Ökonometrie I, WS 2019/20, 1. Teilprüfung am 5.11.2019

Zeit: 105 min

Name:

Matrikelnummer:

Kursnummer:

1 Case Study 1 – Einfache Regression

Die folgenden Daten beschreiben den Konsum von Eistee in New Austin von 1951–03–18 bis 1953–07–11 in monatlichen Abständen. Y ist der Pro-Kopf-Konsum von Eistee (in Gallonen) und X ist die durchschnittliche Tageshöchsttemperatur (in Celsius).

1. Das einfache Regressionsmodell

$$Y = \beta_0 + \beta_1 X + u, \qquad u|X \sim \text{Normal}\left(0, \sigma^2\right), \tag{1}$$

wird via EViews/R geschätzt. Die Ergebnisse finden Sie in Abbildung 1 auf Seite 11 (in R geschätzt) und in Abbildung 7 auf Seite 15 (in EViews geschätzt).

   (a) Wie lautet der OLS-Schätzer von β_0? Wie kann diese Zahl interpretiert werden?

   (b) Wie lautet der OLS-Schätzung von β_1? Wie kann diese Zahl interpretiert werden?

(c) Was ist der erwartete Eisteekonsum bei einer durchschnittlichen Tageshöchsttemperatur von 20 Grad Celsius?

(d) Um wie viel ändert sich der erwartete Eisteekonsum, wenn die durchschnittliche Tageshöchsttemperatur um 5 Grad fällt?

(e) Was ist $\sum_{i=1}^{N} y_i - \hat{y}_i$ in diesem Beispiel?

(f) Betrachten Sie das Streudiagramm der Daten samt Regressionsgerade in Abbildung 2 (R) oder Abbildung 8 (EViews), und das Streudiagramm der Residuen in Abbildung 3 (R) oder Abbildung 9 (EViews). Welche der üblichen Annahmen für das lineare Regressionsmodell scheinen erfüllt (bzw. nicht erfüllt)?

2. Betrachten wir nun das Regressionsmodell, in dem Y (aber nicht X!) vor dem Schätzen logarithmiert wurde:

$$\log Y = \gamma_0 + \gamma_1 X + \epsilon, \qquad \epsilon | X \sim \text{Normal}\left(0, \sigma_\epsilon^2\right). \tag{2}$$

Die Resultate der OLS-Schätzung finden Sie in Abbildung 4 (R) oder Abbildung 10 (EViews).

   (a) Welche *relative* Änderung des Eisteekonsums ist zu erwarten, wenn die durchschnittliche Tageshöchsttemperatur um 1 Grad Celsius steigt?

   Hinweis: Bestimmen Sie den vorhergesagten Wert $\widehat{\log}(y)$ *für* $X = x$ *und für* $X = x + 1$ *und gehen Sie davon aus, dass näherungsweise* $\widehat{\log}(y) \approx \log(\hat{y})$ *gilt. Zudem dürfen Sie benutzen, dass näherungsweise* $\log(\hat{y}) \approx \hat{y} - 1$ *gilt.*

   (b) Ist das Modell (1) oder (2) besser zur Beschreibung der Daten geeignet? Begründen Sie anhand der beiden Streudiagramme in den Abbildungen 2 und 3 (R) oder Abbildungen 8 und 9 (EViews).

2 Case Study 2 – Multiple Regression

Aufgrund wachsender Bedenken über die Energieverschwendung durch Gebäude und deren Umweltfolgen ist der Energiebedarf von Gebäuden von aktueller Relevanz. Während des Sommers entfällt der Großteil des Energieverbrauchs in Gebäuden auf Klimaanlagen. Deswegen ist die Planung von Gebäuden mit vorteilhaftem Energieverhalten, wie zum Beispiel einer niedrigen Kühllast, so bedeutsam. Die Kühllast entspricht der Rate, mit der man Wärmeenergie aus einem Raum entfernen muss, um diesen in einem akzeptablen Temperaturbereich zu halten.

