

# **Bachelorarbeit**

im Sommersemester 2022

Analyse von Risswachstum anhand eines Multivariaten Linearen Modells

von Kaya Bayer  
26. August 2022

Dozent: Dr. Christine Müller

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Problemstellung</b>	<b>2</b>
2.1	Ziele . . . . .	2
2.2	Bauwerk und Erhebung der Daten . . . . .	2
2.3	Beschreibung des Datensatzes . . . . .	4
<b>3</b>	<b>Statistische Methoden</b>	<b>5</b>
3.1	Multivariate Lineare Regression . . . . .	6
3.1.1	Modellgleichung . . . . .	6
3.1.2	Modellannahmen . . . . .	7
3.1.3	Schätzung von $\beta$ . . . . .	8
3.1.4	Robuste Regressions-LTS-Schätzung von $\beta$ . . . . .	9
3.1.5	Schätzung von $\Sigma$ . . . . .	9
3.1.6	Modellierung der Rissbreiten . . . . .	9
3.2	Gütekriterien Multivariater Linearer Modelle . . . . .	10
3.2.1	Wilk's $\Lambda$ . . . . .	11
3.2.2	Weitere Multivariate Gütekriterien . . . . .	13
3.2.3	Minimum-Volume-Ellipsoid-Schätzung . . . . .	14
3.2.4	Minimum-Kovarianz-Determinant-Schätzung . . . . .	15
<b>4</b>	<b>Auswertung</b>	<b>15</b>
4.1	Deskriptive Analyse . . . . .	15
4.2	Anpassung der Multivariaten Linearen Modelle . . . . .	18
4.3	Modellgüte mittels Wilk's $\Lambda$ . . . . .	21
4.4	Modellgüte mittels $r_t$ und $r_d$ . . . . .	22
4.5	Anpassung der robusten Multivariaten Linearen Modelle . . . . .	25
4.6	Modellgüte basierend auf dem MVE-Schätzer . . . . .	27
4.7	Modellgüte basierend auf dem MCD-Schätzer . . . . .	29
4.8	Betrachtung des Sommer 2017 . . . . .	31
<b>5</b>	<b>Zusammenfassung</b>	<b>32</b>
	<b>Literaturverzeichnis</b>	<b>35</b>

# 1 Einleitung

Nach mehreren tragischen Brückeneinstürzen in der Vergangenheit ist die Sicherheit von Brücken von besonderer Bedeutung. Als Ursache für Einstürze kommen unter anderem Wetterextrema, Baumängel oder Überbelastungen in Frage. Da mit einem Brückeneinsturz oftmals der Verlust von Menschenleben einhergeht, muss die Sicherheit dieser Bauwerke systematisch geprüft und überwacht werden. Somit kann im Fall von Mängeln rechtzeitig gehandelt werden. Brücken sind so konstruiert, dass gewisse Schäden verkraftet werden können. Ab einem bestimmten Grad an Mängeln ist die Sicherheit jedoch nicht mehr gewährleistet und eine Brücke gilt als einsturzgefährdet. Demzufolge ist es notwendig, die Schadstellen besonders zu beobachten und dann einzugreifen, wenn die Sicherheit akut gefährdet ist. Durch diese gezielte Beobachtung und Kontrolle der Mängel lässt sich der Zeitpunkt, zu dem Reparaturarbeiten erforderlich sind, präziser bestimmen und damit Investitionen effizienter planen.

Folgende Arbeit beschäftigt sich mit der Analyse multivariater Zeitreihen von Rissbreiten eines Brückenmonitorings. Die Daten stammen dabei von einer Überwachung von Rissen der Brücke der Wittener Straße, die von Osten nach Westen über die L 705 (Sheffield Ring) in Bochum führt. Im Rahmen einer Routineuntersuchung sind dort Risse im Material festgestellt worden, woraufhin Messinstrumente installiert worden sind, welche die Breite der Risse messen. Bei voriger Analyse der Daten wurde bereits festgestellt, dass unter anderem die Temperatur einen Einfluss auf die Rissbreite hat. Mit Hilfe einer multivariaten Zeitreihenanalyse wird im Folgenden die Modellanpassung von 2 Rissen für ausgewählte Stunden des Tages getrennt analysiert. Dabei wird herausgearbeitet, wie gut sich die beiden Risse durch ein multivariates Lineares Modell modellieren lassen.

