Notiz über Auflistungen der Annahmen bei der Methode der kleinsten Quadrate in einigen U.S. Lehrbüchern der Ökonometrie

(im Verhältnis zur Vorlesung und zum Buch von L. v. Auer)

P. v. d. Lippe (03.06.2013)

Im Folgenden stellen wir zusammen, was in einigen U.S. Lehrbücher über die erforderlichen Annahmen für die Least Squares (LS) Methode gesagt wird. Dabei zeigt sich, dass oft das, was auf dem ersten Blick unterschiedlich erscheint, doch bei genauerer Betrachtung auf das Gleiche hinausläuft. Die Bücher sind aber trotzdem auch in vielen anderen Punkten, insbesondere was Aufbau und die Darstellungsweise betrifft, sehr unterschiedlich (und m.E. auch sehr unterschiedlich als Einführung geeignet).[1]

D. N. Gujarati, D. C. Porter, Basic Econometrics

Sehr zu empfehlendes Buch! Es beginnt in ch. 3 (S. 61ff), mit dem Gaussian -. standard- oder classical linear regression model, also so wie auch wir das in der Vorlesung tun (bzw. v. Auer in seinem Lehrbuch, nach dem die Vorlesung aufgebaut ist). Es wird ausführlich begründet, warum zuerst von nichtstochastische Regressoren ausgegangen wird (S. 62). Stochastic regressors (also die Aufhebung der Annahme C1 in unserer [v. Auers] Terminologie) werden erst in ch. 13 [ab S. 467] behandelt. Das Buch listet die folgenden sieben Annahmen auf:

1. linear in the parameters
2. fixed x values or x values independent of the error term (entspricht C1)
3. zero mean value of disturbance u (eigentl. distubances u_i), d.h. $E(u_i|X_i) = 0$ wenn X stochastisch, und $E(u_i) = 0$ wenn x nonstochastic ist;
   Gezeigt wird auch die große Relevanz dieser Annahme und dass sie "no specification bias" (unsere Annahme A1) impliziert.[2]
4. homoscedasticity or constant variance of u_i (gleiche weitere Hinweise wie bei assumption 3, jetzt auch wieder die Unterscheidung $\mathrm{var}(u_i|X_i) = \sigma^2$ und $\mathrm{var}(u_i) = \sigma^2$ und ausführliche Bemerkungen zur Unterscheidung zwischen conditional und unconditional variance of u)
5. no autocorrelation between disturbances $\mathrm{cov}(u_i\, u_j|X_i\, X_j) = 0$ bzw. $\mathrm{cov}(u_i\, u_j) = 0$
6. the number of observations n must be greater than the number of parameters to be estimated (alternatively: the number of explanatory variables), wie assumption 7 wichtig für Rang der Momentenmatrix **X'X**; bei uns ist dies in Annahme C2 [keine offene Multikollinearität] enthalten
7. the nature of X-variables: the X variables in a given sample must not all be the same (technically $\mathrm{var}(x) > 0$). Geht auch auf die Annahme "that there are no outliers" ein und weist besonders darauf hin, dass die Annahmen 1 bis 7 für die PRF, nicht für die SRF[3]

M. P. Murray, Econometrics, A Modern Introduction

Insgesamt zwar didaktisch sehr bemüht, aber im Aufbau eher verwirrend, weil Verf. offenbar vermutet, dass diverse Exkurse das Verständnis erleichtern (er führt z.B. für die Steigung β bei einfacher Regression Schätzfunktionen $\hat{\beta}_{g1}$ bis $\hat{\beta}_{g4}$ ein und demonstriert deren Eigenschaften), was aber wohl eher nicht der Fall sein dürfte. Das Buch enthält zwar z.T. gute Erklärungen, es ist aber – wie gesagt – nicht leicht, sich darin zurecht zu finden.

---

[1] Vieles mag aber auch Geschmackssache sein: für jeden wird es Bücher geben, die ihm zusagen und solche, mit denen er weniger anfangen kann.

