5. Lineare Regressionsmodelle mit Zeitreihendaten

5.1 Vorüberlegungen

Besonderheiten von Zeitreihendaten im Vergleich zu Querschnittsdaten:

- Bei Zeitreihendaten werden eine oder mehrere Variablen über mehrere aufeinander folgende Zeitperioden erhoben, so dass im Gegensatz zu Querschnittsdaten eine explizite zeitliche Anordnung der Beobachtungen vorliegt
- Dabei können zwar zurückliegende Werte zukünftige Werte beeinflussen, jedoch nicht umgekehrt
- Querschnittsdaten werden häufig als Stichprobe vom Umfang n aus der Grundgesamtheit betrachtet, so dass man bei unterschiedlichen Stichproben in der Regel unterschiedliche Ausprägungen von abhängigen und erklärenden Variablen (z.B. Bildung, Berufserfahrung, Geschlecht, Löhne) erhält. Aber auch ökonomische Zeitreihendaten können als Zufallsvariablen aufgefasst werden, da ihre zukünftigen Realisationen im Vorfeld nicht vorhersehbar sind.
- Formal werden Sequenzen von Zufallsvariablen über die Zeit als stochastische Prozesse oder Zeitreihenprozesse aufgefasst
- Bei der Erhebung von Zeitreihendaten beobachtet man jeweils nur eine einzige Realisation des zugrunde liegenden stochastischen Prozesses, so dass nur diese spezifische Realisation empirisch untersucht werden kann

- Falls die zugrundeliegenden Bedingungen in der Vergangenheit anders gewesen wären, hätte man aber eine andere Realisation des stochastischen Prozesses erhalten. Dies ist der zentrale Grund, warum Zeitreihendaten als Realisationen von Zufallsvariablen aufgefasst werden können.
- Beispiele: Makroökonomische Daten (z.B. Einkommen, Konsum, Investitionen, Geldmengen, Preisindizes, Inflationsraten), Finanzmarktdaten (z.B. Aktienkurse, Aktienrenditen)
- Im Gegensatz zur Querschnittsanalyse werden bei der Betrachtung von Zeitreihendaten üblicherweise nicht die Indizes i = 1,…, n für die einzelnen Beobachtungen verwendet, sondern t = 1,…, n für den stochastischen Prozess $\{(x_{t1}, x_{t2},\ldots, x_{tk}, y_t)\}$, wobei n nun die Anzahl der Zeitperioden ist

Damit ergibt sich die folgende Formulierung eines multiplen linearen Regressionsmodells:

$$(5.1) \quad y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t \quad \text{für } t = 1,\ldots, n$$

Dabei bezeichnet $\{u_t: t = 1,\ldots, n\}$ die Sequenz der Störterme und x_{tj} (in Analogie zu linearen Regressionsmodellen mit Querschnittsdaten) den Wert der erklärenden Variablen j = 1,…, k in Zeitperiode t. Im Folgenden beinhalten der k-dimensionale Vektor $x_t = (x_{t1},\ldots, x_{tk})$ die erklärenden Variablen in t sowie die (nxk)-dimensionale Matrix X sämtliche erklärende Variablen über alle Perioden, wobei x_t die t-te Zeile von X darstellt.

Beispiel: Erklärung von Wohnungsbauinvestitionen

Mit Hilfe eines linearen Regressionsmodells soll der Effekt des Zinssatzes für Wohnungsbaukredite in % (hlintrate), des BIP pro Kopf in Euro (GDPcapita) und des Immobilienpreisindexes relativ zu 2010 (hpriceind) auf die Wohnungsbauinvestitionen in % des Bruttoinlandsprodukts für Deutschland untersucht werden. Dabei liegen Daten für n = 8 Quartale von 2004 bis 2005 vor, so dass für X Folgendes gilt:

t	hlintrate	GDPcapita	hpriceind
1/2004	5,063	6854	100,1
2/2004	4,928	6956	98,7
3/2004	4,971	7111	97,7
4/2004	4,765	7229	96,1
1/2005	4,510	6893	100,4
2/2005	4,370	7083	98,6
3/2005	4,125	7242	100,7
4/2005	4,148	7376	97,8

5.2 Eigenschaften von OLS-Schätzern

→ Die unbekannten Parameter in multiplen linearen Regressionsmodellen können grundsätzlich ebenfalls mit der OLS-Methode geschätzt werden. Zur Ableitung der Eigenschaften von OLS-Schätzern bei endlichen Stichprobenumfängen sowie ihrer asymptotischen Eigenschaften werden erneut verschiedene Annahmen getroffen, die sich teilweise von den Annahmen bei Querschnittsanalysen unterscheiden.