Hierzu wird der vorliegende Datensatz im nächsten Abschnitt 2 zunächst beschrieben. Es werden die einzelnen Messvariablen vorgestellt und Auffälligkeiten erläutert. In Abschnitt 3 wird ein Multivariates Lineares Modell aufgestellt, mit dem die Rissbreite vorhergesagt werden soll. Insbesondere werden verschiedene Gütekriterien vorgestellt, mit denen die Anpassung des Modells an die Daten ermittelt wird. Darüber hinaus wird herausgearbeitet, ob es auffällige Zeiträume gibt, in denen sich die Modellgüte von der Modellgüte in anderen Zeiträumen unterscheidet. Die Ergebnisse der Analyse werden in dem darauf folgenden Abschnitt 4 vorgestellt. In Abschnitt 5 werden diese erneut zusammengefasst und die gesamte Analyse wird kritisch betrachtet. Dabei wird auf mögliche Verbesserungsvorschläge bezüglich des Modells eingegangen.

## 2 Problemstellung

### 2.1 Ziele

Das Ziel des Projekts ist die multivariate Modellanpassung von 2 Rissen mittels gleitender Fenster von 28 Tagen für die verschiedenen Stunden eines Tages getrennt über einen Zeitraum vom 01.06.2016 00:00:00 Uhr bis zum 31.10.2018 23:00:00 Uhr. Dies bedeutet, dass für jedes Zeitfenster die Daten einer Stunde über 28 Tage vorliegen. Mit diesen Daten wird jeweils ein Multivariates Lineares Modell erstellt, sodass schlussendlich für alle Zeitfenster jeder Stunde ein Modell erhalten wird. Die Modellgüte über die Zeit wird folglich für die verschiedenen Stunden getrennt analysiert. Die Modellanpassung wird mit unterschiedlichen Gütekriterien herausgearbeitet und analysiert. Zusätzlich wird dies mit robusten Methoden durchgeführt. Dazu wird das Modell mit robusten Methoden aufgestellt und die Modellanpassung wird mit robusten Gütekriterien herausgearbeitet. Dabei wird insbesondere untersucht, ob es auffällige Zeiträume gibt. Hierbei wird speziell auf auffällige jahreszeitliche Veränderungen oder auf Stunden mit besonders auffälligen Ergebnissen geachtet.

### 2.2 Bauwerk und Erhebung der Daten

Die Daten stammen von einem Brückenmonitoring der Brücke der Wittener Straße, die von Osten nach Westen über die L 705 (Sheffield Ring) in Bochum führt. Diese Brücke wurde im Jahr 1961 errichtet und besitzt zwei getrennte Überbauten je Fahrtrichtung. Pro Überbau besitzt die Brücke 2 Fahrbahnen für Kraftfahrzeuge und eine Spur für eine Straßenbahnlinie. Bei einer Routineuntersuchung wurden Mängel in Form von Rissen in dem Baumaterial festgestellt, woraufhin 14 Wegaufnehmer an der Unterseite der Brücke installiert worden sind. Dabei handelt es sich um Geräte, welche die Breite eines Risses im Beton physikalisch messen können. Die Messergebnisse sind bis auf einen Tausendstel Millimeter genau. Die folgende Abbildung 1 zeigt eine Skizze des Bauwerks und die Position der daran befestigten Messinstrumente.



Abbildung 1: schematische Darstellung des Bauwerks und die Position der daran befestigten Messinstrumente (Müller; 2022)

Im westlichen Abschnitt des nördlichen Überbaus sind die Wegaufnehmer WWN1 bis WWN4 montiert. Für den östlichen Abschnitt sind die drei Wegaufnehmer WON1 bis WON3 zuständig. Im westlichen Abschnitt des südlichen Überbaus sind die 3 Wegaufnehmer WWS2 bis WWS4 installiert. Diese werden ergänzt durch die 4 Wegaufnehmer WOS1 bis WOS4 im östlichen Abschnitt.