[2] eigentlich nicht ganz korrekt, denn $E(u_i) \neq 0$, bzw. bei uns $E(u_t) \neq 0$, kann auch unter anderen Voraussetzungen als omitted variables (z.B. bei Nichtlinearität, Messfehler, censored oder truncated data) auftreten. Richtig ist allerdings, dass unsere Annahmen A1 (keine Fehlspezifikation) und B1 ($E(u_t) = 0,\ \forall\ t$) nicht völlig unabhängig (im nicht-technischen Sinne also im Sinne von "nichts mit einander zu tun haben") von einander sind (das gilt auch in anderen Fällen für das Verhältnis von zwei Annahmen zu einander; Nichterfüllen von A2 [oder oben Annahme 1] kann z.B. wie omitted variables auch zur Verletzung von B1 führen).

[3] PRF = population regression function, SRF = sample regression function.

Auch Murray beginnt bei den "**Gauss-Markov Assumptions**" mit dem Fall nichtstochastischer Regressoren (S. 91) und listet dabei vier Annahmen auf

1. $E(\varepsilon_i)=0$,
2. $var(\varepsilon_i) = \sigma^2$ (jeweils $\forall i$),
3. $cov(\varepsilon_i\varepsilon_j) = 0$ ($i \neq j$) und
4. each X_i is fixed across samples (entspr. C1 bei uns),

er fügt dann aber auch gleich einen Abschnitt "Replacing Fixed X's with Stochastic X's" (S. 93) ein, geht aber auf diesen Fall stochast. Regressoren erst später (ab S. 507) ein.

Er unterscheidet ein "First Gauss Markov Theorem" und das allgemeinere "Second…" oder "More General… Theorem " (S. 221, 383f), dass dann $\hat{Y}$ "the BLUE estimator"[4] für gleich dem bedingten Erwartungswert $E(Y|X_1,\ldots, X_k)$ ist (S. 221). Noch etwas allgemeiner (S. 383) wird erklärt, dass das Theorem nicht nur für LS Schätzer $\hat{\beta}_1,\ldots, \hat{\beta}_k$, sondern auch für Linearkombinationen dieser Schätzer gilt. Es wird auch (m.W. nur hier) darauf hingewiesen, dass bestimmte Eigenschaften der LS Methode als Entsprechungen (die SRF betreffend) von Annahmen, die – wie alle Annahmen – die PRF betreffen interpretiert werden können, so etwa $\frac{1}{n}\sum\hat{\varepsilon}_i = 0$ (bei uns $\frac{1}{T}\sum\hat{u}_t = 0$) in Analogie zu Annahme 1 (es gibt aber z.B. nichts analog zu Annahme 3).

W. H. Greene, Econometric Analysis, A Modern Introduction

Ein sehr ausführliches und gutes (aber auch besonders anspruchsvolles) Buch, das dem Problem dieser Notiz sogar ein eigenes Kapitel (Kap. 6. The Classical Multiple Linear Regression Model, S. 210ff) widmet und die Annahmen sehr ausführlich kommentiert. Die Darstellung erfolgt aber durchwegs in Matrixschreibweise und sowohl finite sample als auch large sample (asymptotische) Eigenschaften der LS Schätzer werden detailliert hergeleitet. Das Buch liefert damit also mehr als andere Bücher auch Begründungen dafür, warum und an welcher Stelle bei welchen Herleitungen bestimmte Annahmen erforderlich sind und es zeigt auch mehr als andere Bücher, welche Beziehungen zwischen den Annahmen bestehen.

