fixed effects'* Modell besteht in der *Annahme*, dass der zeitinvariante individuenspezifische Fehler μ_i *nicht* mit den x-Variablen korreliert ist.

Die dahinter liegende Vorstellung ist, dass eine Stichprobe aus einer sehr großen Grundgesamtheit gezogen wird, und dass *'omitted variables'* kein Problem darstellen.

Die konkreten Annahmen über den Störterm sind

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{\upsilon}) &= 0 \\
\mathrm{E}(\boldsymbol{\upsilon}\boldsymbol{\upsilon}') &= \sigma_\upsilon^2 \\
\mathrm{E}(\mu_i) &= 0 \\
\mathrm{E}(\mu_i\mu_i) &= \sigma_\mu^2 \\
\mathrm{E}(\mu_i\mu_j) &= 0 \quad \text{für } i \neq j \\
\mathrm{E}(\mu_i\upsilon_j t) &= 0
\end{aligned}
$$

(alle Erwartungen bedingt auf $\boldsymbol{X}$)

Das bedeutet, es wird homoskedastische Varianz und Equi-Korrelation innerhalb der Individuen angenommen

$$\text{cov}(\varepsilon_{it}\varepsilon_{js}) = \begin{cases} \sigma_\mu^2 + \sigma_\upsilon^2 & \text{für } i = j \text{ und } t = s \\ \sigma_\mu^2, & \text{für } i = j \text{ und } t \neq s \\ 0, & \text{für } i \neq j \text{ und } t \neq s \end{cases}$$

Die $T \times T$ Varianz-Kovarianzmatrix für ein einzelnes Individuum ist

$$\boldsymbol{\Sigma}_i = \text{E}(\boldsymbol{\mu}_i \boldsymbol{\mu}_i') = \sigma_\upsilon^2 \boldsymbol{I}_T + \sigma_\mu^2 \boldsymbol{\iota\iota'} = \begin{pmatrix} \sigma_\mu^2 + \sigma_\upsilon^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_\upsilon^2 & \dots & \sigma_\mu^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 + \sigma_\upsilon^2 \end{pmatrix}$$

wobei $\boldsymbol{\iota}$ ein $T \times 1$ Einsenvektor ist. Dies ist die Varianz-Kovarianzmatrix für ein einzelnes Individuum, für das gestackte Gesamtmodell ergibt sich damit eine $NT \times NT$ Varianz-Kovarianzmatrix

$$\boldsymbol{\Omega} = \text{E}(\boldsymbol{\mu\mu'}) = \boldsymbol{I}_N \otimes \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{pmatrix}$$

Der GLS Schätzer ist

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$$

Für erforderliche Datentransformation benötigen wir $\boldsymbol{\Omega}^{-1/2} = (\boldsymbol{I} \otimes \boldsymbol{\Sigma})^{-1/2}$.

$\boldsymbol{\Omega}$ hat die Dimension $NT \times NT$, aber die blockdiagonale Struktur erlaubt die Berechnung der Inversen mit Hilfe des Wansbeek, Kapteyn (1982, 1983)-Tricks.

Man kann zeigen, dass

$$\boldsymbol{\Sigma}^{-1/2} = \frac{1}{\sigma_\upsilon}\left[\boldsymbol{I}_T - \left(\frac{1-\theta}{T}\boldsymbol{\iota\iota'}\right)\right]$$

mit

$$\theta = \sqrt{\frac{\sigma_\upsilon^2}{T\sigma_\mu^2} + \sigma_\upsilon^2}$$

Feasible GLS: σ_μ^2 und σ_υ^2 sind unbekannte Parameter und müssen geschätzt werden.

- Wallace, Hussein (1969): OLS Residuen
- Amemiya (1971): LSDV-Residuen
- Swamy, Arora (1972): Within- und between Schätzer.

Johnston and Dinardo (1996, p. 394)

$$\begin{aligned} s_v^2 &= \frac{\mathbf{e}_w'\mathbf{e}_w}{NT - NK - N} \\ s_b^2 &= \frac{\mathbf{e}_b'\mathbf{e}_b}{N - K} \\ s_\mu^2 &= s_b^2 - \frac{s_v^2}{T} \end{aligned}$$

wobei $\mathbf{e}_w$ die Residuen der *'within'* Regression und $\mathbf{e}_b$ die Residuen der *'between'* Regression sind. Der GLS Schätzer für $\boldsymbol{\beta}$ ist in diesem Fall effizienter als der OLS Schätzer, aber der OLS Schätzer ist ebenfalls konsistent.

