

Bitte tragen Sie Ihre Daten sorgfältig und leserlich ein:

Matrikelnummer

Nachname

Studiengang

Vorname

## Bearbeitungshinweise:

Diese Klausur besteht aus fünf Aufgaben, von welchen **vier Aufgaben Ihrer Wahl** zu bearbeiten sind.

Bearbeiten Sie alle Aufgaben, so werden nur die ersten vier bewertet.

Alle Antworten sind zu begründen.

Bitte verwenden sie einen Kugelschreiber oder keinen zu starken Filzstift.

Für jede der fünf Aufgaben sind maximal je 18 Punkte zu erreichen.

Die Bearbeitungszeit beträgt 90 Minuten.

## Erlaubte Hilfsmittel:

- Taschenrechner (nicht programmierbar)

**Viel Erfolg!**

Vom Prüfer auszufüllen:

Punkte Aufgabe 1  / 18

Punkte Aufgabe 2  / 18

Punkte Aufgabe 3  / 18

Punkte Aufgabe 4  / 18

Punkte Aufgabe 5  / 18

Gesamtpunkte  / 72

Note:

# Aufgabe 1 zu Kapitel 3 Multiple Lineare Regression

Betrachtet sei das Ökonometrische Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

unter den Annahmen **MLR 1** bis **MLR 5**.

- a) Was versteht man unter dem Residuum  $\hat{u}_i$ ? Geben Sie die Formel an und erklären Sie kurz, was es ökonomisch bedeutet.
- b) Sei  $\mathbf{b} \in \mathbb{R}^{K+1}$  eine Vermutung über  $\boldsymbol{\beta} \in \mathbb{R}^{K+1}$ . Wie lautet die Summe der quadrierten Residuen  $S(\mathbf{b} \mid \mathbf{y}, \mathbf{X})$  in Matrixschreibweise? Welche Größe wird bei der KQ-Schätzung minimiert?
- c) Wie lautet die Formel für den KQ-Schätzer  $\hat{\boldsymbol{\beta}}$ ? Welche Bedingung (MLR-Annahme) muss erfüllt sein, damit  $(\mathbf{X}'\mathbf{X})^{-1}$  existiert?
- d) Was bezeichnet man als „Momentenmatrix“? Geben Sie an, welche Ordnung (Zeilen  $\times$  Spalten) sie hat, wenn  $\mathbf{X}$  von der Ordnung  $n \times (K + 1)$  ist.
- e) Geben Sie die Formeln für  $SST$ ,  $SSE$  und  $SSR$  an. Wie lautet die Streuungszerlegung, die diese drei Größen verbindet?
- f) Wie ist das Bestimmtheitsmaß  $R^2$  definiert? Welchen Wertebereich kann  $R^2$  annehmen und wie interpretiert man  $R^2$  inhaltlich?

# Aufgabe 2 zu Kapitel 4 Inferenz

Betrachtet sei das Ökonometrische Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

unter den Annahmen **MLR 1** bis **MLR 6**.

- a) Welche Verteilung besitzt der OLS-Schätzer  $\hat{\boldsymbol{\beta}}$ , bedingt auf die Regressormatrix  $\mathbf{X}$ ?  
Nennen Sie ebenfalls den Erwartungswert  $E[\hat{\boldsymbol{\beta}}]$  und die Varianz-Kovarianz-Matrix  $\mathbb{V}(\hat{\boldsymbol{\beta}})$ .

- b) Welche Verteilung besitzt

$$\frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)}$$

Nennen Sie ebenfalls den Erwartungswert und die Anzahl der Freiheitsgrade.

- c) Gegeben ist folgende Schätzung für die abhängige Variable *colGPA* bei 141 Beobachtungen:

| Variable | Koeffizient | Std. Fehler |
|----------|-------------|-------------|
| const    | 1.390       | 0.331       |
| hsGPA    | 0.412       | 0.094       |
| ACT      | 0.015       | 0.011       |
| skipped  | -0.083      | 0.026       |

Berechnen Sie die t-Statistik für den Parameter  $\beta_3$  für *skipped* unter der Nullhypothese  $H_0 : \beta_3 = 0$  mit  $H_1 : \beta_3 \neq 0$  an und testen Sie diese Nullhypothese auf dem 5%-Niveau.

- d) Berechnen Sie das symmetrische 95%-Konfidenzintervall für  $\beta_3$  und interpretieren Sie dieses Intervall.
- e) Wie lautet die Nullhypothese des Globalen F-Tests?  
Geben Sie Teststatistik in Abhängigkeit von  $R^2$  an.
- f) Für das Modell aus Teil c) betrage das Bestimmtheitsmaß  $R^2 = 0,234$ . Testen Sie die Nullhypothese des globalen F-Tests auf dem 5%-Niveau.

# Aufgabe 3 zu Kapitel 8 Heteroskedastizität

Betrachtet sei das Ökonometrische Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

unter den Annahmen **MLR 1** bis **MLR 4**.

- a) Definieren Sie reine Heteroskedastizität bei unkorrelierten Fehlertermen formal und stellen Sie die Varianz-Kovarianz-Matrix  $\Sigma(\mathbf{u}|\mathbf{X}) = \sigma^2\boldsymbol{\Omega}$  für diesen Fall dar.
- b) Geben Sie die Matrix  $\mathbf{P} = \boldsymbol{\Omega}^{-\frac{1}{2}}$  an, die zur GLS-Transformation verwendet wird und geben Sie an, wie  $y_i$ ,  $x_{ij}$  und  $u_i$  transformiert werden müssen, damit das transformierte Modell die Annahmen **MLR 1** bis **MLR 5** erfüllt.
- c) Formulieren Sie die Nullhypothese des Breusch-Pagan-Tests und begründen Sie, warum unter Homoskedastizität  $E[u_i^2|X] = \sigma^2$  gilt.
- d) Stellen Sie die Hilfsregression des Breusch-Pagan-Tests auf und formulieren Sie die zu testenden Restriktionen.
- e) Beschreiben Sie die Hilfsregression des White-Tests für ein Modell mit drei Regressoren.
- f) Erläutern Sie Vorteile und Nachteile des White-Tests gegenüber dem Breusch-Pagan-Test.

# Aufgabe 4 zu Kapitel 12 Serielle Korrelation in Zeitreihen

Betrachtet sei das Ökonometrische Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

unter den Annahmen **MLR 1** bis **MLR 4**.

- a) Erklären Sie, warum der OLS-Schätzer  $\hat{\boldsymbol{\beta}}$  weiterhin unverzerrt ist.
- b) Erläutern Sie, warum die üblichen OLS-Standardfehler und Tests nicht valide sind.
- c) Erklären Sie, warum OLS bei serieller Korrelation im Allgemeinen nicht effizient ist.

In einer Zeitreihenregression werden OLS-Residuen auf ihren ersten Lag regressiert, und es ergibt sich

$$\hat{u}_t = -0,1118 + 0,5725\hat{u}_{t-1}.$$

Der p-Wert des Koeffizienten von  $\hat{u}_{t-1}$  beträgt  $2,43 \cdot 10^{-6}$ .

- d) Interpretieren Sie den geschätzten Wert 0,5725 und beurteilen Sie, ob ein Hinweis auf Autokorrelation vorliegt.
- e) Erklären Sie, welche Probleme mit dem t-Test in dieser Hilfsregression einher gehen.
- f) Nennen Sie zwei alternative beziehungsweise ergänzende Verfahren zur Diagnose serieller Korrelation. Beschreiben Sie kurz das jeweilige Vorgehen.