Das Modell lautet $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Bei K Regressoren und n Beobachtungen haben die Spaltenvektoren $\mathbf{y}$ und $\boldsymbol{\varepsilon}$ die Dimension $n\times1$, der Parametervektor $\boldsymbol{\beta}$ ist ein $K\times1$ Vektor und die X ist eine $n\times K$ Matrix. Die Annahmen von William H. Greene sind (S. 213ff):

1. Linearity of the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. (im Sinne von intrisisch linear, vgl. Vorles.; viele entspr. Modelle werden gezeigt: log-log [constant elasticity] semilog [log-lin] und translog model (S. 217)
2. $\mathbf{X}$ is full (column) rank (nennt das identification condition[5]); ist bei $n < K$ nicht erfüllt
3. $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$ (mean of each ε conditional on all observations is zero) implies $cov[\boldsymbol{\varepsilon}|\mathbf{X}] = 0$ (Unkorreliertheit, nicht notwendig Unabhängigkeit) und ist eine strengere Forderung als $E(\varepsilon_i) = 0$ bei nichtstochast. Regressoren.[6]
4. $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\mathbf{X}] = \boldsymbol{\Omega} = \sigma^2\mathbf{I}$ (mit $\boldsymbol{\Omega}$ als Varianz-Kovarianz-Matrix der Störgrößen; das entspricht unseren Annahmen B2 [Homoskedastizität] und B3 [keine Autokorrelation])
5. klassisches Modell (nonstochast. regressors): $\mathbf{X}$ is a nonstochastic matrix (a known $n\times K$ matrix of constants[7]). Wenn dies gilt folgt: Annahmen 3 und 4 "can be made

[4] eigentlich steht ja das E in BLU<u>E</u> schon für estimator.
[5] sehr missverständlich; betrifft Berechenbarkeit der Parameter, nicht (wie sonst üblich) Identifikation bei einem Mehrgleichungsmodell
[6] Wenn $E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$ für alle Beobachtungen $\mathbf{X}$, dann auch $E(\varepsilon_i) = 0$ für alle i (aber Umkehrung gilt nicht). Gute Interpretation: wenn Annahme 3 nicht erfüllt, dann sind bei einfacher Regression $y = \alpha+\beta x + \varepsilon$ und $y = \alpha^*+\beta x + \varepsilon^*$ beobachtungsäquivalent mit $\alpha^* = \alpha + \mu$ und $\varepsilon^* = \varepsilon - \mu$.
[7] Bietet auch folgende alternative Sprachregelung an: observations on X are fixed in repeated samples

unconditional, also E[$\boldsymbol{\varepsilon}$] = **0** und E[$\boldsymbol{\varepsilon\varepsilon}'$] = $\sigma^2\mathbf{I}$ und $\lim(\frac{1}{n}\mathbf{X}'\mathbf{X}) = \mathbf{Q}$ gilt statt (bei stoch. Regress.) $\text{plim}(\frac{1}{n}\mathbf{X}'\mathbf{X}) = \mathbf{Q}$ mit einer positiv definiten Matrix **Q**.[8]

6. Normality $\boldsymbol{\varepsilon}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$; bemerkt, dass diese Annahme von einigen Autoren für überflüssig gehalten wird (sie ist nicht notwendig für den Beweis der Optimalitätseigenschaften der LS Schätzer, wohl aber "useful in constructing test statistics" (und confidence intervals) (S. 222).

J. H. Stock, M. W. Watson, Introduction to Econometrics

Einfach geschrieben (ohne Matrixalgebra, insofern viel leichter als z.B. Greene) und mit Hinweisen auf Anwendungen, Fachdiskussionen, Wirtschaftspolitik. Damit vom Stil ähnlich wie Murray aber übersichtlicher (aber immer noch viel weniger übersichtlich als die nächsten beiden Bücher, die allerdings schon etwas älter sind). Geht in vielen Punkten eigene Wege (etwa in Abschn. 9.1 und v.a. auch in dem hier allein interessierenden Punkt der Modellannahmen). Viele verbale Erklärungen, die nicht selten nicht sehr klar sind. Zwar nicht unbekannt, aber offenbar nicht mainstream und oft eher unorthodox. Sehr beispiel-orientiert, aber sehr fraglich, ob damit der rote Faden stets erkennbar bleibt und dem Leser die korrekte Einordnung der Gegenstände nach der sonst üblichen Systematik der Lehrbücher gelingt.