Abbildung 2: Foto des WWN2 (Thunich; 2017)

Abbildung 2 zeigt ein Foto des WWN2. Es ist zu erkennen, dass der Wegaufnehmer nahe an der unteren Kante des Trägers angebracht ist. Der Abstand zu dieser beträgt etwa 4,5 cm. Zu den Wegaufnehmern kommen 2 Lufttemperatursensoren hinzu. Davon ist einer unterhalb der Brücke montiert und einer oberhalb der Brücke an einem Bürgersteiggeländer befestigt. Jede Messung

der Wegaufnehmer und der Lufttemperatursensoren wurde alle 2 Sekunden wiederholt. Da der südliche Überbau vor dem nördlichen Überbau abgerissen worden ist, ist der Erhebungszeitraum des südlichen Teils kürzer. Hierfür stehen Daten vom 01.06.2016 00:00:00 Uhr bis zum 24.10.2017 00:00:00 Uhr zur Verfügung, wohingegen für den nördlichen Überbau Messungen vom 01.06.2016 00:00:00 Uhr bis zum 31.10.2018 23:00:00 Uhr vorliegen (Schürmann; 2018).

## 2.3 Beschreibung des Datensatzes

In dieser Arbeit sind die Messungen stundenweise von Interesse. In dem ursprünglichen Datensatz befinden sich jedoch Messwerte im Abstand von 2 Sekunden. Aus diesem Grund wurden die Daten unter Verwendung des Medians, dem Wert in der Mitte der sortierten Daten, in stündliche Daten umgewandelt. Um einen genauen Überblick über die verschiedenen Tageszeitpunkte zu bekommen, wird jede Stunde bei der Modellierung der Rissbreiten betrachtet.

Die multivariate Modellanpassung wird an den 2 größten Rissen des nördlichen Überbaus ermittelt, d.h. es werden die Messungen des WWN2 und WWN3 betrachtet. Die Messungen an den weiteren Wegaufnehmern sind für diese Arbeit nicht von Interesse.

Anzumerken ist, dass die Messstationen zu einigen Zeitpunkten ausgefallen sind und der Datensatz demzufolge einige fehlende Werte enthält. Teilweise erstreckt sich der Ausfall der Messinstrumente lediglich über einen kurzen Zeitraum, sodass der Messwert von nur einer Stunde fehlt. In 2 Zeitperioden kam es bei den Messungen jedoch zu einem längerfristigen Ausfall, sodass für die stundenweisen Daten keine Messergebnisse zu den Zeitpunkten vom 24.10.2017 um 09:30 Uhr bis zum 24.10.2017 um 23:30 Uhr und vom 30.11.2017 um 04:30 Uhr bis zum 30.11.2017 um 23:30 Uhr vorliegen. Insgesamt fehlen die Daten zu 64 stündlichen Zeitpunkten. Da fehlende Daten im weiteren Verlauf der Analyse zu Problemen führen, werden die fehlenden Werte imputiert. Hierbei wird das Verfahren der linearen Interpolation angewendet, wobei das arithmetische Mittel, der Durchschnitt der beiden Datenpunkte unmittelbar vor und nach Ausfall der Messinstrumente, bestimmt wird und den fehlenden Wert ersetzt. Bei mehreren aufeinanderfolgenden fehlenden Werten werden alle fehlenden Werte durch den selben Wert, welcher sich aus dem Durchschnitt des vor und nach dem Ausfall ergebendem Wert bildet, ersetzt.

Für die Modellierung der Risse wird der Datensatz weiter bearbeitet und angepasst. Dazu gehört das Hinzufügen einiger neuer Variablen, die sich aus Kombination der bereits vorhandenen Daten ergeben. Überflüssige Variablen werden entfernt. Eine Übersicht über die Variablen des zur Modellierung verwendeten Datensatzes liefert folgende Tabelle.