# Aufgabe 5 zu Kapitel 15 Instrumente

Betrachten Sie die wahre Lohnregression

$$\ln(w_i) = \gamma_0 + \gamma_1 educ_i + \gamma_2 a_i + e_i,$$

wobei  $a_i$  individuelle Fähigkeiten beschreibt.

Angenommen,  $a_i$  ist unbeobachtbar und es wird stattdessen das reduzierte Modell

$$\ln(w_i) = \beta_0 + \beta_1 educ_i + u_i$$

geschätzt.

- a) Drücken Sie den Fehlerterm  $u_i$  des reduzierten Modells durch  $a_i$  und  $e_i$  aus und leiten Sie  $Cov(educ_i, u_i)$  her.
- b) Nennen Sie die Bedingungen, unter denen  $educ_i$  endogen wäre, und erörtern Sie, ob diese Bedingungen im Lohnbeispiel plausibel erfüllt sein könnten.
- c) Definieren Sie die beiden Bedingungen, die eine gültige Instrumentalvariable  $z_i$  für  $educ_i$  erfüllen muss und erklären Sie, welche dieser Bedingungen statistisch getestet werden kann und welche theoretisch begründet werden muss.
- d) Leiten Sie aus  $Cov(\ln(w_i), z_i)$  den theoretischen IV-Ausdruck für  $\beta_1$  her.
- e) Geben Sie den einfachen IV-Schätzer  $\hat{\beta}_1^{IV}$  auf Basis von Stichprobenkovarianzen an und erklären Sie, warum OLS ein Spezialfall von IV ist.
- f) Erklären Sie das Weak-Instruments-Problem.





















## Kritische Werte der $t$ -Verteilung

|                     |          | Signifikanzniveau         |            |           |            |          |
|---------------------|----------|---------------------------|------------|-----------|------------|----------|
|                     |          | einseitig:<br>zweiseitig: | 10%<br>20% | 5%<br>10% | 2,5%<br>5% | 1%<br>2% |
| Freiheits-<br>grade | 1        | 3,078                     | 6,314      | 12,706    | 31,821     | 63,657   |
|                     | 2        | 1,886                     | 2,920      | 4,303     | 6,965      | 9,925    |
|                     | 3        | 1,638                     | 2,353      | 3,182     | 4,541      | 5,841    |
|                     | 4        | 1,533                     | 2,132      | 2,776     | 3,747      | 4,604    |
|                     | 5        | 1,476                     | 2,015      | 2,571     | 3,365      | 4,032    |
|                     | 6        | 1,440                     | 1,943      | 2,447     | 3,143      | 3,707    |
|                     | 7        | 1,415                     | 1,895      | 2,365     | 2,998      | 3,499    |
|                     | 8        | 1,397                     | 1,860      | 2,306     | 2,896      | 3,355    |
|                     | 9        | 1,383                     | 1,833      | 2,262     | 2,821      | 3,250    |
|                     | 10       | 1,372                     | 1,812      | 2,228     | 2,764      | 3,169    |
|                     | 11       | 1,363                     | 1,796      | 2,201     | 2,718      | 3,106    |
|                     | 12       | 1,356                     | 1,782      | 2,179     | 2,681      | 3,055    |
|                     | 13       | 1,350                     | 1,771      | 2,160     | 2,650      | 3,012    |
|                     | 14       | 1,345                     | 1,761      | 2,145     | 2,624      | 2,977    |
|                     | 15       | 1,341                     | 1,753      | 2,131     | 2,602      | 2,947    |
|                     | 16       | 1,337                     | 1,746      | 2,120     | 2,583      | 2,921    |
|                     | 17       | 1,333                     | 1,740      | 2,110     | 2,567      | 2,898    |
|                     | 18       | 1,330                     | 1,734      | 2,101     | 2,552      | 2,878    |
|                     | 19       | 1,328                     | 1,729      | 2,093     | 2,539      | 2,861    |
|                     | 20       | 1,325                     | 1,725      | 2,086     | 2,528      | 2,845    |
|                     | 21       | 1,323                     | 1,721      | 2,080     | 2,518      | 2,831    |
|                     | 22       | 1,321                     | 1,717      | 2,074     | 2,508      | 2,819    |
|                     | 23       | 1,319                     | 1,714      | 2,069     | 2,500      | 2,807    |
|                     | 24       | 1,318                     | 1,711      | 2,064     | 2,492      | 2,797    |
|                     | 25       | 1,316                     | 1,708      | 2,060     | 2,485      | 2,787    |
|                     | 26       | 1,315                     | 1,706      | 2,056     | 2,479      | 2,779    |
|                     | 27       | 1,314                     | 1,703      | 2,052     | 2,473      | 2,771    |
|                     | 28       | 1,313                     | 1,701      | 2,048     | 2,467      | 2,763    |
|                     | 29       | 1,311                     | 1,699      | 2,045     | 2,462      | 2,756    |
|                     | 30       | 1,310                     | 1,697      | 2,042     | 2,457      | 2,750    |
|                     | 40       | 1,303                     | 1,684      | 2,021     | 2,423      | 2,704    |
|                     | 60       | 1,296                     | 1,671      | 2,000     | 2,390      | 2,660    |
|                     | 90       | 1,291                     | 1,662      | 1,987     | 2,368      | 2,632    |
|                     | 120      | 1,289                     | 1,658      | 1,980     | 2,358      | 2,617    |
|                     | $\infty$ | 1,282                     | 1,645      | 1,960     | 2,326      | 2,576    |