11.7 Das Modell mit fixen Effekten *(Fixed Effects Model)*

Die große Beliebtheit, derer sich Panelmodelle unter den ÖkonometrikerInnen erfreuen, beruht zu einem guten Teil aus den Eigenschaften des *'fixed effects'* Modells. Bisher schien das Panelmodell eine etwas 'aufgeblasene' Version des altbekannten OLS (bzw. GLS) Modells, das zwar eine etwas genauere Schätzung der Parameter erlaubt, aber sonst wenig Neues zu bieten hat.

Die Popularität des *'fixed effects'* Modells beruht vor allem darauf, dass es unter bestimmten Umständen eine Lösung für eines der größeren Probleme von ÖkonometrikerInnnen verspricht, das Problem unbeobachtbarer oder fehlender Variablen. Aber beginnen wir von vorne.

Für eine erste intuitive Erklärung beginnen wir wieder mit einem sehr einfachen Modell, mit einem Modell für zwei Zeitperioden ($t = 1, 2$)

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_{it}$$

oder in 'gestackter' Matrixschreibweise

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

wobei $\boldsymbol{X}$ eine Matrix ist, deren Werte über die Zeit und die Individuen streuen, und $\boldsymbol{Z}$ ist eine Matrix mit Variablen, die *nur zwischen* den Individuen streuen, aber 'innerhalb' der Individuen gleich bleiben (z.B. das Geschlecht bei Personen). Solche Variablen nennen wir *zeitinvariant.*

Für den Störterm gelten im wesentlichen die gleichen Annahmen wie für das *'random effects'* Modell

$$\varepsilon_{it} = \mu_i + \upsilon_{it}$$

mit

$$\begin{aligned} \mathrm{E}(\boldsymbol{\upsilon}) &= 0 \\ \mathrm{E}(\boldsymbol{\upsilon}\boldsymbol{\upsilon}') &= \sigma_\upsilon^2 \\ \mathrm{E}(\mu_i) &= 0 \\ \mathrm{E}(\mu_i\mu_i) &= \sigma_\mu^2 \\ \mathrm{E}(\mu_i\mu_j) &= 0 \quad \text{für } i \neq j \\ \mathrm{E}(\mu_i\upsilon_j t) &= 0 \end{aligned}$$

Das bedeutet, es wird homoskedastische Varianz und Equi-Korrelation innerhalb der Individuen angenommen (alle Erwartungen wieder bedingt auf $\boldsymbol{X}$).

Der große Unterschied zum *'random effects'* Modell besteht in der *Annahme* über die Korrelation des zeitinvarianten individuenspezifischen Fehlers μ_i mit den unabhängigen Variablen, das *'fixed effects'* Modell erlaubt nämlich eine solche Korrelation! Wenn wir aus $\boldsymbol{X}$ und $\boldsymbol{Z}$ eine Matrix $\boldsymbol{W}$ bilden, d.h. $\boldsymbol{W} = [\boldsymbol{X}\ \boldsymbol{Z}]$, erlaubt das *'fixed effects'* Modell

$$\mathrm{E}(\boldsymbol{W}'_{it}\boldsymbol{\varepsilon}_{it}) \neq 0$$

Zu befürchten ist vor allem, dass die zeitinvarianten $\boldsymbol{z}'_i$ mit den individuenspezifischen Störtermen μ_i korreliert sind.

Wie soll das funktionieren, wir wissen doch, dass eine Korrelation zwischen erklärenden Variablen und Störtermen immer zu verzerrten Schätzern führt?

Nun, wenn wir die Gleichung *nur* für die erste oder nur für die zweite Periode schätzen stimmt dies auch, sowohl

$$y_{i1} = \boldsymbol{x}'_{i1}\boldsymbol{\beta} + \boldsymbol{z}'_i\boldsymbol{\delta} + \varepsilon_{i1}$$

als auch

$$y_{i2} = \boldsymbol{x}'_{i2}\boldsymbol{\beta} + \boldsymbol{z}'_i\boldsymbol{\delta} + \varepsilon_{i2}$$

liefern verzerrte Schätzungen.