Tabelle 1: Variablen des zur Modellierung der Rissbreiten verwendeten Datensatzes

Variable	Skalenniveau	Einheit	Beschreibung
Zeit	kardinal	Stunden	gibt den Messzeitpunkt an
$R_{1(t)}^s$	kardinal	mm	Rissbreite des Risses von WWN2 am Tag $t$ zur Stunde $s$
$R_{2(t)}^s$	kardinal	mm	Rissbreite des Risses von WWN3 am Tag $t$ zur Stunde $s$
$R_{1(t-1)}^s$	kardinal	mm	Rissbreite des Risses von WWN2 am Tag $t-1$ zur Stunde $s$
$R_{2(t-1)}^s$	kardinal	mm	Rissbreite des Risses von WWN3 am Tag $t-1$ zur Stunde $s$
$R_{(t-1)}^s$	kardinal	mm	Durchschnitt der Rissbreite des Risses von WWN2 und WWN3 am Tag $t-1$ zur Stunde $s$ , dh. $\frac{R_{1(t-1)}^s + R_{2(t-1)}^s}{2}$
$TB4_{(t)}^s$	kardinal	Celsius	Temperatur unterhalb der Brücke 4 Stunden vor Stunde $s$ am Tag $t$
$TS4_{(t)}^s$	kardinal	Celsius	Temperatur oberhalb der Brücke 4 Stunden vor Stunde $s$ am Tag $t$
$TBM_{(t)}^s$	kardinal	Celsius	gemittelte Temperatur unterhalb der Brücke über die letzten $24 \cdot 7$ Stunden vor Stunde $s$ am Tag $t$
$TSM_{(t)}^s$	kardinal	Celsius	gemittelte Temperatur oberhalb der Brücke über die letzten $24 \cdot 7$ Stunden vor Stunde $s$ am Tag $t$

Da  $s$  den Messzeitpunkt stündlich angibt, ist  $s \in 1, 2, \dots, 24$ , wobei  $s = 1$  den Zeitpunkt 00:30 Uhr bezeichnet,  $s = 2$  den Zeitpunkt 1:30 Uhr usw. Der Messzeitpunkt 00:30 Uhr meint dabei den Median der im Abstand von 2 Sekunden gemessenen Daten von 00:00 Uhr bis zum letzten gemessenen Wert vor 01:00 Uhr.

Da die Variablen  $TBM_{(t)}^s$  und  $TSM_{(t)}^s$  die gemittelte Temperatur der letzten 7 Tage vor Tag  $t$  zur Stunde  $s$  betrachten, ist  $t \in 8, 9, \dots, 883$ . Somit bedeutet  $t = 8$  für diese beiden Variablen, dass die gemittelte Temperatur der Tage 1 – 7 betrachtet wird.

### 3 Statistische Methoden

Vor der Analyse der Daten folgt ein Überblick über die verwendeten statistischen Methoden.

Im Folgenden seien  $X^\top$  das Transponierte von  $X$ ,  $X^{-1}$  die Inverse von  $X$ ,  $|X|$  die Determinante von  $X$ ,  $\text{rg}(X)$  der Rang und  $\text{tr}(X)$  die Spur von  $X$ , sofern  $X$  eine Matrix ist.

## 3.1 Multivariate Lineare Regression

### 3.1.1 Modellgleichung

Mit Hilfe des Multivariaten Linearen Modells soll eine multivariate stetige Zielvariable  $Y : \Omega \rightarrow \mathbb{R}^{n \times p}$  anhand von multiplen stetigen oder binären Kovariablen  $X \in \mathbb{R}^{n \times (q+1)}$  modelliert werden. Jedes Set der  $p$  Zielvariablen  $Y^{(1)}, \dots, Y^{(p)}$  soll von allen Kovariablen  $X_1, \dots, X_q$  vorhergesagt werden. Die  $n$  beobachteten Werte der Zufallsmatrix  $Y$  können als folgende Matrix  $y$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix} = (y^{(1)}, \dots, y^{(p)})$$

aufgeschrieben werden. Jede Spalte von  $y$  enthält jeweils einen Wert für eine der  $p$  abhängigen Zielvariablen. Jede Spalte enthält  $n$  Beobachtungen von einer der  $p$  Zielvariablen.