In dieser Fallstudie verwenden wir ein ein multiples lineares Regressionsmodell, um den Effekt der folgenden vier Variablen auf die Kühllast (`coolload`) in kW eines Gebäudes zu bestimmen: Wir betrachten den relativen Kompaktheitsindex (`relcomp`) in %, die Wandfläche (`wallarea`) in m^2, die Dachfläche (`roofarea`) in m^2 und den Verglasungsindex in %. Beachten Sie, dass der Verglasungsindex ein Prozentsatz zwischen 0 und 100 ist, der dem Flächenanteil von Fenstern an der Gesamtwandfläche entspricht. (Ein Wert von 0 % bedeutet zum Beispiel, dass es überhaupt keine Fenster gibt, wohingegen ein Wert von 100 % komplett verglasten Wänden entspricht.) Die Stichprobengröße beträgt $N = 768$.

Beantworten Sie folgende Fragen an Hand von Abbildung 13 (EViews output) oder 1 (R output) .

1. Was ist die OLS-Schätzung des zur Prädiktionsvariable Dachfläche gehörigen Regressionskoeffizienten? Wie kann man diese interpretieren?

2. Was ist die OLS-Schätzung des Achsenabschnitts? Was ist deren übliche Interpretation und ist diese bei diesem Beispiel sinnvoll? Erklären Sie!

3. Bestimmen Sie die erwartete Kühllast eines Gebäudes mit $220\,m^2$ Dachfläche, einem relativen Kompaktheitsindex von $70\,\%$, einer Wandfläche von $310\,m^2$ und einem Verglasungsindex von $40\,\%$.

4. Wie viel größer wäre die Kühllast des Gebäudes aus Aufgabe 3, wenn alle Wände komplett aus Glas bestünden?

5. Berechnen Sie die unverzerrte (unbiased) Schätzung der Fehlervarianz ($\hat{\sigma}^2$) unter Verwendung der berichteten Summe der quadrierten Residuen (SSR).

6. Was ist der Wert des Bestimmtheitsmaßes (coefficient of determination) und wie kann man diesen interpretieren?

7. Die Firmenleitung eines Bauunternehmens überlegt, warum die Gesamtoberfläche (`surfacearea`) nicht als zusätzlicher Prädiktor mit einbezogen wird. Bedenken Sie, dass die Gesamtoberfläche als

   ```
   surfacearea = wallarea + 2*roofarea
   ```

   definiert werden kann. Würden Sie dem Vorschlag der Firmenleitung folgen und die Gesamtoberfläche als zusätzlichen Prädiktor mit einbeziehen oder nicht? Erklären Sie Ihre Antwort.

8. Wir schätzen ein weiteres Modell mit der zusätzlichen Variable 'Wandhöhe' (`height`) in m. Das sich daraus ergebende Bestimmtheitsmaßist $R^2 = 0.8876$. Wie würden Sie sich entscheiden, welches Modell Sie wählen?

3 Richtig oder falsch?

Richtig oder falsch? Wo liegt bei einer falschen Aussage der Fehler? Die Aufgabe gilt als richtig gelöst, wenn Sie eine richtige Aussage als richtig erkennen bzw. wenn es Ihnen gelingt, eine falsche Aussage richtig zu stellen.

1. Betrachten Sie das einfache lineare Regressionsmodell

$$Y = \beta_0 + \beta_1 X + u,$$

wobei $\mathbb{E}(u|X) = 0$. Nehmen Sie weiters an, dass $\beta_0 = 2$, $\beta_1 = -3$ und $\mathbb{E}(X) = 1$.

(a) Falls X um 2 Einheiten fällt, erwarten wir, dass Y um 6 Einheiten steigt.

(b) Der Erwartungswert von Y ist 2.