Annahmen zur Betrachtung des Erwartungswerts von OLS-Schätzern:

- Annahme B1: Linearität in den Parametern
  Der Zeitreihenprozess $\{(x_{t1},\ldots, x_{tk}, y_t): t = 1,\ldots, n\}$ folgt dem linearen Modell
  $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$
- Annahme B2: Keine perfekte Kollinearität
  In der Stichprobe (und damit auch im zugrundeliegenden Zeitreihenprozess) ist keine der erklärenden Variablen konstant, und es besteht keine exakte lineare Beziehung zwischen den erklärenden Variablen
- Annahme B3: Bedingter Erwartungswert von u_t ist null
  Für jede Zeitperiode t ist der bedingte Erwartungswert von u_t, gegeben die erklärenden Variablen für alle Perioden $t = 1,\ldots, n$, gleich null, d.h.

$$(5.2) \quad E(u_t|X) = 0 \quad t = 1,\ldots, n$$

Zu B3 als wesentliche Annahme:

- Sie impliziert, dass der Störterm u_t in einer Zeitperiode t mit jeder erklärenden Variablen x_{tj} (j = 1,…, k) in jeder Periode t = 1,…, n unkorreliert ist. Falls u_t unabhängig von X und $E(u_t) = 0$, dann gilt (5.2) automatisch.
- Im folgenden Fall sind die x_{tj} lediglich kontemporär exogen:

  $$(5.3) \quad E(u_t|x_{t1},\ldots, x_{tk}) = E(u_t|x_t) = 0 \quad \text{für } t = 1,..., n$$

  In diesem Fall sind die u_t und die erklärenden Variablen in derselben Periode unkorreliert, d.h. $Corr(x_{tj}, u_t) = 0$ für alle j = 1,…, k.
- Annahme B3 erfordert jedoch mehr als kontemporäre Exogenität, da u_t mit x_{sj} selbst für s ≠ t unkorreliert sein muss. In diesem Fall spricht man von strikter Exogenität der erklärenden Variablen, wenngleich kontemporäre Exogenität für einige Eigenschaften von OLS-Schätzern ausreichend ist.
- Bei Querschnittsanalysen ist die Annahme der Unkorreliertheit des Störtermes u_i mit den erklärenden Variablen x_{sj} für andere Beobachtungen (s ≠ i) nicht notwendig, sofern eine Zufallsstichprobe unterstellt wird, da diese automatisch die Unabhängigkeit zwischen u_i und x_{sj} (s ≠ i) impliziert
- Bei linearen Regressionsanalysen mit Zeitreihendaten ist die Annahme von Zufallsstichproben dagegen fast nie angemessen, so dass hier mit B3 eine spezifische und strenge Annahme getroffen werden muss. Falls aber der Störterm u_t doch mit einer erklärenden Variablen in irgendeiner Zeitperiode korreliert ist, liegt ein Verstoß gegen Annahme B3 vor.

- Annahme B3 ist in linearen Regressionsanalysen mit Zeitreihendaten jedoch sehr häufig unrealistisch und gilt nur in vergleichsweise wenigen Fällen. Dennoch soll diese Annahme zunächst betrachtet werden.

Auch unter den drei Annahmen B1 bis B3 sind alle mit der OLS-Methode geschätzten Parameter sowohl bedingt auf X als auch unbedingt erwartungstreu, d.h. der Erwartungswert der Schätzer ist gleich den unbekannten Regressionsparametern:

$$(5.4) \quad E(\hat{\beta}_j) = \beta_j \quad \text{für } j = 0, 1, \ldots, k$$

Annahmen zur Betrachtung der Varianz von OLS-Schätzern:

- Annahmen B1 bis B3 (wie sie bereits bei der Betrachtung des Erwartungswertes von OLS-Schätzern getroffen werden)
- Annahme B4: Homoskedastizität
  Die bedingte Varianz des Fehlerterms u_t ist konstant über alle Zeitperioden $t = 1, \ldots, n$, d.h. es gilt $Var(u_t|X) = Var(u_t) = \sigma^2$. Falls diese Annahme verletzt ist, liegt wie bei Querschnittsanalysen Heteroskedastizität vor.
- Annahme B5: Keine Autokorrelation
  Bedingt auf X sind die Störterme für beliebige Zeitperioden unkorreliert, d.h. es gilt $Corr(u_t,u_s|X) = 0$ für alle $t \neq s$. Zur Vereinfachung der Darstellung und Interpretation wird häufig von der Bedingung auf X abstrahiert:

$$(5.5) \quad Corr(u_t,u_s) = 0 \quad \text{für alle } t \neq s$$

Zu Annahme B5:

- Falls B5 nicht erfüllt ist, liegt bei den Störtermen Autokorrelation bzw. serielle Korrelation über die Zeit vor
- Wenn z.B. bei einem positiven Wert von u_{t-1} im Durchschnitt auch u_t in der folgenden Zeitperiode positiv ist, gilt $Corr(u_t,u_{t-1}) > 0$, so dass die Störterme (positiv) autokorreliert sind. Dies impliziert, dass ein sehr hoher Wert der abhängigen Variablen (z.B. ein Zinssatz) in t-1 im Durchschnitt (bei konstanten Werten der erklärenden Variablen) im Durchschnitt zu hohen Werten in der folgenden Periode t führt.
- Zu beachten ist, dass Annahme B5 keine Implikationen für intertemporale Korrelationen in den erklärenden Variablen hat, d.h. solche Korrelationen sind mit B5 vereinbar
- Bei Querschnittsanalysen ist Annahme B5 nicht erforderlich, sofern eine Zufallsstichprobe unterstellt wird, da diese automatisch die Unabhängigkeit der Störterme u_i und u_h für zwei beliebige Beobachtungen i und h impliziert
- Autokorrelation wird daher vornehmlich als Problem von Zeitreihenanalysen betrachtet, so dass die Annahmen B1 bis B5 in diesem Kontext passende Gauss-Markov-Annahmen darstellen (allerdings können diese Annahmen vereinzelt auch bei Querschnittsanalysen relevant werden, falls die Annahme einer Zufallsstichprobe nicht adäquat ist, z.B. wenn die untersuchten Beobachtungen einen großen Anteil der Grundgesamtheit ausmachen)

Damit ergibt sich unter den Annahmen B1 bis B5 für die bedingte Varianz der mit OLS geschätzten Steigungsparameter:

$$(5.6) \quad \mathrm{Var}(\hat{\beta}_j|X) = \frac{\sigma^2}{(1-R_j^2)\sum_{t=1}^{n}(x_{tj}-\bar{x}_j)^2} = \frac{\sigma^2}{(1-R_j^2)SST_j} \quad \text{für } j = 1,\ldots, k$$

Dabei stellt R_j^2 das Bestimmtheitsmaß bei einer Regression von x_j auf alle übrigen erklärenden Variablen (einschließlich einer Konstanten) dar.

Des Weiteren gilt:

- Unter den Annahmen B1 bis B5 ergibt sich folgender erwartungstreuer Schätzer für die Varianz σ^2 des Fehlerterms u_t:

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\sum_{t=1}^{n}\hat{u}_t^{\,2} = \frac{SSR}{n-k-1}$$

- Unter den Annahmen B1 bis B5 sind die OLS-Schätzer die besten linearen unverzerrten Schätzer unter der Bedingung von X

Somit weisen lineare Regressionsanalysen mit Zeitreihendaten unter den Annahmen B1 bis B5 dieselben wünschenswerten Eigenschaften bei endlichen Stichproben auf wie Querschnittsanalysen unter den Annahmen A1 bis A5.

→ Für die Ableitung der geschätzten Varianzen der mit OLS geschätzten Steigungsparameter sowie von t- und F-Statistiken muss aber zunächst erneut eine zusätzliche Annahme getroffen werden.

Annahmen:

- Annahmen B1 bis B5 (die bei der Betrachtung der Varianz von OLS-Schätzern verwendet wurden)
- Annahme B6: Normalverteilung
  Die Störterme u_t sind unabhängig von X und unabhängig und identisch normalverteilt mit Erwartungswert null und Varianz σ^2, d.h. es gilt: $u_t \sim N(0; \sigma^2)$. Dabei impliziert B6 die Annahmen B3 bis B5, jedoch ist diese Annahme aufgrund der Unabhängigkeit und Normalverteilung strenger.

Unter den klassischen linearen Modellannahmen B1 bis B6 in linearen Regressionsanalysen mit Zeitreihendaten gilt:

Die OLS-Schätzer sind unter der Bedingung von X normalverteilt, die konventionelle Konstruktion von Konfidenzintervallen ist gültig und unter den entsprechenden Nullhypothesen folgen die t- und F-Statistiken der t- und F-Verteilung

Damit ergeben sich mit B1 bis B6 dieselben Folgerungen wie unter den Annahmen A1 bis A6. Jedoch sind B1 bis B6 für lineare Regressionsanalysen mit Zeitreihendaten restriktiver (insbesondere strikte Exogenität und keine Autokorrelation sind häufig unrealistisch) als A1 bis A6 für Querschnittsanalysen.