Die  $n$  beobachteten Werte der Kovariablen  $x_1, \dots, x_q$  können in folgender Matrix  $X$

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1q} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nq} \end{pmatrix} = (1_n, x^{(1)}, \dots, x^{(q)})$$

dargestellt werden. Die Matrix  $X$  besteht aus  $q + 1$  Spalten. Die 2. bis  $q + 1$ . Spalte geben die  $n$  Beobachtungen jeweils einer Kovariablen an. Die 1. Spalte der Matrix besteht aus konstanten 1en und wird hinzugefügt, um den Interzept des Modells zu modellieren.

Da jedes der  $y$ 's auf unterschiedliche Weise von den  $x$ 's abhängt, wird für jede Spalte von  $y$  ein Parameter  $\beta$  angegeben, der den Zusammenhang der  $y$ 's und der  $x$ 's angibt. Diese Parameter können in einer  $(q + 1) \times p$  - Matrix  $B$

$$B = \begin{pmatrix} \beta_{01} & \cdots & \beta_{0p} \\ \vdots & & \vdots \\ \beta_{q1} & \cdots & \beta_{qp} \end{pmatrix} = (\beta^{(1)}, \dots, \beta^{(p)})$$

dargestellt werden. In jeder Zeile von  $B$  werden die Parameter für eine der Kovariablen angegeben. Die 1. Spalte von  $B$  gibt dabei den Interzept des Modells an.

Da die Regression die wahren Beobachtungen selten genau vorhersagt, wird eine additive Störgröße  $\mathcal{E}$  für jede Beobachtung hinzugefügt. Somit ergibt sich die Matrix  $\mathcal{E}$

$$\mathcal{E} = \begin{pmatrix} \mathcal{E}_{11} & \cdots & \mathcal{E}_{1p} \\ \vdots & & \vdots \\ \mathcal{E}_{n1} & \cdots & \mathcal{E}_{np} \end{pmatrix},$$

welche dem Modell durch eine Addition beigelegt wird.

Folglich lässt sich das Multivariate Lineare Modell zusammengefasst als

$$Y = XB + \mathcal{E}$$

und ausführlich mit Realisierungen als

$$y = \begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1q} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_{01} & \cdots & \beta_{0p} \\ \vdots & & \vdots \\ \beta_{q1} & \cdots & \beta_{qp} \end{pmatrix} + \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1p} \\ \vdots & & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{np} \end{pmatrix}$$

darstellen (Rencher und Christensen; 2012, S.337/388).

### 3.1.2 Modellannahmen

Um eine möglichst gute Vorhersage der Zielvariablen durch die Kovariablen im Multivariaten Linearen Modell zu erhalten, werden folgende Annahmen getroffen:

1.  $E(Y) = XB$  bzw.  $E(\mathcal{E}) = 0_{n \times p}$
2.  $\text{cov}(Y_i) = \Sigma \forall i = 1, \dots, n$ , wobei  $Y_i$  die  $i$ -te Zeile von  $Y$  ist
3.  $\text{cov}(Y_i, Y_l) = 0_{p \times p}, \forall i \neq l, l = 1, \dots, n$ .

Die 1. Annahme sagt aus, dass  $Y$  im Erwartungswert gleich  $XB$  entspricht. Dies bedeutet, dass sich die Zielvariable durch eine Linearkombination der Kovariablen beschreiben lässt und der Fehlerterm im Erwartungswert 0 ist. Daraus folgt, dass die Vorhersage der Beobachtungen nach unten und oben im Mittel gleich viel von der wahren Beobachtung abweichen.

Die 2. Annahme besagt, dass alle  $n$  Beobachtungsvektoren der Zielvariable die gleiche Kovarianzmatrix  $\Sigma$  haben.  $\Sigma$  entspricht dabei

$$\Sigma = \text{cov}(Y_i) = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & & \vdots \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$$

und enthält die Varianzen und Kovarianzen von allen  $Y_{i1}, \dots, Y_{ip}$  in jedem Vektor  $Y_i, i = 1, \dots, n$ .