Aber wenn diese beiden Gleichungen eine zutreffende Beschreibung des Zusammenhangs liefern, dann gilt dies auch für jede Linearkombination dieser Gleichungen. Wir könnten z.B. die Differenz dieser beiden Gleichungen betrachten

$$\begin{aligned}
y_{i1} &= \boldsymbol{x}'_{i1}\boldsymbol{\beta} + \boldsymbol{z}'_i\boldsymbol{\delta} + \varepsilon_{i1} \\
y_{i2} &= \boldsymbol{x}'_{i2}\boldsymbol{\beta} + \boldsymbol{z}'_i\boldsymbol{\delta} + \varepsilon_{i2} \\
y_{i2} - y_{i1} &= (\boldsymbol{x}'_{i2} - \boldsymbol{x}'_{i1})\boldsymbol{\beta} + (\boldsymbol{z}'_i - \boldsymbol{z}'_i)\boldsymbol{\delta} + \varepsilon_{i2} - \varepsilon_{i1}
\end{aligned}$$

Wenn $\boldsymbol{z}'_i$ in beiden Perioden gleich ist, also zeitinvariant ist, dann fällt $\boldsymbol{z}'_i$ bei der Differenzenbildung heraus! Unter Verwendung des Differenzenoperators Δ ist also

$$\Delta\boldsymbol{y} = \Delta\boldsymbol{X}\boldsymbol{\beta} + \Delta\boldsymbol{v}$$

Damit ist aber auch eine allfällige Korrelation zwischen $\boldsymbol{z}_i$ und μ_i beseitigt, und darin liegt der riesengroße Vorteil von Paneldaten.

Für das transformierte Modell muss nur gelten

$$\mathrm{E}(\Delta\boldsymbol{X}\,\Delta\boldsymbol{v}) = 0$$

Die Differenzenbildung ist nur eine mögliche Datentransformation, die diese Korrelation mit dem zeitinvarianten Individuenstörterm μ_i beseitigt. Jede der möglichen Transformation führt zu einem *'fixed effects'* Schätzer.

Tatsächlich kann man zeigen, dass diese Differenzenbildung zu Problemen mit dem transformierten Störterm führt (der folgt einem MA Prozess), was die Anwendung eines GLS Schätzers erforderlich macht.

Deshalb wird meist eine andere Transformation angewandt, nämlich die *'within'* Transformation, die wir bereits kennengelernt haben. Tatsächlich ist der *'within'* Schätzer ein möglicher *'fixed effects'* Schätzer. Die verschiedenen *'fixed effects'* Schätzer müssen nicht alle zum gleichen numerischen Ergebnis führen, aber große Unterschiede zwischen verschiedenen Schätzern sollten zumindest skeptisch stimmen.

Alle *'fixed effects'* Schätzer teilen einige wichtige Eigenschaften:

1. **Da durch die Transformation alle zeitinvarianten Variablen herausfallen, können mit dem *'fixed effects'* Modell keine Koeffizienten für zeitinvariante Variablen geschätzt werden!** Wenn es z.B. gilt die Einkommensunterschiede zwischen Männern und Frauen zu schätzen ist *'fixed effects'* Modell in der üblichen Form ungeeignet, da das Geschlecht (fast immer) eine zeitinvariante Eigenschaft ist.

2. Der große Vorteil des *'fixed effects'* Modells besteht darin, dass dieser Schätzer unempfindlich gegenüber einer Korrelation von zeitinvarianten Variablen mit dem Störterm ist. Da bei der Transformation alle zeitinvarianten Variablen herausfallen, ist **der *'fixed effects'* Schätzer unempfindlich gegenüber der Nicht-Berücksichtigung zeitinvarianter Variablen** (*'omitted variables bias'*). Zum Beispiel sind in Länderquerschnittsregressionen Faktoren wie Kultur, Geschichte, Religion usw. schwer zu messen. Aber die schätzbaren Parameter eines *'fixed effects'* Modells werden durch die Nicht-Berücksichtigung solcher Variablen nicht verzerrt geschätzt.