→ Aufgrund der zuvor betrachteten bei linearen Regressionsanalysen mit Zeitreihendaten sehr restriktiven Annahmen, die für wünschenswerte Eigenschaften von OLS-Schätzern bei endlichen Stichprobenumfängen erforderlich sind, gewinnen asymptotische Eigenschaften unter weniger restriktiven Annahmen bei Zeitreihenanalysen eine noch größere Bedeutung als bei Querschnittsanalysen

Annahmen zur Betrachtung der Konsistenz von OLS-Schätzern:

- Annahme B1‘: Linearität, Stationarität und schwache Abhängigkeit
  Es gilt Annahme B1, d.h. der Zeitreihenprozess $\{(x_t, y_t): t = 1,\ldots, n\}$ folgt dem linearen Modell $y_t = \beta_0 + \beta_1 x_{t1} + \ldots + \beta_k x_{tk} + u_t$. Zusätzlich ist der Zeitreihenprozess aber stationär und schwach abhängig.
- Annahme B2‘: Keine perfekte Kollinearität (so dass Annahme B2 gilt)
- Annahme B3‘: Bedingter Erwartungswert von u_t ist null
  Im Gegensatz zu Annahme B3 wird nun nicht mehr die strikte Exogenität der erklärenden Variablen, sondern lediglich die kontemporäre Exogenität entsprechend (5.3) betrachtet, d.h. $E(u_t|x_t) = 0$. Häufig wird für die Konsistenzeigenschaft auch lediglich folgende schwächere Bedingung angenommen:

$$(5.7) \quad E(u_t) = 0,\ Cov(x_{tj}, u_t) = 0 \quad j = 1,\ldots, k$$

Unter diesen drei Annahmen sind die OLS-Schätzer $\hat{\beta}_j$ konsistent (wenngleich nicht notwendigerweise erwartungstreu), d.h. $plim(\hat{\beta}_j) = \beta_j$ für $j = 0,1,\ldots, k$.

Die Annahmen zur Ableitung der asymptotischen Normalverteilung von Funktionen von OLS-Schätzern in linearen Regressionsanalysen mit Zeitreihendaten und damit zur Durchführung von Testverfahren sind etwas weniger restriktiv als die klassischen linearen Modellannahmen B1 bis B6 für endliche Stichproben:

- Annahmen B1‘ bis B3‘ (die bereits für die Konsistenz von OLS-Schätzern erforderlich sind)
- Annahme B4‘: Kontemporäre Homoskedastizität
  Die bedingte Varianz des Fehlerterms u_t ist konstant, wobei nicht mehr auf die erklärenden Variablen in allen Zeitperioden t = 1,…, n, sondern nur noch in der jeweiligen Zeitperiode t konditioniert wird: $Var(u_t|x_t) = Var(u_t) = \sigma^2$.
- Annahme B5‘: Keine Autokorrelation
  Für alle $t \neq s$ gilt $E(u_t u_s|x_t, x_s) = 0$, d.h. es wird nur noch auf die erklärenden Variablen in den Zeitperioden von u_t und u_s konditioniert. Auch bei dieser Annahme wird zur einfacheren Interpretation oft von der Bedingung abstrahiert und lediglich die Unkorreliertheit von u_t und u_s betrachtet.

Unter den Annahmen B1‘ bis B5‘ ergibt sich, dass Funktionen der OLS-Schätzer asymptotisch normalverteilt und die OLS-Schätzer asymptotisch effizient sind sowie dass die t- und F-Statistiken asymptotisch t- und F-verteilt sind. Dadurch können dann die konventionellen t- und F-Tests durchgeführt und Konfidenzintervalle konstruiert werden.

Beispiel: Erklärung von kurzfristigen Zinssätzen

Mit Hilfe eines linearen Regressionsmodells soll für die Jahre 1950 bis 2024 in Deutschland der Effekt der mit dem Verbraucherpreisindex gemessenen jährlichen Inflationsrate in % (infrate) und der Staatsverschuldung in % vom BIP (debtgdp) auf den Zinssatz für Treasury Bills (d.h. den kurzfristigen Geldmarktzins, „three-month T-bill rate“) in % (intrate) untersucht werden. Dabei haben sich mit STATA folgende OLS-Schätzergebnisse gezeigt (n = 75, R^2 = 0,513):

```
reg intrate infrate debtgdp
------------------------------------------------------------------------------
     intrate | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     infrate |    .524205    .114426     4.58   0.000      .296101     .752309
     debtgdp |    -.07806   .0128798    -6.06   0.000    -.1037354   -.0523847
       _cons |   6.276597   .6972849     9.00   0.000     4.886586    7.666609
------------------------------------------------------------------------------
```

Interpretation:

- Die jährliche Inflationsrate hat einen signifikant positiven Effekt und die Staatsverschuldung hat einen signifikant negativen Effekt auf die kurzfristigen Zinssätze
- Zum Beispiel führt eine Erhöhung der jährlichen Inflationsrate um einen Prozentpunkt ceteris paribus zu einem geschätzten Anstieg des kurzfristigen Geldmarktzinses um 0,524 Prozentpunkte

5.3 Dummy-Variablen, Zeittrends und Saisonalität

→ Auch in linearen Regressionsanalysen mit Zeitreihendaten werden häufig Dummy-Variablen, Interaktionsterme sowie logarithmierte abhängige und erklärende Variablen verwendet