Bei der einfachen Regression werden drei Annahmen (S. 126ff), zu denen im Fall der multiplen Regression (S. 202ff) noch die Annahme 4 hinzukommt:

1. The conditional distribution of u_i given X_i, bzw. $X_{1i}, X_{2i}, \ldots, X_{ki}$ has a mean of zero (gemeint ist wohl expectation of zero); dies entspricht Annahme 3 bei Gujarati/Porter und 3 bei Greene[9]
2. ($X_{1i}, X_{2i}, \ldots X_{ki}, Y_i$) are i.i.d. Damit werden unabhängige Züge aus einer identischen k+1 dimensionalen Verteilung gefordert (es wird zutreffend bemerkt, dass diese Forderung, die sonst in der Literatur nicht speziell erwähnt wird, darauf hinausläuft, dass ein simple random sample gezogen wurde;[10] es ist also eine Frage der Qualität der Stichprobe); ähnlich speziell (in anderen Lehrbüchern nicht zu finden) ist auch die folgende dritte Annahme
3. Large outliers are unlikely (die Annahme scheint mir schon deshalb problemtisch zu sein, weil nicht klar ist, ab welcher Wahrscheinlichkeit etwas von likely zu unlikely wird)[11]; üblicher ist dagegen die Forderung
4. No perfect multicollinearity.

Die beiden folgenden älteren Lehrbücher scheinen mir im Aufbau und in der Klarheit der Erläuterungen besser zu sein als viele modernere Bücher. Die Darstellung wird jedoch, anders

---

[8] Positiv definit heißt, dass die quadratische Form (ein Skalar) x'Qx positiv (größer als 0) ist.

[9] Es ist auch richtig, dass dies hinausläuft auf "assuming that the population regression line is the conditional mean of Y_i given X_i" (S. 127. aber warum hier wieder mean statt korrekt expection und E(Y|X = x)?), oder dass dies "the key assumption that makes the OLS estimators unbiased" (S.203) ist

[10] "This assumption holds automatically if the data are collected by simple random sampling" (S. 203). Man findet die Annahme sonst in der Literatur nicht, weil es offenbar als selbstverständlich beim Schätzen und Testen vorausgesetzt wird {was allerdings in der Tat bei Anwendungen oft nicht gerechtfertigt ist}, dass eine "echte" Stichprobe (Zufallsauswahl, random sample) gezogen wurde. Problematisch bzw. Missverständnisse provozierend dürfte jedoch der Hinweis sein, dass die Annahme nicht erfüllt ist, wenn X_i nonrandom (also nichtstochastisch) ist, wie beim "classical linear regression model", das bei uns und z.B. bei Gujarati/Porter zunächst im Vordergrund der Betrachtung steht oder auch bei Zeitreihendaten (die ja quasi der Normalfall von Daten in der Ökonometrie sind) nicht erfüllt ist (S. 128f). Problematisch ist die Annahme (als eine speziell zu fordernde Annahme) auch, wenn es gleichzeitig heißt, dass die für "i.i.d. regressors" (wie öfter in diesem Buch, eine sehr schlampige Sprechweise: als ob die Regressoren $X_1, \ldots, X_k$ identisch verteilt wären) entwickelten Ergebnis "are also true if the regressors are nonrandom" (S. 129), also in einem Fall, in dem die Annahme 2 ausdrücklich nicht erfüllt sein soll. Wenn die Ergebnisse bei Nichterfüllung einer Annahme genauso gelten wie bei Erfüllung der Annahme, fragt es sich natürlich, warum man extra eine solche Annahme fordert.

[11] Sie wird jedoch dahingehend konkretisiert, dass endliche vierte Momente der Regressoren existieren (S. 129, dort wird aber auch $E(Y^4) < \infty$ gefordert). Das läuft auf das hinaus, was bei Greene unter Nr. 5 erwähnt wurde.

als Murray oder Stock/Watson weniger durch aktuelle Beispiele aus Anwendungen und Hinweise auf wirtschaftspolitische Themen aufgelockert.

