Aufgabe 1:

Der folgende EViews-Output liefert Ihnen Informationen über den Zusammenhang zwischen der Anzahl der Verbrechen in einem Gebiet X auf der einen Seite und der Anzahl der Gefangenen, der Anzahl der Polizisten, der Arbeitslosenrate und dem nominalem Einkommen desselben Gebietes X auf der anderen Seite. Dabei wurden folgende Variablen verwendet:

crime:	**Anzahl Verbrechen pro 100'000 Einwohner**
log(crime):	**Natürlicher Logarithmus von crime**
pris:	**Anzahl Gefangene pro 100'000 Einwohner**
log(pris):	**Natürlicher Logarithmus von pris**
police:	**Anzahl Polizisten pro 100'000 Einwohner**
unem:	**Arbeitslosenrate in Prozent**
income:	**Pro Kopf Einkommen, nominal**
log(income):	**Natürlicher Logarithmus von income**

Der aus den USA stammende Datensatz beinhaltet 714 Beobachtungen aus 51 Bundesstaaten.

Dependent Variable: LOG(CRIME)
Method: Least Squares
Date: 14/12/04 Time: 08.59
Sample: 1 714
Included observations: 714

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C (b_1)	3.817823	0.357300	10.68519	0.0000
LOG(PRIS) (b_2)	0.102309	0.018234	5.610776	0.0000
POLICE (b_3)	0.001335	0.000116	11.53013	0.0000
UNEM (b_4)	-0.483089		-1.128397	0.2595
LOG(INCOME) (b_5)	-0.090287	0.039386	-2.292377	0.0222

R-squared		Mean dependent var	3.803649
Adjusted R-squared		S.D. dependent var	0.265534
S.E. of regression		Akaike info criterion	-0.153690
Sum squared resid	35.35041	Schwarz criterion	-0.121681
Log likelihood	59.86745	F-statistic	74.82040
Durbin-Watson stat	0.264170	Prob(F-statistic)	0.000000

a) Bestimmen Sie den Standardfehler für b_4 und den Standardfehler der Regression.

– $se(b_4) = \frac{b_4 - 0}{t_4} = \frac{-0.483089}{-1.128397} = \mathbf{0.428120}$

– S.E. of regression: $\hat{\sigma} = \sqrt{\frac{\sum \hat{e}^2}{T-K}} = \sqrt{\frac{35.35041}{714-5}} = \mathbf{0.223292}$

b) Interpretieren Sie die Koeffizienten b_1, b_2 und b_3, und beschreiben Sie die geschätzten Zusammenhänge zwischen den erklärenden Variablen und der abhängigen Variablen. Erläutern Sie, ob Ihnen die Ergebnisse plausibel erscheinen.

– b_1 (Konstante C): Der Parameter β_1 ist der Achsenabschnittsterm, der Effekt der "Variable" x_{t1}=1. Sind alle anderen Parameter Null, so ist ln(y) = 3.81, d.h. es gibt 45,2 Verbrechen pro 100,000 Einwohner

($\ln(y) = 3.81 \;\Rightarrow\; y = e^{3.81} \cong 45.2$). Generell kritisch zu interpretieren;
→ hoch signifikant (p-Wert von 0).

– b_2 (log(PRIS)): Steigt die Anzahl der Gefangenen um 1%, so steigt die Anzahl der Verbrechen um 0,1%. Der Zusammenhang dürfte aber umgekehrt sein, es ist plausibler, dass die Anzahl der Verbrechen die Anzahl der Gefangenen determiniert als umgekehrt. → OLS gibt keine Auskunft über Kausalität;
→ hoch signifikant (p-Wert von 0).

– b_3 (POLICE): Steigt die Anzahl der Polizisten um eins, so steigt die Anzahl der Verbrechen um 0,001%. → Semielastizität;
mögliche Interpretation: Der Nutzen der Polizei durch Verbrechensaufdeckung scheint größer zu sein als die Abschreckung, die durch mehr Polizisten verursacht wird.
→ hoch signifikant (p-Wert von 0).

c) Bestimmen Sie das 99% Konfidenzintervall für b_3 (geben Sie die für die Berechnung korrekten Freiheitsgrade an, unterstellen Sie jedoch für Ihre Berechnung $DF = \infty$). Interpretieren Sie Ihr Ergebnis.