2. Betrachten Sie das einfache lineare Regressionsmodell

$$Y = \beta_0 + \beta_1 X + u, \qquad \text{wobei } \mathbb{E}(u|X) = 0, \quad \mathbb{V}(u|X) = \sigma^2 > 0.$$

Es sei (y_i, x_i), $i = 1, \ldots, N$, eine zufällige Stichprobe der Grundgesamtheit. Seien $\hat{\beta}_0$ und $\hat{\beta}_1$ die OLS-Schätzer von β_0 und β_1.

(a) Je kleiner die Stichprobenvariation $s_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$ des Regressors ist, desto kleiner ist die Varianz des Schätzers $\hat{\beta}_1$.

(b) Betrachten Sie nun ein weiteres Modell

$$Y = \gamma_0 + \gamma_1 Z + u,$$

wobei $Z = -X$. Die OLS-Schätzer $\hat{\gamma}_0$ und $\hat{\gamma}_1$ für γ_0, γ_1 erfüllen die folgende Bedingung

$$\hat{\gamma}_0 = \hat{\beta}_0, \qquad \hat{\gamma}_1 = -\hat{\beta}_1.$$

3. Betrachten Sie das multiple lineare Regressionsmodell

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u,$$

wobei $\mathbb{E}(u|X_1, X_2, X_3) = 0$, $\mathbb{V}(u|X_1, X_2, X_3) = \sigma^2 > 0$. Sei $(y_i, x_{1,i}, x_{2,i}, x_{3,i})$, $i = 1, \ldots, N$, eine zufällige Stichprobe.

Sei $\hat{\boldsymbol{\beta}}$ der OLS-Schätzer für $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$. Wir definieren die Residuen $\hat{u}_i = y_i - \hat{y}_i$, $i = 1, \ldots, N$, wobei $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i} + \hat{\beta}_3 x_{3,i}$.

(a) Falls N gerade ist, so sind gleich viele Residuen positiv wie negativ.

(b) Es gilt, dass $\mathbb{E}(\hat{\beta}_j) = 0$ für alle $j = 0, 1, 2, 3$.

(c) Nehmen wir nun an, Y entspreche dem Einkommen einer Person, X_1 stehe für das Einkommen von deren Mutter, X_2 für das des Vaters und X_3 sei der Mittelwert der beiden elterlichen Einkommen. Sei $(y_i, x_{1,i}, x_{2,i}, x_{3,i})$, $i = 1, \ldots, N$, eine zufällige Stichprobe.

Dann kann der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ von $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ berechnet werden als

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \qquad \text{where} \qquad \boldsymbol{y} = (y_1, y_2, \ldots, y_N)'.$$

(d) Das Bestimmtheitsmaß R^2 nimmt immer ab, wenn man einen Regressor aus dem Modell herausnimmt.

A R output

A.1 Case Study 1

```
Call:
lm(formula = cons ~ temp_c, data = iceT)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.0393 -0.7996 -0.2193  0.2396  3.2941 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  0.24832    0.33122   0.750  0.45967   
temp_c       0.08779    0.02535   3.463  0.00174 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.245 on 28 degrees of freedom
Multiple R-squared:  0.2999,	Adjusted R-squared:  0.2749 
F-statistic: 11.99 on 1 and 28 DF,  p-value: 0.001735
```