# Kritische Werte der F-Verteilung zum Signifikanzniveau von 1%

|             |          | Anzahl der Restriktionen |      |      |      |      |      |      |      |      |      |
|-------------|----------|--------------------------|------|------|------|------|------|------|------|------|------|
|             |          | 1                        | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| $n - k - 1$ | 10       | 10,04                    | 7,56 | 6,55 | 5,99 | 5,64 | 5,39 | 5,20 | 5,06 | 4,94 | 4,85 |
|             | 11       | 9,65                     | 7,21 | 6,22 | 5,67 | 5,32 | 5,07 | 4,89 | 4,74 | 4,63 | 4,54 |
|             | 12       | 9,33                     | 6,93 | 5,95 | 5,41 | 5,06 | 4,82 | 4,64 | 4,50 | 4,39 | 4,30 |
|             | 13       | 9,07                     | 6,70 | 5,74 | 5,21 | 4,86 | 4,62 | 4,44 | 4,30 | 4,19 | 4,10 |
|             | 14       | 8,86                     | 6,51 | 5,56 | 5,04 | 4,69 | 4,46 | 4,28 | 4,14 | 4,03 | 3,94 |
|             | 15       | 8,68                     | 6,36 | 5,42 | 4,89 | 4,56 | 4,32 | 4,14 | 4,00 | 3,89 | 3,80 |
|             | 16       | 8,53                     | 6,23 | 5,29 | 4,77 | 4,44 | 4,20 | 4,03 | 3,89 | 3,78 | 3,69 |
|             | 17       | 8,40                     | 6,11 | 5,18 | 4,67 | 4,34 | 4,10 | 3,93 | 3,79 | 3,68 | 3,59 |
|             | 18       | 8,29                     | 6,01 | 5,09 | 4,58 | 4,25 | 4,01 | 3,84 | 3,71 | 3,60 | 3,51 |
|             | 19       | 8,18                     | 5,93 | 5,01 | 4,50 | 4,17 | 3,94 | 3,77 | 3,63 | 3,52 | 3,43 |
|             | 20       | 8,10                     | 5,85 | 4,94 | 4,43 | 4,10 | 3,87 | 3,70 | 3,56 | 3,46 | 3,37 |
|             | 21       | 8,02                     | 5,78 | 4,87 | 4,37 | 4,04 | 3,81 | 3,64 | 3,51 | 3,40 | 3,31 |
|             | 22       | 7,95                     | 5,72 | 4,82 | 4,31 | 3,99 | 3,76 | 3,59 | 3,45 | 3,35 | 3,26 |
|             | 23       | 7,88                     | 5,66 | 4,76 | 4,26 | 3,94 | 3,71 | 3,54 | 3,41 | 3,30 | 3,21 |
|             | 24       | 7,82                     | 5,61 | 4,72 | 4,22 | 3,90 | 3,67 | 3,50 | 3,36 | 3,26 | 3,17 |
|             | 25       | 7,77                     | 5,57 | 4,68 | 4,18 | 3,85 | 3,63 | 3,46 | 3,32 | 3,22 | 3,13 |
|             | 26       | 7,72                     | 5,53 | 4,64 | 4,14 | 3,82 | 3,59 | 3,42 | 3,29 | 3,18 | 3,09 |
|             | 27       | 7,68                     | 5,49 | 4,60 | 4,11 | 3,78 | 3,56 | 3,39 | 3,26 | 3,15 | 3,06 |
|             | 28       | 7,64                     | 5,45 | 4,57 | 4,07 | 3,75 | 3,53 | 3,36 | 3,23 | 3,12 | 3,03 |
|             | 29       | 7,60                     | 5,42 | 4,54 | 4,04 | 3,73 | 3,50 | 3,33 | 3,20 | 3,09 | 3,00 |
|             | 30       | 7,56                     | 5,39 | 4,51 | 4,02 | 3,70 | 3,47 | 3,30 | 3,17 | 3,07 | 2,98 |
|             | 40       | 7,31                     | 5,18 | 4,31 | 3,83 | 3,51 | 3,29 | 3,12 | 2,99 | 2,89 | 2,80 |
|             | 60       | 7,08                     | 4,98 | 4,13 | 3,65 | 3,34 | 3,12 | 2,95 | 2,82 | 2,72 | 2,63 |
|             | 90       | 6,93                     | 4,85 | 4,01 | 3,54 | 3,23 | 3,01 | 2,84 | 2,72 | 2,61 | 2,52 |
|             | 120      | 6,85                     | 4,79 | 3,95 | 3,48 | 3,17 | 2,96 | 2,79 | 2,66 | 2,56 | 2,47 |
|             | $\infty$ | 6,63                     | 4,61 | 3,78 | 3,32 | 3,02 | 2,80 | 2,64 | 2,51 | 2,41 | 2,32 |

## Kritische Werte der F-Verteilung zum Signifikanzniveau von 5%

|             |          | Anzahl der Restriktionen |      |      |      |      |      |      |      |      |      |
|-------------|----------|--------------------------|------|------|------|------|------|------|------|------|------|
|             |          | 1                        | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
| $n - k - 1$ | 10       | 4,96                     | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,14 | 3,07 | 3,02 | 2,98 |
|             | 11       | 4,84                     | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 3,01 | 2,95 | 2,90 | 2,85 |
|             | 12       | 4,75                     | 3,89 | 3,49 | 3,26 | 3,11 | 3,00 | 2,91 | 2,85 | 2,80 | 2,75 |
|             | 13       | 4,67                     | 3,81 | 3,41 | 3,18 | 3,03 | 2,92 | 2,83 | 2,77 | 2,71 | 2,67 |
|             | 14       | 4,60                     | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,76 | 2,70 | 2,65 | 2,60 |
|             | 15       | 4,54                     | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,71 | 2,64 | 2,59 | 2,54 |
|             | 16       | 4,49                     | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 |
|             | 17       | 4,45                     | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,61 | 2,55 | 2,49 | 2,45 |
|             | 18       | 4,41                     | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 |
|             | 19       | 4,38                     | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,54 | 2,48 | 2,42 | 2,38 |
|             | 20       | 4,35                     | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,51 | 2,45 | 2,39 | 2,35 |
|             | 21       | 4,32                     | 3,47 | 3,07 | 2,84 | 2,68 | 2,57 | 2,49 | 2,42 | 2,37 | 2,32 |
|             | 22       | 4,30                     | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,46 | 2,40 | 2,34 | 2,30 |
|             | 23       | 4,28                     | 3,42 | 3,03 | 2,80 | 2,64 | 2,53 | 2,44 | 2,37 | 2,32 | 2,27 |
|             | 24       | 4,26                     | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,42 | 2,36 | 2,30 | 2,25 |
|             | 25       | 4,24                     | 3,39 | 2,99 | 2,76 | 2,60 | 2,49 | 2,40 | 2,34 | 2,28 | 2,24 |
|             | 26       | 4,23                     | 3,37 | 2,98 | 2,74 | 2,59 | 2,47 | 2,39 | 2,32 | 2,27 | 2,22 |
|             | 27       | 4,21                     | 3,35 | 2,96 | 2,73 | 2,57 | 2,46 | 2,37 | 2,31 | 2,25 | 2,20 |
|             | 28       | 4,20                     | 3,34 | 2,95 | 2,71 | 2,56 | 2,45 | 2,36 | 2,29 | 2,24 | 2,19 |
|             | 29       | 4,18                     | 3,33 | 2,93 | 2,70 | 2,55 | 2,43 | 2,35 | 2,28 | 2,22 | 2,18 |
|             | 30       | 4,17                     | 3,32 | 2,92 | 2,69 | 2,53 | 2,42 | 2,33 | 2,27 | 2,21 | 2,16 |
|             | 40       | 4,08                     | 3,23 | 2,84 | 2,61 | 2,45 | 2,34 | 2,25 | 2,18 | 2,12 | 2,08 |
|             | 60       | 4,00                     | 3,15 | 2,76 | 2,53 | 2,37 | 2,25 | 2,17 | 2,10 | 2,04 | 1,99 |
|             | 90       | 3,95                     | 3,10 | 2,71 | 2,47 | 2,32 | 2,20 | 2,11 | 2,04 | 1,99 | 1,94 |
|             | 120      | 3,92                     | 3,07 | 2,68 | 2,45 | 2,29 | 2,17 | 2,09 | 2,02 | 1,96 | 1,91 |
|             | $\infty$ | 3,84                     | 3,00 | 2,60 | 2,37 | 2,21 | 2,10 | 2,01 | 1,94 | 1,88 | 1,83 |