3. Wann immer die Annahmen des *'random effects'* Modells erfüllt sind, wird auch des *'fixed effects'* Modell konsistente Ergebnisse liefern. Allerdings wird in diesem Fall der *'fixed effects'* Schätzer nicht effizient sein, da zumindest $N - 1$ Freiheitsgrade durch die Transformation verloren gehen.

11.8 Schätzung der Parameter

Das allgemeine Modell ist

$$\begin{aligned} y_{it} &= \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \\ \varepsilon_{it} &= \mu_i + \upsilon_{it} \end{aligned}$$

Beim *'fixed effects'* Modell können die zeitinvarianten individuenspezifischen Effekte μ_i als zu schätzende Parameter betrachtet werden, deshalb wird das Modell häufig folgendermaßen geschrieben

$$y_{it} = \alpha + \mu_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \upsilon_{it}$$

wenn das Interzept als getrennte Variable α geschrieben wird (und deshalb nicht in $\boldsymbol{x}$ enthalten ist).

Die einfachste Möglichkeit dieses Modell zu schätzen besteht darin, für die μ_i Dummy Variablen zu verwenden. Dies führt zum **LSDV** Modell ('*Least Squares Dummy Variable* Modell).

Um dies intuitiv zu verstehen erinnern wir uns, dass eine Regression auf ein Set von Individuendummies die Mittelwerte der Variablen entfernt. Nach dem Frisch-Waugh-Lovell Theorem (siehe z.B. Skript Abschnitt 5.8.1) macht es für die Schätzung der Koeffizienten keinen Unterschied, ob die in einem ersten Schritt alle Variablen auf die Dummies regressiert werden, und für die eigentliche Schätzung die Residuen aus diesen Schätzungen verwendet werden, oder ob die Dummies gleich als zusätzliche Variablen in die Schätzung aufgenommen werden.

Wenn $\boldsymbol{\iota}$ wieder ein $T \times 1$ Einsen-Spaltenvektor ist kann das Dummy Modell für Individuum i geschrieben werden als

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\iota}\,\mu_i + \boldsymbol{v}_i$$

oder in 'gestackter' Form

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\iota} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\iota} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{\iota} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix} + \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \\ \vdots \\ \boldsymbol{v}_N \end{pmatrix}$$

oder einfacher

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\mu} + \boldsymbol{v}$$

Allerdings ist Schätzung des *'fixed effects'* Modells mit Hilfe von Dummy Variablen (*LSDV*) in großen Modellen schwierig, da dies die zusätzliche Schätzung von $N-1$ Koeffizienten der Individuendummies erfordern würde. In einem Panel mit mehreren Tausend Individuen würde man dabei schnell an die Grenzen der Rechenkapazität stoßen.

Deshalb verwenden die meisten Programme die *'within'* Transformation, die auch numerisch exakt die gleichen Resultate liefert wie das LSDV Modell (dies folgt aus dem Frisch-Waugh-Lovell Theorem).

Die residuenerzeugende Matrix ist

$$\boldsymbol{M}_D = \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$$

Wenn $\boldsymbol{Z}$ wieder die zeitinvarianten Variablen enthält ist

$$\boldsymbol{M}_D\boldsymbol{y} = \boldsymbol{M}_D\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{M}_D\boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{M}_D\boldsymbol{\varepsilon} = \boldsymbol{M}_D\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{M}_D\boldsymbol{\varepsilon}$$

da $\boldsymbol{M}_D\boldsymbol{Z} = 0$ (die Abweichung der zeitinvarianten Variablen sind Null).

Der aus dieser Transformation folgende *'fixed effects'* Schätzer ist der übliche *'within'* Schätzer

$$\hat{\boldsymbol{\beta}}^{\text{within}} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$$

Falls die individuenspezifischen Interzepte (das sind die Koeffizienten der Dummy Variablen im LSDV Modell) benötigt werden, können sie nachträglich berechnet werden

$$\alpha = \bar{\bar{y}}_{..} - \bar{\bar{\boldsymbol{x}}}'_{..}\boldsymbol{\beta}, \quad \mu_i = \bar{y}_{i.} - \alpha - \bar{\boldsymbol{x}}'_{i.}\boldsymbol{\beta}$$

(mit Normierung $\sum_{i=1}^{N} \mu_i = 0$)

Falls T fix und $N \to \infty$ ist nur der Schätzer für $\boldsymbol{\beta}$ konsistent, aber nicht die Schätzer für μ_i (*incidental parameter problem*).