Besonderheiten der Einbeziehung von Dummy-Variablen in Zeitreihenanalysen:

Da eine Beobachtungseinheit jeweils einer Zeitperiode entspricht, zeigen Dummy-Variablen an, ob sich ein bestimmtes Ereignis in einer oder mehreren Perioden eingetreten ist

Ausgehend von (5.1) werden nun zusätzlich qualitative erklärende Variablen mit q verschiedenen Ausprägungen betrachtet. In diesem Fall können (maximal) q-1 Dummy-Variablen $x_{t01}, x_{t02},\ldots, x_{t,0,q-1}$ (neben den nun k-q+1 quantitativen erklärenden Variablen $x_{t1}, x_{t2},\ldots, x_{t,k-q+1}$) einbezogen werden:

$$(5.8) \quad y_t = \beta_0 + \beta_1 x_{t01} + \beta_2 x_{t02} + \cdots + \beta_{q-1} x_{t,0,q-1} + \beta_q x_{t1} + \beta_{q+1} x_{t2} + \cdots + \beta_k x_{t,k-q+1} + u_t$$

Die q-te Ausprägung der qualitativen Variablen (d.h. die nicht einbezogene Dummy-Variable x_{t0q}) dient dabei als Basisgruppe.

→ In linearen Regressionsanalysen mit Zeitreihendaten werden Dummy-Variablen häufig verwendet, um einzelne Zeitperioden, die sich systematisch von anderen Perioden unterscheiden, zu isolieren

Beispiel: Erklärung von Fertilitätsraten

Mit Hilfe eines linearen Regressionsmodells soll für die Jahre von 1955 bis 2023 für Westdeutschland der Effekt der Arbeitslosenquote in % (unemprate) sowie zweier Dummy-Variablen für die Jahre vor der Einführung der Antibaby-pille im Jahr 1961 (beforepill) und nach der Wiedervereinigung 1990 (reunion) auf die zusammengefasste Geburtenziffer (totfertrate), d.h. die durchschnittliche Anzahl der pro Frau geborenen Kinder, untersucht werden. Dabei haben sich mit STATA folgende OLS-Schätzergebnisse gezeigt (n = 69, R^2 = 0,713):

```
reg totfertrate unemprate beforepill reunion
------------------------------------------------------------------------------
 totfertrate | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   unemprate |  -.0953304   .0109728    -8.69   0.000    -.1172446   -.0734162
  beforepill |   .4516149   .1008972     4.48   0.000     .2501092    .6531206
     reunion |   .0024726   .0690162     0.04   0.972    -.1353623    .1403075
       _cons |   2.160141   .0610268    35.40   0.000     2.038262     2.28202
------------------------------------------------------------------------------
```

Die geschätzten Steigungsparameter implizieren, dass (bei konstanter Arbeitslosenquote) die durchschnittliche Anzahl der pro Frau geborenen Kinder in den Jahren vor der Einführung der Pille um 0,45 höher war als in der Referenzperiode von 1961 bis 1990. Dagegen unterscheidet sich die zusammengefasste Geburtenziffer in den Jahren nach der deutschen Wiedervereinigung nicht signifikant von derjenigen in der Referenzperiode.

→ Eine besondere Bedeutung kommt der Einbeziehung von Dummy-Variablen bei (finanzökonomischen) Event-Studien zu. Dabei wird untersucht, ob ein spezifisches Ereignis einen Einfluss auf Aktienrenditen hat.

Versionen linearer Regressionsmodelle in Event-Studien:

$$R_t = \beta_0 + \beta_1 R_t^m + \beta_2 d_t + u_t$$

$$R_t = \beta_0 + \beta_1 R_t^m + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 d_t + u_t$$

Dabei bezeichnen R_t die Aktienrendite eines Unternehmens (oder eines Portfolios), R^m_t die Marktrendite sowie SMB_t und HML_t zwei weitere Kapitalmarktfaktoren (nach Fama und French) zur Erklärung von Aktienrenditen. Die Dummy-Variable d_t symbolisiert schließlich die Zeitperiode(n), in denen das spezifische Ereignis stattfindet

Beispiele für (neue überraschende und zuvor nicht bekannte) Ereignisse:

- Zusammenschlüsse oder Übernahmen von Unternehmen
- Aufnahmen in Aktienindizes (z.B. Nachhaltigkeitsindizes)
- Stör- oder Unfälle eines Unternehmens oder einer Branche (z.B. Unfall auf Ölplattform Deepwater Horizon, Atomkatastrophe von Fukushima)
- Phasen von Regulierungen wie z.B. Umweltgesetzgebungen
- Informationen über umwelt- oder sozialrelevante unternehmerische Maßnahmen („Corporate Social Responsibility“)