## Kritische Werte der $\chi^2$ -Verteilung

|                     |    | Signifikanzniveau |       |       |
|---------------------|----|-------------------|-------|-------|
|                     |    | 10%               | 5%    | 1%    |
| Freiheits-<br>Grade | 1  | 2,71              | 3,84  | 6,63  |
|                     | 2  | 4,61              | 5,99  | 9,21  |
|                     | 3  | 6,25              | 7,81  | 11,34 |
|                     | 4  | 7,78              | 9,49  | 13,28 |
|                     | 5  | 9,24              | 11,07 | 15,09 |
|                     | 6  | 10,64             | 12,59 | 16,81 |
|                     | 7  | 12,02             | 14,07 | 18,48 |
|                     | 8  | 13,36             | 15,51 | 20,09 |
|                     | 9  | 14,68             | 16,92 | 21,67 |
|                     | 10 | 15,99             | 18,31 | 23,21 |
|                     | 11 | 17,28             | 19,68 | 24,72 |
|                     | 12 | 18,55             | 21,03 | 26,22 |
|                     | 13 | 19,81             | 22,36 | 27,69 |
|                     | 14 | 21,06             | 23,68 | 29,14 |
|                     | 15 | 22,31             | 25,00 | 30,58 |
|                     | 16 | 23,54             | 26,30 | 32,00 |
|                     | 17 | 24,77             | 27,59 | 33,41 |
|                     | 18 | 25,99             | 28,87 | 34,81 |
|                     | 19 | 27,20             | 30,14 | 36,19 |
|                     | 20 | 28,41             | 31,41 | 37,57 |
|                     | 21 | 29,62             | 32,67 | 38,93 |
|                     | 22 | 30,81             | 33,92 | 40,29 |
|                     | 23 | 32,01             | 35,17 | 41,64 |
|                     | 24 | 33,20             | 36,42 | 42,98 |
|                     | 25 | 34,38             | 37,65 | 44,31 |
|                     | 26 | 35,56             | 38,89 | 45,64 |
|                     | 27 | 36,74             | 40,11 | 46,96 |
|                     | 28 | 37,92             | 41,34 | 48,28 |
|                     | 29 | 39,09             | 42,56 | 49,59 |
|                     | 30 | 40,26             | 43,77 | 50,89 |

## Aufgabe 1: Multiple Lineare Regression

Betrachtet wird das ökonometrische Modell

$$y = X\beta + u$$

unter den Annahmen MLR 1 bis MLR 5.

### a) Residuum $\hat{u}_i$

Das Residuum  $\hat{u}_i$  ist die Differenz zwischen dem beobachteten Wert  $y_i$  und dem durch das Modell geschätzten Wert  $\hat{y}_i$ .

Es gilt

$$\hat{u}_i = y_i - \hat{y}_i.$$

Da

$$\hat{y}_i = x_i' \hat{\beta}$$

gilt, kann man auch schreiben

$$\hat{u}_i = y_i - x_i' \hat{\beta}.$$

In Matrixschreibweise lautet der Residuenvektor

$$\hat{u} = y - X\hat{\beta}.$$

Ökonomisch misst das Residuum den Teil der abhängigen Variable, der durch die im Modell enthaltenen Regressoren nicht erklärt wird.

Ein positives Residuum bedeutet, dass der beobachtete Wert größer ist als der vorhergesagte Wert.

Ein negatives Residuum bedeutet, dass der beobachtete Wert kleiner ist als der vorhergesagte Wert.

**Bewertung: 3 Punkte**

- 1 Punkt für die Formel  $\hat{u}_i = y_i - \hat{y}_i$ .
- 1 Punkt für die Darstellung mit  $x_i' \hat{\beta}$  oder in Matrixform.
- 1 Punkt für die ökonomische Interpretation.

### b) Summe der quadrierten Residuen

Für eine Vermutung  $b \in \mathbb{R}^{K+1}$  über  $\beta \in \mathbb{R}^{K+1}$  lautet der Residuenvektor

$$y - Xb.$$

Die Summe der quadrierten Residuen ist

$$S(b|y, X) = (y - Xb)'(y - Xb).$$

Äquivalent gilt

$$S(b|y, X) = \sum_{i=1}^n (y_i - x_i' b)^2.$$

Bei der KQ-Schätzung wird diese Summe der quadrierten Residuen minimiert.

Der KQ-Schätzer ist daher definiert als

$$\hat{\beta} = \arg \min_{b \in \mathbb{R}^{K+1}} S(b|y, X).$$

**Bewertung: 3 Punkte**

- 1 Punkt für den Residuenvektor  $y - Xb$ .
- 1 Punkt für  $S(b|y, X) = (y - Xb)'(y - Xb)$ .
- 1 Punkt für die Minimierung dieser Größe.

**c) KQ-Schätzer und Existenz von  $(X'X)^{-1}$**

Der KQ-Schätzer lautet

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

Damit  $(X'X)^{-1}$  existiert, muss  $X$  vollen Spaltenrang besitzen.  
Die entsprechende Annahme ist MLR 3.  
Formal gilt

$$rk(X) = K + 1.$$

Diese Bedingung schließt perfekte Multikollinearität aus.

**Bewertung: 3 Punkte**

- 1 Punkt für die Formel  $\hat{\beta} = (X'X)^{-1}X'y$ .
- 1 Punkt für die Bedingung  $rk(X) = K + 1$ .
- 1 Punkt für die Nennung von MLR 3 oder den Ausschluss perfekter Multikollinearität.

**d) Momentenmatrix**

Als Momentenmatrix bezeichnet man in diesem Kontext typischerweise die Matrix

$$X'X$$

oder die normierte Variante

$$\frac{1}{n}X'X.$$

Diese Matrix enthält die Kreuzprodukte beziehungsweise Stichprobenmomente der Regressoren.  
Wenn  $X$  die Ordnung

$$n \times (K + 1)$$

hat, dann hat  $X'$  die Ordnung

$$(K + 1) \times n.$$

Daher hat  $X'X$  die Ordnung

$$(K + 1) \times (K + 1).$$

Auch  $\frac{1}{n}X'X$  hat die Ordnung  $(K + 1) \times (K + 1)$ .

**Bewertung: 3 Punkte**

- 1 Punkt für die Nennung von  $X'X$  oder  $\frac{1}{n}X'X$ .
- 1 Punkt für die Erklärung als Kreuzprodukt- beziehungsweise Momentenmatrix der Regressoren.
- 1 Punkt für die korrekte Ordnung  $(K + 1) \times (K + 1)$ .

## e) SST, SSE, SSR und Streuungszerlegung

Die totale Quadratsumme lautet

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2.$$

Die erklärte Quadratsumme lautet

$$SSE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

Die Residuenquadratsumme lautet

$$SSR = \sum_{i=1}^n \hat{u}_i^2.$$

Die Streuungszerlegung lautet

$$SST = SSE + SSR.$$

Diese Zerlegung gilt im OLS-Modell mit Konstante.

**Bewertung: 3 Punkte**

- 1 Punkt für  $SST$ .
- 1 Punkt für  $SSE$  und  $SSR$ .
- 1 Punkt für  $SST = SSE + SSR$ .

## f) Bestimmtheitsmaß $R^2$

Das Bestimmtheitsmaß ist definiert als

$$R^2 = \frac{SSE}{SST}.$$

Äquivalent gilt

$$R^2 = 1 - \frac{SSR}{SST}.$$

Im Modell mit Konstante gilt

$$0 \leq R^2 \leq 1.$$

Das Bestimmtheitsmaß misst den Anteil der gesamten Variation der abhängigen Variable, der durch das Regressionsmodell erklärt wird.