– $t_{(\alpha/2=0.005;709)} \approx 2.576 \qquad KI:\ b_i \pm t_i * se(b_i)$

– 99%-KI um b_3: $b_3 \pm t_c \cdot \sigma_{b_3} = 0.001335 \pm 2.576 \cdot 0{,}000116 = \left[0.001036\ ;\ 0.001634\right]$

d) Nehmen Sie an, dass der im Anschluss einer linearen Regression durchgeführten Jarque-Bera Test den Wert 19.459447 ergibt und der entsprechende p-Wert 0.003 ist. Was überprüft dieser Test typischerweise? Zu welchem Schluss kommen Sie aufgrund der genannten Ergebnisse? Erläutern Sie, ob und gegebenenfalls welche Konsequenzen Ihr Ergebnis für Ihre Lösung der Teilaufgabe c) hat.

– Der Jarque-Bera Test überprüft die Residuen auf Normalverteilung. H_0 ist dabei die Hypothese, dass die Residuen einer Normalverteilung entsprechen. Der p-Wert zeigt die Wahrscheinlichkeit an, dass diese Daten unter der Normalverteilung auftreten. Mit einem p-Wert von 0.003 können wir auch auf 1% Signifikanzniveau die Normalverteilungsannahme verwerfen. Die in Teilaufgabe c) berechneten Werte haben dementsprechend keine Gültigkeit.

e) Bestimmen Sie das R^2, wenn Sie wissen, dass $\sum y_t^2 = 10380.2427$. Was sagt es aus?

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

mit: $SST = \sum\left(y_t - \bar{y}\right)^2$; $\quad SSR = \sum\left(\hat{y}_t - \bar{y}\right)^2$; $\quad SSE = \sum \hat{e}_t^2$

$$R^2 = 1 - \frac{\sum \hat{e}_t^2}{\sum\left(y_t - \bar{y}\right)^2} = 1 - \frac{\sum \hat{e}_t^2}{\sum y_t^2 - 2\bar{y}\sum y_t + T\bar{y}^2} = 1 - \frac{\sum \hat{e}_t^2}{\sum y_t^2 - T\bar{y}^2}$$

–

$$= 1 - \frac{35.35041}{10380.2427 - 714 \cdot 3.803649^2} = \mathbf{0.296823}$$

– Das R^2 ist ein Maß, welches den Anteil der durch das Modell erklärten Variation $\left(b_1 + b_2 x_t\right)$ an der Gesamtvariation y_t beschreibt. Die Werte für das R^2 liegen zwischen 0 und 1.

Aufgabe 2:

Ihnen liegt ein Datensatz mit Informationen über Pensionäre vor. Sie regressieren die jährlich ausgezahlten Pensionen (P) einer Pensionskasse (gemessen in CHF) auf die Anzahl der Jahre mit Beitragszahlung (BJ) der Versicherten, sowie auf das letzte Jahreseinkommen (JE) (gemessen in CHF), das die Pensionäre verdient haben:

$$P_t = a_0 + a_1 BJ_t + a_2 JE_t + e_t$$

a) Wie interpretieren Sie den Koeffizienten a_1 wenn er beispielsweise den Wert 2'000.8 annimmt?

− $BJ + 1 \quad \Rightarrow \quad P + 2000.8\, CHF \;\; p.a.$

b) Ihr Kollege behauptet, dass nicht die Anzahl der Jahre mit Beitragszahlung die relevante Grösse darstellt, sondern vielmehr die Quartale, in denen Beiträge bezahlt wurden. Wie ändert sich der Koeffizient a_1 wenn Sie diese Anpassung in der Variable BJ vornehmen?

− a_1 wird sich auf ¼ seines Wertes reduzieren (500.2).

c) Sie verwerfen den Vorschlag des Kollegen und legen Ihre ursprünglichen Schätzergebnisse Ihrem Vorgesetzten vor, der daraufhin vorschlägt, die abhängige Variable statt mit ihrem jährlichen mit ihrem monatlichen Wert in der Schätzung zu benutzen. Welche Wirkung hat dies auf die drei geschätzten Koeffizienten?

− alle Koeffizienten *1/12.

Aufgabe 3:

Sie schätzen ein einfaches lineares Regressionsmodell und betrachten anschließend die Residuen. Dabei stellen Sie fest, dass die Residuen bei den sehr niedrigen und den sehr hohen Werten der erklärenden Variable negative Werte aufweisen, während sie im mittleren Bereich meist positiv sind.

a) Wie interpretieren Sie diesen Befund?