Meist werden in diesem Zusammenhang jedoch alternative zweistufige Ansätze verwendet. Dabei werden zunächst vor dem Ereignis für ein Unternehmen z.B. folgende lineare Regressionsmodelle betrachtet:

$$R_t = \beta_0 + \beta_1 R_t^m + u_t$$

$$R_t = \beta_0 + \beta_1 R_t^m + \beta_2 SMB_t + \beta_3 HML_t + u_t$$

Die mit OLS geschätzten Regressionsparameter $\hat{\beta}_0$ und $\hat{\beta}_1$ im ersten Ansatz oder $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ und $\hat{\beta}_3$ im zweiten Ansatz werden anschließend zur Schätzung der abnormalen Renditen AR_t in einer Zeitperiode t innerhalb des Event-Fensters („event window“), d.h. während der Zeit, in der das Ereignis wirkt, verwendet:

$$A\hat{R}_t = R_t - \hat{\beta}_0 - \hat{\beta}_1 R_t^m$$

$$A\hat{R}_t = R_t - \hat{\beta}_0 - \hat{\beta}_1 R_t^m - \hat{\beta}_2 SMB_t - \hat{\beta}_3 HML_t$$

Diese geschätzten abnormalen Renditen können dann über mehrere Zeitperioden im Event-Fenster und/oder über verschiedene Unternehmen aggregiert werden. Auf diese Weise erhält man geschätzte durchschnittliche, geschätzte kumulierte und geschätzte durchschnittliche kumulierte abnormale Renditen. Mit Hilfe verschiedener statistischer Testverfahren wird schließlich überprüft, ob diese geschätzten abnormalen Renditen signifikant von null verschieden sind.

Zeittrends:

Zeitreihendaten können im Zeitverlauf einen fallenden und vor allem einen steigenden Trend aufweisen (z.B. Arbeitsproduktivität, nominale Importe)

Einfaches Modell für lineare Zeittrends eines stochastischen Prozesses $\{y_t\}$:

$$(5.9) \quad y_t = \alpha_0 + \alpha_1 t + e_t \quad \text{für } t = 1, 2, ...$$

Im einfachsten Fall ist $\{e_t\}$ eine unabhängig und identisch verteilte zufällige Sequenz mit $E(e_t) = 0$ und $Var(e_t) = \sigma_e^2$. Falls alle Faktoren in e_t konstant sind mit $\Delta e_t = 0$, bildet α_1 die durchschnittliche Veränderung von y von einer Zeitperiode zur nächsten ab, d.h. $\Delta y_t = y_t - y_{t-1} = \alpha_1$. Zudem ergibt sich:

$$(5.10) \quad E(y_t) = \alpha_0 + \alpha_1 t$$

Bei $\alpha_1 > 0$ weist die Zeitreihe im Durchschnitt einen wachsenden Trend und bei $\alpha_1 < 0$ einen sinkenden Trend auf. Im Gegensatz zum Erwartungswert ist die Varianz von y_t über die Zeit konstant.

→ Falls bei abhängigen und/oder erklärenden Variablen in Regressionsanalysen mit Zeitreihendaten Trends vorliegen, sollten diese explizit berücksichtigt werden, da andernfalls scheinbare Zusammenhänge („spurious regression“) auftreten können. Ohne die Einbeziehung relevanter Trendvariablen werden wichtige relevante erklärende Variablen vernachlässigt, was zu verzerrten Schätzungen der Regressionsparameter führen kann („omitted variable bias“).

Zusätzliche Einbeziehung eines linearen Zeittrends in (5.1) (neben nun k-1 weiteren erklärenden Variablen $x_{t1}, x_{t2}, \ldots, x_{t,k-1}$):

$$(5.11) \quad y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_{k-1} x_{t,k-1} + \beta_k t + u_t$$

Anmerkungen:

- Neben linearen Zeittrends können auch quadratische Zeittrends sowie weitere Polynome in t einbezogen werden, wenngleich die Trendkomponenten möglichst einfach gehalten werden sollten
- Falls der Zeittrend signifikant von null verschieden ist und sich die Schätzergebnisse deutlich verändern, sollten die geschätzten Regressionsparameter aus linearen Regressionsmodellen mit Zeitreihendaten ohne Einbeziehung der Trendvariablen nur sehr vorsichtig interpretiert werden
- Die Einbeziehung eines Zeittrends als erklärende Variable führt zu einer Trendbereinigung, d.h. die geschätzten Steigungsparameter der übrigen erklärenden Variablen können dann als geschätzte Effekte ohne den Zeittrend interpretiert werden
- Konventionelle und angepasste Bestimmtheitsmaße können in Regressionsanalysen mit Zeitreihendaten artifiziell sehr hohe Werte aufweisen, falls die abhängige Variable einen Trend aufweist. Dies bedeutet jedoch nicht zwingend, dass ein hoher Anteil der Variation der abhängigen Variablen durch die Variation der erklärenden Variablen erklärt wird, da die Varianz der abhängigen Variablen nicht mehr unverzerrt oder konsistent geschätzt wird.