Ein  $R^2$  von 0,60 bedeutet beispielsweise, dass 60 Prozent der Variation von  $y$  durch die Regressoren erklärt werden.

Ein hohes  $R^2$  impliziert jedoch keine kausale Interpretation der Koeffizienten.

**Bewertung: 3 Punkte**

- 1 Punkt für die Formel.
- 1 Punkt für den Wertebereich.
- 1 Punkt für die inhaltliche Interpretation.

## Aufgabe 2: Inferenz

Betrachtet wird das ökonometrische Modell

$$y = X\beta + u$$

unter den Annahmen MLR 1 bis MLR 6.

### a) Verteilung des OLS-Schätzers

Unter MLR 1 bis MLR 6 besitzt der OLS-Schätzer bedingt auf  $X$  die Verteilung

$$\hat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1}).$$

Der Erwartungswert lautet

$$E[\hat{\beta}|X] = \beta.$$

Die Varianz-Kovarianz-Matrix lautet

$$\Sigma(\hat{\beta}|X) = \text{Var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}.$$

Die Normalverteilung folgt aus der Normalitätsannahme für die Fehlerterme.

**Bewertung: 3 Punkte**

- 1 Punkt für die Normalverteilung.
- 1 Punkt für  $E[\hat{\beta}|X] = \beta$ .
- 1 Punkt für  $\sigma^2(X'X)^{-1}$ .

### b) Verteilung der t-Statistik

Die Größe

$$\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)}$$

ist t-verteilt.

Es gilt

$$\frac{\hat{\beta}_j - \beta_j}{\text{se}(\hat{\beta}_j)} \sim t_{n-K-1}.$$

Der Erwartungswert dieser t-Verteilung ist null.

Die Anzahl der Freiheitsgrade beträgt

$$df = n - K - 1.$$

**Bewertung: 3 Punkte**

- 1 Punkt für die t-Verteilung.
- 1 Punkt für den Erwartungswert 0.
- 1 Punkt für  $n - K - 1$  Freiheitsgrade.

### c) t-Test für den Parameter von skipped

Für die Variable `skipped` gilt

$$\hat{\beta}_3 = -0.083$$

und

$$se(\hat{\beta}_3) = 0.026.$$

Getestet wird

$$H_0 : \beta_3 = 0$$

gegen

$$H_1 : \beta_3 \neq 0.$$

Die Teststatistik lautet

$$t = \frac{\hat{\beta}_3 - 0}{se(\hat{\beta}_3)} = \frac{-0.083}{0.026} \approx -3.19.$$

Es gibt  $n = 141$  Beobachtungen und  $K = 3$  Regressoren ohne Konstante.

Die Freiheitsgrade lauten daher

$$df = 141 - 3 - 1 = 137.$$

Für einen zweiseitigen Test auf dem 5-Prozent-Niveau liegt der kritische Wert näherungsweise bei 1.98.

Da

$$|t| = 3.19 > 1.98$$

gilt, wird die Nullhypothese verworfen.

Die Variable `skipped` ist auf dem 5-Prozent-Niveau statistisch signifikant von null verschieden.

Der negative Koeffizient spricht dafür, dass mehr ausgelassene Veranstaltungen mit einer geringeren `colGPA` verbunden sind.

**Bewertung: 3 Punkte**

- 1 Punkt für  $t \approx -3.19$ .
- 1 Punkt für den Vergleich mit dem kritischen Wert.
- 1 Punkt für die korrekte Entscheidung und Interpretation.

### d) Symmetrisches 95-Prozent-Konfidenzintervall

Das symmetrische 95-Prozent-Konfidenzintervall lautet

$$\hat{\beta}_3 \pm c \cdot se(\hat{\beta}_3).$$

Mit  $df = 137$  kann näherungsweise

$$c \approx 1.98$$

verwendet werden.

Damit ergibt sich

$$-0.083 \pm 1.98 \cdot 0.026.$$

Es gilt

$$1.98 \cdot 0.026 = 0.05148.$$

Damit lautet das Intervall

$$[-0.083 - 0.05148, -0.083 + 0.05148].$$

Also gilt

$$KI_{95\%} \approx [-0.1345, -0.0315].$$

Interpretation: Mit 95-prozentigem Konfidenzniveau liegt der wahre Effekt von skipped auf colGPA zwischen ungefähr  $-0.135$  und  $-0.032$ .

Da das Intervall die Null nicht enthält, ist der Koeffizient auf dem 5-Prozent-Niveau signifikant.

**Bewertung: 3 Punkte**

- 1 Punkt für die Intervallformel.
- 1 Punkt für die numerische Berechnung.
- 1 Punkt für die Interpretation.

## e) Globaler F-Test

Die Nullhypothese des globalen F-Tests lautet

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_K = 0.$$

Die Alternativhypothese lautet, dass mindestens ein Steigungskoeffizient ungleich null ist.

Die Teststatistik in Abhängigkeit von  $R^2$  lautet

$$F = \frac{R^2/K}{(1 - R^2)/(n - K - 1)}.$$

Unter der Nullhypothese gilt

$$F \sim F_{K, n-K-1}.$$

**Bewertung: 3 Punkte**

- 1 Punkt für die Nullhypothese.
- 1 Punkt für die Interpretation der Alternativhypothese.
- 1 Punkt für die Teststatistik.

## f) Globaler F-Test bei $R^2 = 0,234$

Für das Modell aus Teil c) gilt

$$n = 141, \quad K = 3, \quad R^2 = 0.234.$$

Die Teststatistik lautet

$$F = \frac{0.234/3}{(1 - 0.234)/(141 - 3 - 1)}.$$

Der Zähler ist

$$\frac{0.234}{3} = 0.078.$$

Der Nenner ist

$$\frac{1 - 0.234}{137} = \frac{0.766}{137} \approx 0.00559.$$

Damit ergibt sich

$$F \approx \frac{0.078}{0.00559} \approx 13.95.$$

Der kritische Wert der F-Verteilung bei 3 Restriktionen und ungefähr 137 Freiheitsgraden im Nenner liegt auf dem 5-Prozent-Niveau ungefähr bei 2.68.

Da

$$13.95 > 2.68$$

gilt, wird die Nullhypothese verworfen.

Die Regressoren sind gemeinsam statistisch signifikant.

**Bewertung: 3 Punkte**

- 1 Punkt für das korrekte Einsetzen.
- 1 Punkt für  $F \approx 13.95$ .
- 1 Punkt für die korrekte Testentscheidung.

# Aufgabe 3: Heteroskedastizität

Betrachtet wird das Modell

$$y = X\beta + u$$

unter den Annahmen MLR 1 bis MLR 4.

## a) Reine Heteroskedastizität bei unkorrelierten Fehlertermen

Reine Heteroskedastizität bei unkorrelierten Fehlertermen bedeutet, dass die Fehlervarianzen über die Beobachtungen variieren, während unterschiedliche Fehlerterme unkorreliert sind.

Formal gilt

$$\text{Var}(u_i|X) = \sigma_i^2 > 0$$

und

$$\text{Cov}(u_i, u_j|X) = 0 \quad \text{für } i \neq j.$$

Die Varianz-Kovarianz-Matrix lautet

$$\Sigma(u|X) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}.$$

In der Form  $\Sigma(u|X) = \sigma^2 \Omega$  gilt

$$\Omega = \begin{pmatrix} \frac{\sigma_1^2}{\sigma^2} & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2^2}{\sigma^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_n^2}{\sigma^2} \end{pmatrix}.$$

**Bewertung: 3 Punkte**

- 1 Punkt für  $\text{Var}(u_i|X) = \sigma_i^2$ .
- 1 Punkt für  $\text{Cov}(u_i, u_j|X) = 0$ .
- 1 Punkt für die Matrixdarstellung als  $\sigma^2 \Omega$ .