Abbildung 1: Case Study 1

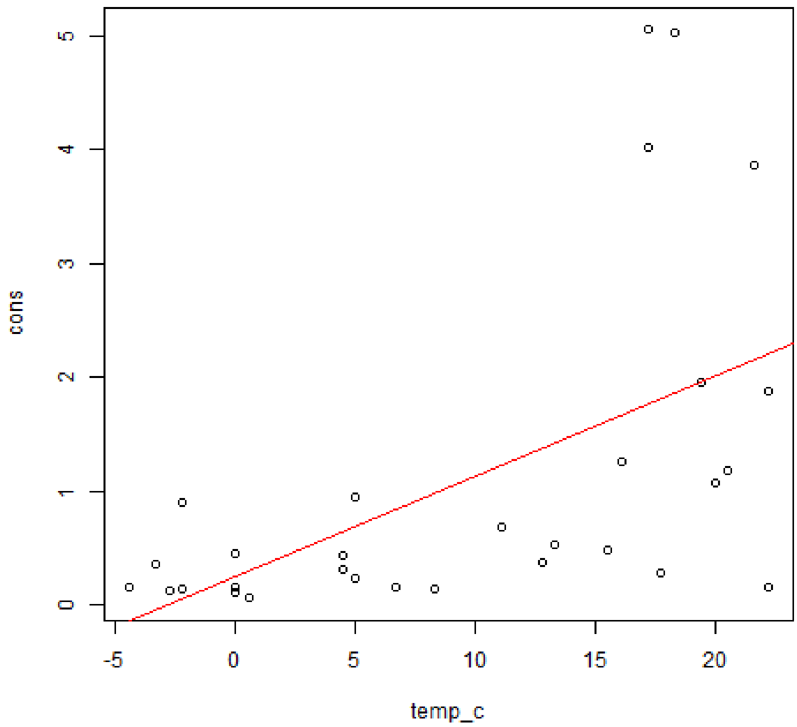

Abbildung 2: Case Study 1

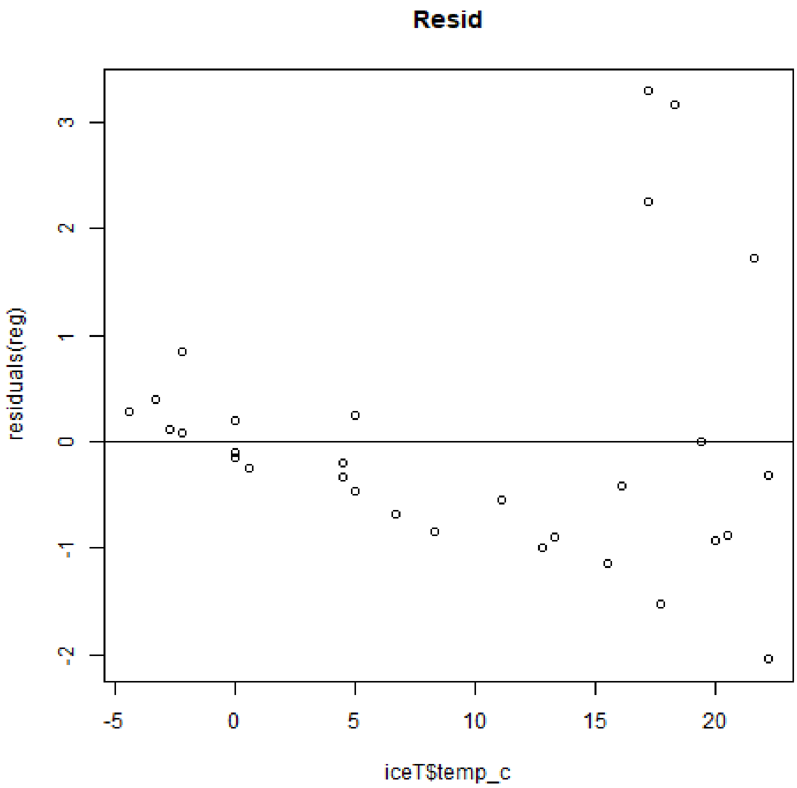

Abbildung 3: Case Study 1

```
Call:
lm(formula = log(cons) ~ temp_c, data = iceT)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.28780 -0.54652 -0.08408  0.65731  1.61052 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.52211    0.25571  -5.953 2.08e-06 ***
temp_c       0.08895    0.01957   4.545 9.61e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9615 on 28 degrees of freedom
Multiple R-squared:  0.4245,    Adjusted R-squared:  0.404 
F-statistic: 20.66 on 1 and 28 DF,  p-value: 9.615e-05
```