Beispiel 1: Erklärung von Wohnungsbauinvestitionen

Mit Hilfe eines linearen Regressionsmodells soll für die 80 Quartale von 2004 bis 2023 für Deutschland der Effekt des Immobilienpreisindexes relativ zu 2010 (hpriceind) sowie des Zinssatzes für Wohnungsbaukredite in % (hlintrate) auf die Wohnungsbauinvestitionen in % des Bruttoinlandsprodukts (hcinvest) untersucht werden. Dabei haben sich mit STATA folgende OLS-Schätzergebnisse gezeigt:

```
reg hcinvest hpriceind hlintrate

      Source |       SS           df       MS      Number of obs   =        80
-------------+----------------------------------   F(2, 77)        =    401.77
       Model |  30.4425112         2  15.2212556   Prob > F        =    0.0000
    Residual |  2.91715403        77  .037885117   R-squared       =    0.9126
-------------+----------------------------------   Adj R-squared   =    0.9103
       Total |  33.3596652        79  .422274243   Root MSE        =    .19464

------------------------------------------------------------------------------
    hcinvest | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   hpriceind |   .0157058   .0009055    17.34   0.000     .0139026    .0175089
   hlintrate |  -.1113963   .0209521    -5.32   0.000    -.1531172   -.0696754
       _cons |   4.126187   .1653679    24.95   0.000     3.796897    4.455476
------------------------------------------------------------------------------
```

Beispiel 1: Erklärung von Wohnungsbauinvestitionen (Fortsetzung)

Es ist jedoch möglich, dass hcinvest einem steigenden Trend folgt, so dass zusätzlich eine entsprechende (lineare) Trendvariable (t) einbezogen wird. Dabei haben sich mit STATA folgende OLS-Schätzergebnisse gezeigt:

```
reg hcinvest hpriceind hlintrate t

      Source |       SS           df       MS      Number of obs   =        80
-------------+----------------------------------   F(3, 76)        =    285.93
       Model |  30.6445603         3  10.2148534   Prob > F        =    0.0000
    Residual |  2.71510493        76  .035725065   R-squared       =    0.9186
-------------+----------------------------------   Adj R-squared   =    0.9154
       Total |  33.3596652        79  .422274243   Root MSE        =    .18901

------------------------------------------------------------------------------
    hcinvest | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   hpriceind |   .0121703   .0017272     7.05   0.000     .0087303    .0156104
   hlintrate |  -.0734871   .0258469    -2.84   0.006    -.1249656   -.0220086
           t |   .0069869   .0029379     2.38   0.020     .0011355    .0128383
       _cons |   4.175554   .1619206    25.79   0.000     3.853062    4.498047
------------------------------------------------------------------------------
```

Die geschätzten Effekte von hpriceind und hlintrate fallen nun schwächer aus als zuvor. Der lineare Zeittrend in hcinvest ist signifikant positiv. Der geschätzte Parameter für t impliziert, dass die Wohnungsbauinvestitionen (ceteris paribus) im Durchschnitt um etwa 0,007 Prozentpunkte des BIP pro Quartal steigen.

Beispiel 2: Erklärung von Fertilitätsraten

Wie zuvor soll wieder mit Hilfe eines linearen Regressionsmodells die zusammengefasste Geburtenziffer analysiert werden. Zusätzlich zu unemprate, beforepill und reunion wird nun zur Erklärung von totfertrate auch eine lineare Zeittrendvariable (t) einbezogen. Dabei haben sich mit STATA folgende OLS-Schätzergebnisse gezeigt (n = 69, R^2 = 0,763):

```
reg totfertrate unemprate beforepill reunion t
------------------------------------------------------------------------------
 totfertrate | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   unemprate |   -.084617   .0103605    -8.17   0.000    -.1053144   -.0639196
  beforepill |   .2555674   .1050559     2.43   0.018      .045694    .4654407
     reunion |   .3177886   .1035511     3.07   0.003     .1109216    .5246556
           t |  -.0112248   .0029326    -3.83   0.000    -.0170833   -.0053663
       _cons |   2.357443   .0757295    31.13   0.000     2.206156     2.50873
------------------------------------------------------------------------------
```

Der geschätzte sinkende lineare Zeittrend ist hochsignifikant von null verschieden. Daraus ergibt sich, dass der geschätzte Parameter von beforepill nun kleiner ist. Wichtiger ist jedoch, dass die geschätzte zusammengefasste Geburtenziffer in den Jahren nach der deutschen Wiedervereinigung nun signifikant höher ist als in der Referenzperiode von 1961 bis 1990.