## b) GLS-Transformation

Im Fall reiner Heteroskedastizität ist die Transformationsmatrix  $P = \Omega^{-1/2}$  diagonal. Sie lautet

$$P = \begin{pmatrix} \frac{\sigma}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma}{\sigma_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma}{\sigma_n} \end{pmatrix}.$$

Die transformierten Größen lauten

$$\tilde{y} = Py, \quad \tilde{X} = PX, \quad \tilde{u} = Pu.$$

Für einzelne Beobachtungen gilt

$$\tilde{y}_i = \frac{\sigma}{\sigma_i} y_i,$$

$$\tilde{x}_{ij} = \frac{\sigma}{\sigma_i} x_{ij},$$

und

$$\tilde{u}_i = \frac{\sigma}{\sigma_i} u_i.$$

Das transformierte Modell ist

$$\tilde{y} = \tilde{X}\beta + \tilde{u}.$$

Für die transformierte Fehlervarianz gilt

$$Var(\tilde{u}_i|X) = Var\left(\frac{\sigma}{\sigma_i} u_i \middle| X\right) = \frac{\sigma^2}{\sigma_i^2} Var(u_i|X) = \sigma^2.$$

Außerdem bleibt für  $i \neq j$

$$Cov(\tilde{u}_i, \tilde{u}_j|X) = 0.$$

Damit erfüllt das transformierte Modell MLR 5.

**Bewertung: 3 Punkte**

- 1 Punkt für die korrekte Matrix  $P$ .
- 1 Punkt für die Transformation von  $y_i$ ,  $x_{ij}$  und  $u_i$ .
- 1 Punkt für die Begründung der Homoskedastizität nach Transformation.

### c) Nullhypothese des Breusch-Pagan-Tests

Die Nullhypothese des Breusch-Pagan-Tests lautet Homoskedastizität.

Formal gilt

$$H_0 : \Sigma(u|X) = \sigma^2 I.$$

Äquivalent gilt

$$H_0 : Var(u_i|X) = \sigma^2 \quad \text{für alle } i.$$

Unter MLR 4 gilt

$$E[u_i|X] = 0.$$

Daher folgt

$$Var(u_i|X) = E[u_i^2|X] - E[u_i|X]^2 = E[u_i^2|X].$$

Unter Homoskedastizität gilt somit

$$E[u_i^2|X] = \sigma^2.$$

**Bewertung: 3 Punkte**

- 1 Punkt für die Nullhypothese.
- 1 Punkt für die Varianzzerlegung.
- 1 Punkt für die Schlussfolgerung  $E[u_i^2|X] = \sigma^2$ .

## d) Hilfsregression des Breusch-Pagan-Tests

Zunächst wird das ursprüngliche Modell per OLS geschätzt und es werden die Residuen  $\hat{u}_i$  gespeichert.

Anschließend werden die quadrierten Residuen auf die Regressoren regressiert.

Die Hilfsregression lautet

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \dots + \delta_K x_{iK} + \varepsilon_i.$$

Die Nullhypothese lautet

$$H_0 : \delta_1 = \delta_2 = \dots = \delta_K = 0.$$

Die Alternativhypothese lautet, dass mindestens ein  $\delta_j$  ungleich null ist.

Der Test wird als globaler F-Test in der Hilfsregression durchgeführt.

**Bewertung: 3 Punkte**

- 1 Punkt für die Hilfsregression.
- 1 Punkt für die Restriktionen.
- 1 Punkt für den globalen F-Test.

## e) Hilfsregression des White-Tests bei drei Regressoren

Bei drei Regressoren werden die quadrierten Residuen auf die Regressoren, deren Quadrate und deren Kreuzprodukte regressiert.

Die Hilfsregression lautet

$$\hat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \delta_2 x_{i2} + \delta_3 x_{i3} + \delta_4 x_{i1}^2 + \delta_5 x_{i2}^2 + \delta_6 x_{i3}^2 + \delta_7 x_{i1}x_{i2} + \delta_8 x_{i1}x_{i3} + \delta_9 x_{i2}x_{i3} + \varepsilon_i.$$

Die Nullhypothese lautet

$$H_0 : \delta_1 = \delta_2 = \dots = \delta_9 = 0.$$

**Bewertung: 3 Punkte**

- 1 Punkt für die Regressoren in Leveln.
- 1 Punkt für die Quadrate.
- 1 Punkt für die Kreuzprodukte und die Restriktionen.

## f) Vorteile und Nachteile des White-Tests

Der Vorteil des White-Tests besteht darin, dass er allgemeinere Formen von Heteroskedastizität erkennen kann als der Breusch-Pagan-Test.

Er berücksichtigt neben den Regressoren auch deren Quadrate und Kreuzprodukte.

Dadurch kann er nichtlineare Zusammenhänge und Interaktionen in der Varianzstruktur erfassen.

Der Nachteil besteht darin, dass viele zusätzliche Regressoren in die Hilfsregression aufgenommen werden.

Bei  $K$  Regressoren entstehen

$$\frac{K(K+3)}{2}$$

Restriktionen.

Dadurch gehen Freiheitsgrade verloren, was insbesondere in kleinen Stichproben problematisch sein kann.

**Bewertung: 3 Punkte**

- 1 Punkt für den Vorteil der größeren Flexibilität.
- 1 Punkt für den Nachteil vieler zusätzlicher Regressoren.
- 1 Punkt für Freiheitsgradverlust oder kleine-Stichproben-Problematik.

## Aufgabe 4: Serielle Korrelation in Zeitreihen

Betrachtet wird das Modell

$$y = X\beta + u$$

unter den Annahmen MLR 1 bis MLR 4.

### a) Unverzerrtheit von OLS

Der OLS-Schätzer lautet

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

Mit  $y = X\beta + u$  folgt

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u.$$

Bedingt auf  $X$  gilt

$$E[\hat{\beta}|X] = \beta + (X'X)^{-1}X'E[u|X].$$

Unter MLR 4 gilt

$$E[u|X] = 0.$$

Daher folgt

$$E[\hat{\beta}|X] = \beta.$$

Serielle Korrelation verletzt nicht automatisch MLR 4.

Solange MLR 4 erfüllt ist, bleibt OLS unverzerrt.

**Bewertung: 3 Punkte**

- 1 Punkt für die Darstellung von  $\hat{\beta}$ .
- 1 Punkt für die Erwartungswertbildung.
- 1 Punkt für die Rolle von MLR 4.

### b) Ungültigkeit üblicher Standardfehler und Tests

Die üblichen OLS-Standardfehler beruhen auf

$$\Sigma(u|X) = \sigma^2 I.$$

Bei serieller Korrelation gilt jedoch

$$\text{Cov}(u_t, u_s|X) \neq 0 \quad \text{für mindestens einige } t \neq s.$$

Die Varianz-Kovarianz-Matrix der Fehler ist also nicht mehr diagonal.