Abbildung 4: Case Study 1

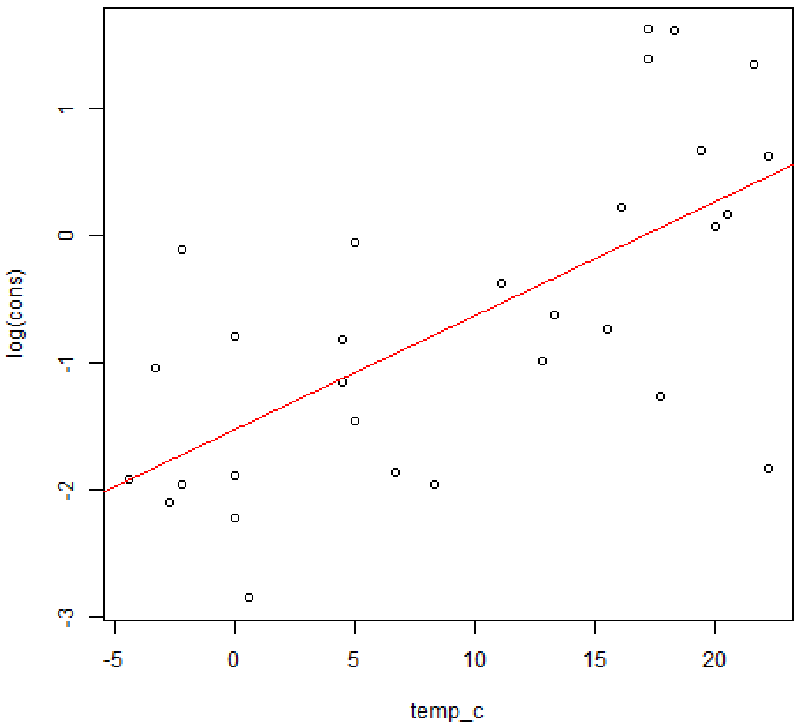

Abbildung 5: Case Study 1

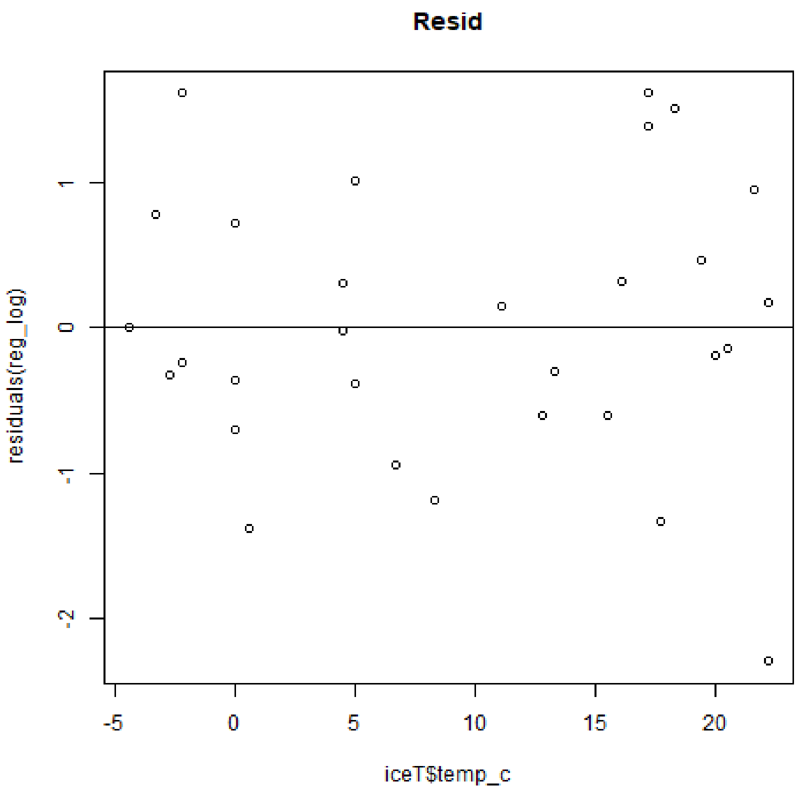

Abbildung 6: Case Study 1

A.2 Case Study 2

```
> fit <- lm(coolload ~ relcomp + wallarea + roofarea +  glazing, data = energy)
> summary(fit)

Call:
lm(formula = coolload ~ relcomp + wallarea + roofarea + glazing,
data = energy)

Residuals:
Min      1Q  Median      3Q     Max
-7.2274 -1.7729 -0.8312  1.0400 12.8147

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 279.970327  14.746220   18.99   <2e-16 ***
relcomp      -1.541746   0.093561  -16.48   <2e-16 ***
wallarea     -0.145381   0.011745  -12.38   <2e-16 ***
roofarea     -0.536435   0.022432  -23.91   <2e-16 ***
glazing       0.148180   0.009406   15.75   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.47 on xxx degrees of freedom
Multiple R-squared:  0.8676,Adjusted R-squared:  0.8669
F-statistic:  1250 on 4 and 763 DF,  p-value: < 2.2e-16

> SSR <- sum(fit$residuals^2)
> SSR
[1] 9188.46
```

Tabelle 1: Case Study 2 Output

B EViews output

B.1 Case Study 1

Dependent Variable: CONS
Method: Least Squares
Date: 10/29/19 Time: 16:26
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.248324	0.331215	0.749736	0.4597
TEMP_C	0.087789	0.025350	3.463053	0.0017

R-squared	0.299873	Mean dependent var	1.082320
Adjusted R-squared	0.274868	S.D. dependent var	1.462591
S.E. of regression	1.245465	Akaike info criterion	3.341235
Sum squared resid	43.43310	Schwarz criterion	3.434648
Log likelihood	-48.11852	Hannan-Quinn criter.	3.371119
F-statistic	11.99274	Durbin-Watson stat	2.406628
Prob(F-statistic)	0.001735		