Beispiel 2: Erklärung von Fertilitätsraten (Fortsetzung)

Wie zuvor soll wieder mit Hilfe eines linearen Regressionsmodells die zusammengefasste Geburtenziffer analysiert werden. Zusätzlich zu unemprate, beforepill und reunion wird nun neben t allerdings auch eine quadrierte lineare Zeittrendvariable (sqt) zur Erklärung von totfertrate einbezogen. Dabei haben sich mit STATA folgende OLS-Schätzergebnisse gezeigt (n = 69, R^2 = 0,882):

```
reg totfertrate unemprate beforepill reunion t tsq
------------------------------------------------------------------------------
 totfertrate | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   unemprate |  -.0084679   .0126963    -0.67   0.507    -.0338393    .0169036
  beforepill |   -.521227   .1292379    -4.03   0.000    -.7794885   -.2629656
     reunion |   .3231731   .0759323     4.26   0.000     .1714344    .4749117
           t |  -.0798979   .0094226    -8.48   0.000    -.0987275   -.0610684
         tsq |   .0007982   .0001066     7.49   0.000     .0005851    .0010113
       _cons |   3.092157   .1127688    27.42   0.000     2.866807    3.317508
------------------------------------------------------------------------------
```

Die Schätzergebnisse implizieren einen signifikanten U-förmigen Zeittrend der zusammengefassten Geburtenziffer. Daraus ergibt sich, dass die Arbeitslosenquote (unemprate) keinen signifikanten Einfluss auf die Fertilitätsrate mehr hat.

Saisonalität:

Falls sich Zeitreihendaten auf Monate oder Quartale (bzw. auch Wochen oder Tage) beziehen, kann Saisonalität vorliegen, z.B. Wettereinflüsse bei makroökonomischen Variablen wie Beschäftigungszahlen. Andere Variablen wie z.B. Zins- oder Inflationsraten weisen dagegen selten Saisonalität auf. Für einige Variablen mit Saisonalität können Daten bereits im Vorfeld saisonbereinigt sein.

Falls bei abhängigen und/oder erklärenden Variablen in Regressionsanalysen mit Zeitreihendaten Saisonalität vorliegt und die Daten noch nicht saisonbereinigt wurden, sollten entsprechende saisonale Dummy-Variablen einbezogen werden. Bei Monatsdaten und der zusätzlichen Einbeziehung von elf Dummy-Variablen für die Monate Februar (feb) bis Dezember (dec) in (5.1) (neben nun k-11 sonstigen erklärenden Variablen $x_{t1}, x_{t2}, \ldots, x_{t,k-11}$) ergibt sich:

$$(5.12) \quad y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_{k-11} x_{t,k-11} + \beta_{k-10} feb_t + \cdots + \beta_k dec_t + u_t$$

Anmerkungen:

- Bei Quartalsdaten können Dummy-Variablen für maximal drei Quartale einbezogen werden
- Die Einbeziehung saisonaler Dummy-Variablen führt zu einer Saisonbereinigung, d.h. die geschätzten Steigungsparameter der erklärenden Variablen können als geschätzte Effekte ohne Saisonalität interpretiert werden
- Saisonale Dummy-Variablen können gemeinsam mit Trendvariablen einbezogen werden, so dass dann eine Saison- und Trendbereinigung erfolgt

---

Beispiel: Erklärung von Wohnungsbauinvestitionen

Wie zuvor sollen wieder mit Hilfe eines linearen Regressionsmodells die Wohnungsbauinvestitionen analysiert werden. Allerdings werden nun zusätzlich zu hpriceind, hlintrate und t auch die drei saisonalen Dummy-Variablen quarter1, quarter2 und quarter3 einbezogen. Dabei haben sich mit STATA folgende OLS-Schätzergebnisse gezeigt (n = 80, R^2 = 0,920):

```
reg hcinvest hpriceind hlintrate t quarter1 quarter2 quarter3
------------------------------------------------------------------------------
    hcinvest | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   hpriceind |   .0120938   .0017501     6.91   0.000     .0086058    .0155818
   hlintrate |  -.0722392   .0262101    -2.76   0.007    -.1244758   -.0200026
           t |   .0071801   .0029811     2.41   0.019     .0012388    .0131213
    quarter1 |   .0358074   .0606612     0.59   0.557    -.0850902    .1567049
    quarter2 |   .0611169   .0606017     1.01   0.317    -.0596621     .181896
    quarter3 |     .01381   .0605297     0.23   0.820    -.1068254    .1344455
       _cons |   4.145946   .1683686    24.62   0.000     3.810388    4.481505
------------------------------------------------------------------------------
```

Die geschätzten Parameter für quarter1, quarter2 und quarter3 sind bei allen üblichen Signifikanzniveaus nicht von null verschieden, so dass keine Evidenz für saisonale Effekte bei Wohnungsbauinvestitionen vorliegt.