− Da im Residuum ein systematischer Zusammenhang zwischen y (bzw. e) und x verbleibt, ist die funktionale Form von X nicht optimal abgebildet.

b) Entspricht dieser Befund den Grundannahmen des einfachen linearen Regressionsmodells bezüglich der Residuen?

− Nicht solange $E(e) = 0$, $var(e) = \sigma^2$, $cov(e_i, e_j) = 0$, lineare funktionale Form, und *x* verschiedene Werte annimmt. Problematisch ist hier die Annahme $cov(e_i, e_j) = 0$.

c) Welche Möglichkeit sehen Sie, Ihr Regressionsmodell zu verbessern und wie könnte der Erfolg dieser Maßnahme gemessen werden?

− andere funktionale Form für *x* (ln(*x*); x^2)
− verbessertes R^2/ angepasstes R^2
− zufällig verteilte Residuen

Aufgabe 4:

Der Kleinstquadratschätzer des Modells $y_t = \beta_1 + \beta_2 x_t + e_t$ minimiert eine Funktion der Störterme et, die im X-Y Diagramm als *vertikale* Abweichungen von der Regressionsgeraden dargestellt werden. Wie würden Sie die Parameterschätzer für β_1 und β_2 ableiten, wenn Sie stattdessen die *horizontalen* Abweichungen von der Regressionsgerade nutzen sollten? Geben Sie die zu minimierende Zielfunktion S an.

− $S = \sum e_t^2$
$y_t = \beta_1 + \beta_2 x_t + e_t$

− auf Gerade: $y_t = \beta_1 + \beta_2 x_t$

$$\Leftrightarrow \qquad x_t = -\frac{\beta_1}{\beta_2} + \frac{1}{\beta_2} y_t$$

− Zielfunktion: $S = \sum \left(x_t - \left(-\frac{\beta_1}{\beta_2} + \frac{1}{\beta_2} y_t \right) \right)^2$

bzw. allgemein: Wenn $x_t = a + by_t$

$$S = \sum \left(x - \left(a + by \right) \right)^2$$

Aufgabe 5:

Wahr oder Falsch? Tragen Sie für zutreffende Aussagen den Buchstaben w (für wahr), für nicht zutreffende f (für falsch) ein. (Für jede richtige Antwort gibt es 0,5 Punkte, für jede falsche Antwort werden 0,5 Punkte abgezogen. Die Gesamtpunktzahl kann nicht negativ werden.)

F	Das einfache lineare Regressionsmodell wird unter der Annahme geschätzt, dass die erklärenden Variablen unkorreliert und normalverteilt sind.
F	Bei einem einseitigen t-Test verwerfen wir die Nullhypothese, wenn der p-Wert kleiner als α ist.
W	Das angepasste R^2 kann nicht als Anteil der erklärten Varianz interpretiert werden.
W	Intervallschätzer von Parametern erlauben Rückschlüsse auf deren statistische Signifikanz.
W	Für die Ableitung des Kleinstquadratschätzers ist die Annahme der Normalverteilung unerheblich.
F	Wenn Schätzer erwartungstreu sind, stimmt der geschätzte Wert des Parameters mit dem wahren Wert überein.
F	Ein Typ II Fehler liegt vor, wenn eine Nullhypothese verworfen wurde, obwohl sie zutrifft.
F	Kollineare erklärende Variablen erhöhen die Präzision der Parameterschätzung im multiplen Regressionsmodell.
W	Bei einseitigen Tests von $H_0 : \beta \geq c$ liegt die Verwerfungsregion auf der linken Seite der t-Verteilung.
F	Eine hohe Streuung der abhängigen Variable erhöht die Präzision, mit der Parameter des linearen Modells geschätzt werden können.
W	Die Anzahl der Freiheitsgrade der Chi^2-Verteilung richtet sich nach der Anzahl der aufsummierten standardnormalverteilten und quadrierten Zufallsvariablen, auf deren Basis die Chi^2-verteilte Zufallsvariable bestimmt wurde.
W	R^2 lässt sich im einfachen Regressionsmodell aus dem Stichprobenkorrelationskoeffizienten von y und $\hat{y}$ berechnen.