11.8.1 Fixe Zeit- und Individueneffekte

Dieses Modell kann einfach erweitert werden um zusätzliche Zeiteffekte. Dies sind Effekte, die sich über die Zeit ändern, aber für alle Individuen gleich sind (z.B. Konjunktureffekte für Firmen einer Branche). Wenn wir die Zeiteffekte mit γ_t bezeichnen ist das Modell

$$y_{it} = \mu_i + \gamma_t + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \upsilon_{it}$$

Dieses Modell könnte prinzipiell wieder durch geeignete *'within'* Transformation geschätzt werden, aber da die Anzahl der Zeitperioden meist relativ gering ist werden häufig einfach $T-1$ Zeitdummies (d.h. die erste Dummy hat in der ersten Periode den Wert 1 und Null sonst, die zweite Dummy hat in der zweiten Periode den Wert 1 und Null sonst, usw.) im Modell inkludiert.

11.9 Tests

11.10 Test für fixe Effekte

Die gemeinsame Signifikanz der Individueneffekte, d.h.

$$H_0: \quad \mu_1 = 0, \mu_2 = 0, \ldots, \mu_N = 0$$

kann der übliche F-Test herangezogen werden, im LSDV Modell kann die gemeinsame Signifikanz der Dummy Koeffizienten getestet werden, wenn die *'within'* Transformation angewandt wurde, kann die Quadratsumme der Residuen des restringierten Modells (RSS_r) mit der Quadratsumme der Residuen des nicht restringierten *'within'* Modells (RSS_u) verglichen werden

$$F = \frac{(\mathrm{RSS}_r - \mathrm{RSS}_u)/(N-1)}{\mathrm{RSS}_u/(NT-N-K)} \sim F_{N-1,NT-N-K}$$

11.11 Wu-Hausman Test

Die Grundidee des Hausman Tests besteht in einem Vergleich zweier Schätzer, wovon einer sowohl unter der Null-Hypothese als auch unter der Alternativhypothese konsistent ist, während der zweite nur unter Nullhypothese konsistente Schätzresultate liefert. Ein großer Unterschied zwischen diesen beiden Schätzungen wird als Evidenz zugunsten der Alternativhypothese gedeutet.

Wenn die Annahmen des *'random effects'* Modells erfüllt sind, d.h. die Effekte nicht mit den erklärenden Variablen korreliert sind, ist der *'random effects'* Schätzer konsistent und effizient. In diesem Fall ist der Schätzer des *'fixed effects'* Modells zwar ebenfalls konsistent, aber nicht effizient.

Wenn die Effekte aber mit den den erklärenden Variablen korreliert sind ist der *'fixed effects'* Schätzer konsistent und effizient, aber der *'random effects'* Schätzer ist in diesem Fall nicht konsistent.

Dieser Unterschied wird beim Wu-Hausman Test verwendet. Es gibt einige Formen dieses Tests, aber im Prinzip basieren sie auf

$$H = (\hat{\boldsymbol{\beta}}^{\mathrm{re}} - \hat{\boldsymbol{\beta}}^{\mathrm{fe}})' \left(\boldsymbol{\Sigma}^{\mathrm{fe}} - \boldsymbol{\Sigma}^{\mathrm{re}}\right) (\hat{\boldsymbol{\beta}}^{\mathrm{re}} - \hat{\boldsymbol{\beta}}^{\mathrm{fe}})$$

Diese Wu-Hausman Teststatistik ist unter Gültigkeit der Nullhypothese asymptotisch χ^2-verteilt, wobei die Nullhypothese die Richtigkeit des *'random effects'* Modells unterstellt.

Literaturverzeichnis

Baltagi, B. H. (2008), *Econometric Analysis of Panel Data*, 4 edn, Wiley.

Cameron, A. C. and Trivedi, P. K. (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press.

Greene, W. H. (2007), *Econometric Analysis*, 6th edn, Prentice Hall.

Johnston, J. and Dinardo, J. (1996), *Econometric Methods*, 4 edn, McGraw-Hill/Irwin.