Abbildung 7: Case Study 1

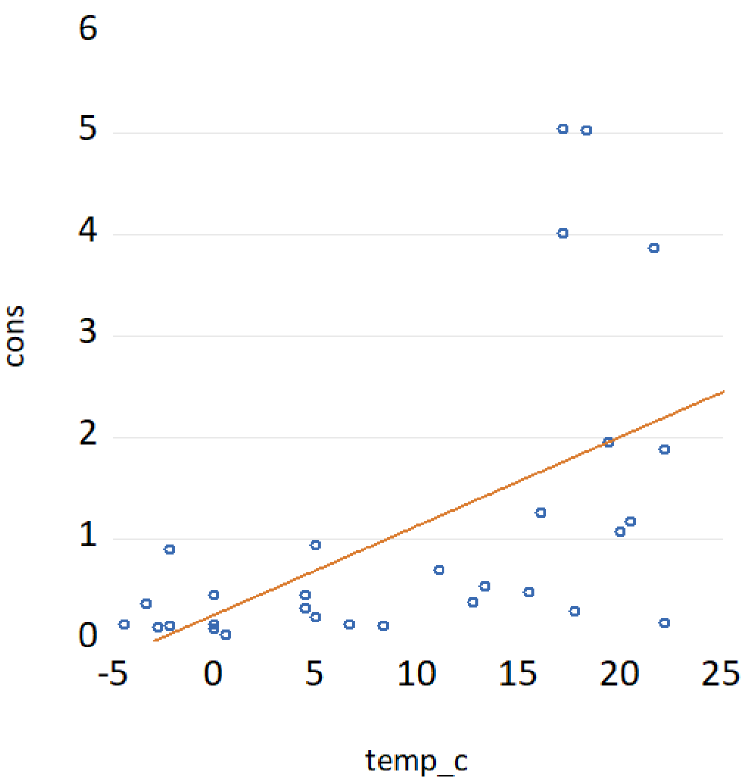

Abbildung 8: Case Study 1

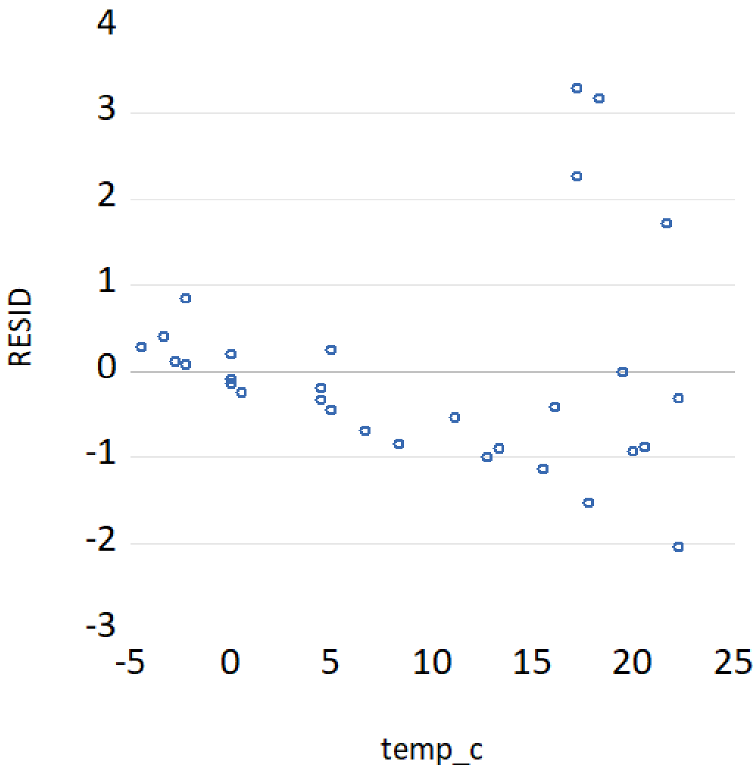

Abbildung 9: Case Study 1

Dependent Variable: LOG(CONS)
Method: Least Squares
Date: 10/29/19 Time: 15:52
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.522113	0.255708	-5.952540	0.0000
TEMP_C	0.088948	0.019571	4.544882	0.0001

R-squared	0.424531	Mean dependent var	-0.677104
Adjusted R-squared	0.403978	S.D. dependent var	1.245474
S.E. of regression	0.961536	Akaike info criterion	2.823771
Sum squared resid	25.88744	Schwarz criterion	2.917184
Log likelihood	-40.35656	Hannan-Quinn criter.	2.853655
F-statistic	20.65595	Durbin-Watson stat	2.132731
Prob(F-statistic)	0.000096		