J. Johnston, J. Dinardo, Econometric Methods

Dieses Buch (und auch das folgende von Maddala) enthält sehr kompakt und übersichtlich alle Gegenstände (auch weiterführende), die üblicherweise in den Lehrbüchern der Ökonometrie dargestellt werden; die Erklärungen sind jedoch i.d.R. kürzer (wobei ja gerade die Bücher von Murray und Stock/Watson zeigen, dass viele Worte die Sache nicht unbedingt klarere machen müssen) und in Matrixschreibweise. Dies gilt auch für die Annahmen:[12]

1. $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ (Linearität und korrekte Spezifikation; bei uns Annahmen A1 und A2 und jeweils die Annahme 1 bei Gujarati/Porter und Greene),
2. u_i are iid and $(0,\sigma^2)$ also white noise bzw. iid and $N(0,\sigma^2)$ (Gaussian white noise)[13] auch $E(\mathbf{uu}') =$,$\sigma^2\mathbf{I}$,
3. $E(X_{it},u_s)$ for all i = 1, …,k, t, und s (k stochastic regressors), bzw. X ist nonstochastic with full rank k (wenn less than full column rank k $\rightarrow$ collinearity).

G.S. Maddala, Introduction to Econometrics

Hier, wie auch in anderen, hier nicht speziell erwähnten Lehrbüchern stehen die bekannten Annahmen über die Störgröße (besser über die Verteilung der Störgrößen u_t t = 1, …, T, also bei uns die B-Annahmen) im Vordergrund. Maddala erwähnt hier 5 Annahmen S. 65, 128)

1. Zero mean $E(u_i) = 0$ for all i (entspricht B1 bei v. Auer und in unserer Vorlesung)
2. Common variance $\mathrm{var}(u_i) = \sigma^2$ for all i ($\forall$i), entspr. B2
3. Independence u_i and u_j for all $i \neq j$, entspr. B3
4. u_i and x_j are independent for all i and j. This assumption automatically follows if x_j are considered nonrandom (… the distribution of u does not depend on the value of x), das entspr. Annahme C1 bei uns
5. u_i are normally distributed (entspricht B4): "In conjunction with assumption 1,2, and 3 this implies that u_i are independently (wegen Annahme 3) and normally distributed $u_i \sim N(0,\sigma^2)$"

Bei multipler Regression kommt das hinzu, was bei uns Annahme C2 ist, nämlich "that x_1, x_2, …, x_k are not collinear, that is, there is no deterministic linear relationship among them". (S. 128).

Zwei deutsche Bücher und Fazit

Die Annahmen B1 bis B3 findet man genauso auch bei Eckey/Kosfeld/Dreger Ökonometrie (Gabler), S.20f. Sehr ausführlich mit einem ganzen Kapitel (Kap. 4 auf S. 58 – 66) ist in diesem Punkt das Buch von P. Hackl, Einführung in die Ökonometrie (Pearson Studium).

Man kann also festhalten: in allen Büchern (mit der einzigen Ausnahme jedoch von Stock & Watson) findet man im Prinzip die gleichen Auflistungen von Annahmen (insbesondere was die Annahmen B über die [Verteilung der] Störgrößen u_t betrifft). Es sind genau die Annahmen bei v. Auer und in unserer Vorlesung.

[12] Es gibt jedoch kurze, sehr klare und wohl auch für eine erste Bekanntschaft mit dem Thema voll ausreichende Bemerkungen (Erläuterungen) in Abschnitt. 4.1 "Specification error" unter den drei Überschriften "Possible Problems with u", "Possible Problems with X", und "Possible Problems with β".

[13] Entspricht den Annahmen 1-3 bei Gujarati/Porter, 1-3 bei Murray und 3-4 bei Greene