Die tatsächliche Varianz des OLS-Schätzers lautet

$$\text{Var}(\hat{\beta}|X) = (X'X)^{-1}X'\Sigma(u|X)X(X'X)^{-1}.$$

Diese Matrix ist im Allgemeinen nicht gleich

$$\sigma^2(X'X)^{-1}.$$

Daher sind die üblichen OLS-Standardfehler falsch.

Folglich sind darauf basierende t-Tests, F-Tests und Konfidenzintervalle nicht valide.

**Bewertung: 3 Punkte**

- 1 Punkt für die Verletzung von  $\Sigma(u|X) = \sigma^2 I$ .
- 1 Punkt für die korrekte Varianzformel.
- 1 Punkt für die Konsequenz für Tests.

### c) Ineffizienz von OLS

OLS berücksichtigt die serielle Korrelationsstruktur der Fehler nicht.

Wenn die Varianz-Kovarianz-Struktur der Fehler bekannt wäre, könnte ein GLS-Schätzer konstruiert werden.

Dieser GLS-Schätzer hätte im Allgemeinen eine kleinere Varianz als OLS.

Daher ist OLS bei serieller Korrelation im Allgemeinen nicht effizient.

**Bewertung: 3 Punkte**

- 1 Punkt für die Aussage, dass OLS die Korrelationsstruktur ignoriert.
- 1 Punkt für den Hinweis auf GLS oder andere effizientere Schätzer.
- 1 Punkt für die Schlussfolgerung, dass OLS nicht effizient ist.

### d) Interpretation von $\hat{\rho} = 0,5725$

Die Hilfsregression lautet

$$\hat{u}_t = -0.1118 + 0.5725\hat{u}_{t-1}.$$

Der Koeffizient 0.5725 ist ein Schätzer für die Autokorrelation erster Ordnung.

Der positive Wert zeigt positive serielle Korrelation an.

Ein positiver Fehler in Periode  $t - 1$  geht tendenziell mit einem positiven Fehler in Periode  $t$  einher.

Der p-Wert beträgt

$$2.43 \cdot 10^{-6}.$$

Dieser p-Wert ist sehr klein.

Daher wird die Nullhypothese keiner Autokorrelation deutlich verworfen.

Es liegt ein starker Hinweis auf positive Autokorrelation vor.

**Bewertung: 3 Punkte**

- 1 Punkt für die Interpretation als Autokorrelationsschätzer.
- 1 Punkt für die positive Autokorrelation.
- 1 Punkt für die Schlussfolgerung aus dem p-Wert.

### e) Probleme des t-Tests in der Hilfsregression

Der t-Test in der Hilfsregression verwendet OLS-Standardfehler.

Diese Standardfehler sind problematisch, weil die verwendeten  $\hat{u}_t$  geschätzte Residuen und nicht die wahren Fehler sind.

Außerdem liegt bei serieller Korrelation gerade eine Verletzung der klassischen Fehlerannahmen vor.

Der gewöhnliche t-Test ist daher in endlichen Stichproben nicht exakt valide.

Er ist lediglich asymptotisch, also für große Stichproben, näherungsweise valide.

**Bewertung: 3 Punkte**

- 1 Punkt für den Hinweis auf geschätzte Residuen.
- 1 Punkt für problematische Standardfehler.
- 1 Punkt für die Aussage, dass der Test nur asymptotisch valide ist.

## f) Alternative Verfahren zur Diagnose serieller Korrelation

Ein erstes Verfahren ist der Durbin-Watson-Test.

Die Durbin-Watson-Statistik lautet

$$DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2}.$$

Es gilt näherungsweise

$$DW \approx 2(1 - \hat{\rho}).$$

Bei positiver Autokorrelation liegt  $DW$  typischerweise unter 2.

Mithilfe kritischer Werte kann entschieden werden, ob die Nullhypothese keiner positiven Autokorrelation verworfen wird.

Ein zweites Verfahren ist der Breusch-Godfrey-Test.

Dabei wird zunächst das Ausgangsmodell per OLS geschätzt und die Residuen  $\hat{u}_t$  werden gespeichert.

Anschließend wird  $\hat{u}_t$  auf  $\hat{u}_{t-1}$  und die ursprünglichen Regressoren  $X$  regressiert.

Aus dieser Hilfsregression wird das Bestimmtheitsmaß  $R^2$  verwendet.

Die Teststatistik lautet bei Autokorrelation erster Ordnung näherungsweise

$$BG = (n - 1)R^2.$$

Asymptotisch gilt

$$BG \sim \chi_1^2.$$

Der Breusch-Godfrey-Test ist flexibler als der einfache residualbasierte t-Test.

**Bewertung: 3 Punkte**

- 1 Punkt für ein korrektes Diagnoseverfahren.
- 1 Punkt für ein zweites korrektes Diagnoseverfahren.
- 1 Punkt für die kurze Beschreibung des Vorgehens.

## Aufgabe 5: Instrumente

Betrachtet wird die wahre Lohnregression

$$\ln(w_i) = \gamma_0 + \gamma_1 \text{educ}_i + \gamma_2 a_i + e_i,$$

wobei  $a_i$  individuelle Fähigkeiten beschreibt.

Da  $a_i$  unbeobachtbar ist, wird stattdessen das reduzierte Modell

$$\ln(w_i) = \beta_0 + \beta_1 \text{educ}_i + u_i$$

geschätzt.

### a) Fehlerterm und Kovarianz

Da  $a_i$  ausgelassen wird, erscheint der Term  $\gamma_2 a_i$  im Fehlerterm des reduzierten Modells.

Der Fehlerterm lautet daher

$$u_i = \gamma_2 a_i + e_i.$$

Damit gilt

$$\text{Cov}(\text{educ}_i, u_i) = \text{Cov}(\text{educ}_i, \gamma_2 a_i + e_i).$$

Wegen der Linearität der Kovarianz folgt

$$\text{Cov}(\text{educ}_i, u_i) = \gamma_2 \text{Cov}(\text{educ}_i, a_i) + \text{Cov}(\text{educ}_i, e_i).$$

Wenn  $\text{Cov}(\text{educ}_i, e_i) = 0$  gilt, vereinfacht sich dies zu

$$\text{Cov}(\text{educ}_i, u_i) = \gamma_2 \text{Cov}(\text{educ}_i, a_i).$$

**Bewertung: 3 Punkte**

- 1 Punkt für  $u_i = \gamma_2 a_i + e_i$ .
- 1 Punkt für die Kovarianzzerlegung.
- 1 Punkt für die vereinfachte Darstellung oder Interpretation.

### b) Bedingungen für Endogenität von $\text{educ}_i$

Der Regressor  $\text{educ}_i$  ist endogen, wenn

$$\text{Cov}(\text{educ}_i, u_i) \neq 0$$

gilt.

Aus Teil a) folgt

$$\text{Cov}(\text{educ}_i, u_i) = \gamma_2 \text{Cov}(\text{educ}_i, a_i) + \text{Cov}(\text{educ}_i, e_i).$$

Wenn  $\text{Cov}(\text{educ}_i, e_i) = 0$  angenommen wird, entsteht Endogenität genau dann, wenn

$$\gamma_2 \neq 0$$

und

$$\text{Cov}(\text{educ}_i, a_i) \neq 0$$

gelten.

Die Bedingung  $\gamma_2 \neq 0$  bedeutet, dass Fähigkeiten einen eigenständigen Einfluss auf den Lohn haben.