Abbildung 10: Case Study 1

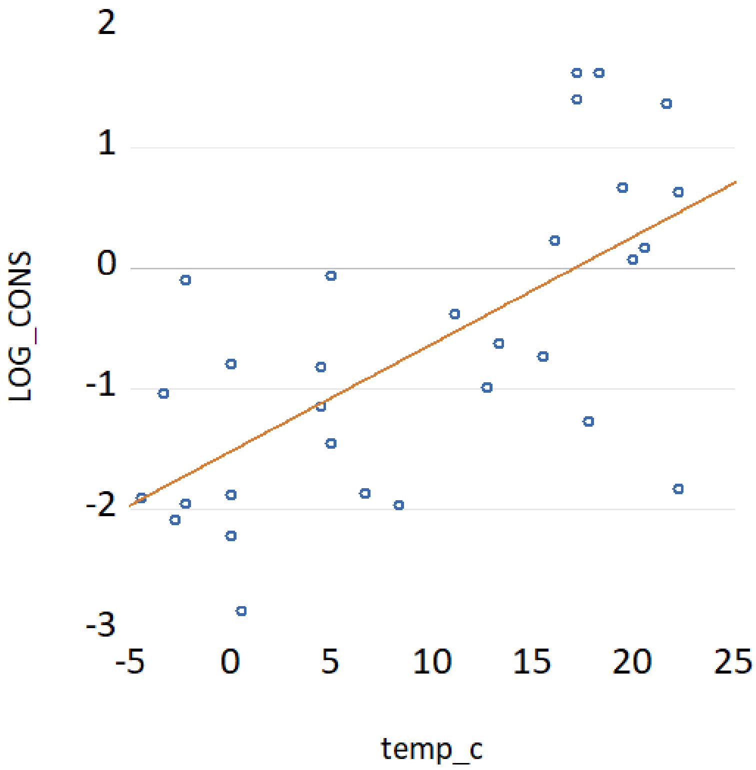

Abbildung 11: Case Study 1

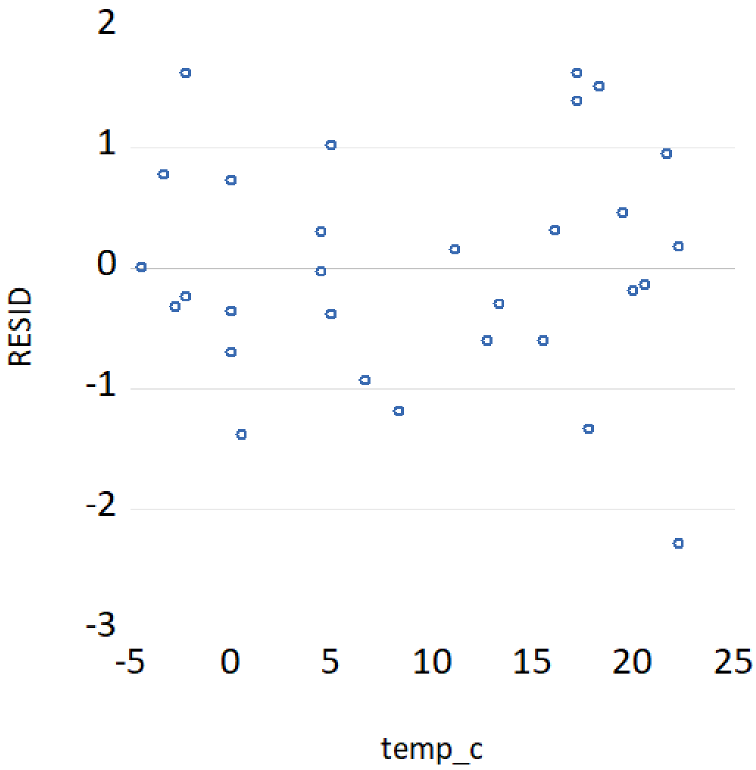

Abbildung 12: Case Study 1

B.2 Case Study 2

Dependent Variable: COOLLOAD
Method: Least Squares
Date: 10/31/19 Time: 10:26
Sample: 1 768
Included observations: 768

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	279.9703	14.74622	18.98590	0.0000
RELCOMP	-1.541746	0.093561	-16.47844	0.0000
WALLAREA	-0.145381	0.011745	-12.37822	0.0000
ROOFAREA	-0.536435	0.022432	-23.91402	0.0000
GLAZING	0.148180	0.009406	15.75428	0.0000

R-squared	0.867632	Mean dependent var	24.58776
Adjusted R-squared	0.866938	S.D. dependent var	9.513306
S.E. of regression	3.470237	Akaike info criterion	5.332812
Sum squared resid	9188.460	Schwarz criterion	5.363045
Log likelihood	-2042.800	Hannan-Quinn criter.	5.344448
F-statistic	1250.303	Durbin-Watson stat	1.072638
Prob(F-statistic)	0.000000		

Abbildung 13: Case Study 2 Output