Die Bedingung  $Cov(educ_i, a_i) \neq 0$  bedeutet, dass Fähigkeiten mit der Bildungsdauer zusammenhängen. Im Lohnbeispiel ist es plausibel, dass fähigere Personen sowohl mehr Bildung erwerben als auch höhere Löhne erzielen.

Daher erscheint Endogenität von  $educ_i$  plausibel.

**Bewertung: 3 Punkte**

- 1 Punkt für  $Cov(educ_i, u_i) \neq 0$ .
- 1 Punkt für  $\gamma_2 \neq 0$  und  $Cov(educ_i, a_i) \neq 0$ .
- 1 Punkt für eine plausible ökonomische Diskussion.

### c) Bedingungen für eine gültige Instrumentalvariable

Eine gültige Instrumentalvariable  $z_i$  für  $educ_i$  muss zwei Bedingungen erfüllen.

Die erste Bedingung ist Instrumentrelevanz:

$$Cov(z_i, educ_i) \neq 0.$$

Das Instrument muss also mit dem endogenen Regressor korreliert sein.

Die zweite Bedingung ist Instrumentexogenität:

$$Cov(z_i, u_i) = 0.$$

Das Instrument darf also nicht mit dem Fehlerterm der strukturellen Gleichung korreliert sein.

Die Instrumentrelevanz kann statistisch geprüft werden, zum Beispiel durch eine Regression von  $educ_i$  auf  $z_i$ .

Die Instrumentexogenität kann im exakt identifizierten Fall nicht statistisch getestet werden.

Sie muss theoretisch beziehungsweise ökonomisch begründet werden.

**Bewertung: 3 Punkte**

- 1 Punkt für Instrumentrelevanz.
- 1 Punkt für Instrumentexogenität.
- 1 Punkt für Testbarkeit beziehungsweise theoretische Begründung.

### d) Theoretischer IV-Ausdruck für $\beta_1$

Das reduzierte Modell lautet

$$\ln(w_i) = \beta_0 + \beta_1 educ_i + u_i.$$

Bildet man die Kovarianz mit  $z_i$ , ergibt sich

$$Cov(\ln(w_i), z_i) = Cov(\beta_0 + \beta_1 educ_i + u_i, z_i).$$

Da die Kovarianz einer Konstanten mit  $z_i$  null ist, folgt

$$Cov(\ln(w_i), z_i) = \beta_1 Cov(educ_i, z_i) + Cov(u_i, z_i).$$

Wegen Instrumentexogenität gilt

$$Cov(u_i, z_i) = 0.$$

Damit bleibt

$$Cov(\ln(w_i), z_i) = \beta_1 Cov(educ_i, z_i).$$

Wegen Instrumentrelevanz gilt

$$\text{Cov}(\text{educ}_i, z_i) \neq 0.$$

Daher erhält man

$$\beta_1 = \frac{\text{Cov}(\ln(w_i), z_i)}{\text{Cov}(\text{educ}_i, z_i)}.$$

**Bewertung: 3 Punkte**

- 1 Punkt für die Kovarianzgleichung.
- 1 Punkt für die Verwendung von  $\text{Cov}(u_i, z_i) = 0$ .
- 1 Punkt für den finalen IV-Ausdruck.

## e) Einfacher IV-Schätzer und OLS als Spezialfall

Der einfache IV-Schätzer ergibt sich, indem die theoretischen Kovarianzen durch Stichprobenkovarianzen ersetzt werden.

Es gilt

$$\hat{\beta}_1^{IV} = \frac{\widehat{\text{Cov}}(z, \ln(w))}{\widehat{\text{Cov}}(z, \text{educ})}.$$

In Summenschreibweise ist

$$\hat{\beta}_1^{IV} = \frac{\frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})(\ln(w_i) - \overline{\ln(w)})}{\frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})(\text{educ}_i - \overline{\text{educ}})}.$$

Da sich die Faktoren  $\frac{1}{n-1}$  kürzen, kann auch geschrieben werden

$$\hat{\beta}_1^{IV} = \frac{\sum_{i=1}^n (z_i - \bar{z})(\ln(w_i) - \overline{\ln(w)})}{\sum_{i=1}^n (z_i - \bar{z})(\text{educ}_i - \overline{\text{educ}})}.$$

OLS ist ein Spezialfall von IV, wenn der Regressor exogen ist und sich selbst instrumentieren kann. Wenn also  $z_i = \text{educ}_i$  gesetzt wird, ergibt sich

$$\hat{\beta}_1^{IV} = \frac{\widehat{\text{Cov}}(\text{educ}, \ln(w))}{\widehat{\text{Var}}(\text{educ})},$$

was dem OLS-Schätzer im einfachen Regressionsmodell entspricht.

**Bewertung: 3 Punkte**

- 1 Punkt für den IV-Schätzer als Kovarianzquotient.
- 1 Punkt für die Summenschreibweise.
- 1 Punkt für OLS als Spezialfall mit  $z_i = x_i$ .

## f) Weak-Instruments-Problem

Das Weak-Instruments-Problem liegt vor, wenn das Instrument nur schwach mit dem endogenen Regressor korreliert ist.

Formal ist dann

$$\text{Cov}(z_i, \text{educ}_i)$$

sehr klein.

Der IV-Schätzer kann geschrieben werden als

$$\hat{\beta}_1^{IV} = \beta_1 + \frac{\widehat{Cov}(z, u)}{\widehat{Cov}(z, educ)}.$$

Wenn der Nenner sehr klein ist, können schon kleine Abweichungen im Zähler zu großen Verzerrungen führen. Schwache Instrumente führen zudem zu großen Standardfehlern und unpräzisen Schätzungen.

In kleinen Stichproben kann IV mit schwachen Instrumenten schlechtere Eigenschaften haben als OLS.

Besonders problematisch sind schwache Instrumente, wenn sie zusätzlich nicht vollständig exogen sind.

Praktisch wird die Stärke der Instrumente häufig über die first-stage Regression und den zugehörigen F-Test beurteilt.

Als Faustregel gilt häufig, dass ein first-stage  $F$ -Wert größer als 10 gegen ein gravierendes Weak-Instruments-Problem spricht.

#### **Bewertung: 3 Punkte**

- 1 Punkt für die Definition schwacher Instrumente.
- 1 Punkt für Folgen wie Verzerrung, große Standardfehler oder geringe Präzision.
- 1 Punkt für den Hinweis auf first-stage F-Test oder die Faustregel  $F > 10$ .

# Punkteübersicht

| Aufgabe       | Teilaufgaben     | Punkte            |
|---------------|------------------|-------------------|
| Aufgabe 1     | a)–f)            | $6 \cdot 3 = 18$  |
| Aufgabe 2     | a)–f)            | $6 \cdot 3 = 18$  |
| Aufgabe 3     | a)–f)            | $6 \cdot 3 = 18$  |
| Aufgabe 4     | a)–f)            | $6 \cdot 3 = 18$  |
| Aufgabe 5     | a)–f)            | $6 \cdot 3 = 18$  |
| Zu bearbeiten | 4 von 5 Aufgaben | $4 \cdot 18 = 72$ |

Die Gesamtpunktzahl der Klausur beträgt maximal 72 Punkte, da vier von fünf Aufgaben bewertet werden.