Die 3. Annahme sagt aus, dass alle  $n$  Beobachtungsvektoren der Zielvariablen untereinander unkorreliert sind. Es wird zwar angenommen, dass die  $p$  Beobachtungen in dem jeweiligen Beobachtungsvektor abhängig voneinander sind, jedoch sind diese unabhängig zu den Beobachtungen in den anderen Vektoren. Dies impliziert, dass die Kovarianzen für jedes  $Y_{i1}, \dots, Y_{ip}$  mit jedem

$Y_{l1}, \dots, Y_{lp}$  gleich 0 ist. Daraus ergibt sich

$$\text{cov}(Y_i, Y_l) = \begin{pmatrix} \text{cov}(Y_{i1}, Y_{l1}) & \cdots & \text{cov}(Y_{i1}, Y_{lp}) \\ \vdots & & \vdots \\ \text{cov}(Y_{ip}, Y_{l1}) & \cdots & \text{cov}(Y_{ip}, Y_{lp}) \end{pmatrix} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

(Rencher und Christensen; 2012, S.339).

### 3.1.3 Schätzung von $\beta$

Für die Schätzung der Parametermatrix  $B$  aus den vorhandenen Daten wird analog zum univariaten Fall die Methode der kleinsten Quadrate verwendet. Es werden die Parameter gesucht, die die Fehlermatrix  $\hat{\mathcal{E}}$  minimieren. Als Lösung ergibt sich der Ausdruck

$$\hat{B} = (X^\top X)^{-1} X^\top Y,$$

welcher  $\hat{\mathcal{E}}$  hinsichtlich der Spur minimiert. Die Spur von  $\hat{\mathcal{E}}$  ist durch

$$\begin{aligned} \text{tr}(\hat{\mathcal{E}}) \text{tr}((Y - X\hat{B})^\top (Y - X\hat{B})) &= \text{tr} \left( \begin{pmatrix} Y^{(1)} - X\hat{B}^{(1)} \\ \vdots \\ Y^{(p)} - X\hat{B}^{(p)} \end{pmatrix} (Y^{(1)} - X\hat{B}^{(1)}, \dots, Y^{(p)} - X\hat{B}^{(p)}) \right) \\ &= \sum_{k=1}^p (Y^{(k)} - X\hat{B}^{(k)})^2 \end{aligned}$$

gegeben.

Der Schätzer  $\hat{B}$  wird auch als *Kleinster Quadrate Schätzer* für  $B$  bezeichnet.

Der *Kleinster Quadrate Schätzer*  $\hat{B}$  kann ohne die Modellannahmen aus Abschnitt 3.1.2 bestimmt werden. Sind diese Annahmen jedoch erfüllt, hat  $\hat{B}$  folgende Eigenschaften:

1. Der Schätzer  $\hat{B}$  ist unverzerrt, sodass  $E(\hat{B}) = B$  gilt. Dies bedeutet, dass, wenn man wiederholt zufällige Stichprobenwerte ziehen würde,  $\hat{B}$  im Durchschnitt  $B$  entsprechen würde.
2. Der *Kleinster Quadrate Schätzer*  $\hat{\beta}_{jk}, k = 1, \dots, p, j = 1, \dots, q + 1$  in  $\hat{B}$  hat die minimale Varianz von allen möglichen linearen unverzerrten Schätzern für  $B$ .
3. Alle der  $\hat{\beta}_{jk}$ 's in  $\hat{B}$  sind miteinander korreliert. Diese Eigenschaft ergibt sich aus der Korreliertheit der Variablen innerhalb von  $Y$ . Die Korrelation der  $Y^{(1)}, \dots, Y^{(p)}$  führt zur Korrelation der  $\hat{\beta}$ 's in einer Spalte mit den  $\hat{\beta}$ 's in den anderen Spalten

(Rencher und Christensen; 2012, S.339/340).

### 3.1.4 Robuste Regressions-LTS-Schätzung von $\beta$

Ein robustes Schätzverfahren für die Schätzung der Koeffizientenmatrix  $B$  reagiert nicht sensibel auf Ausreißer in den Daten.

Seien  $R \in \mathbb{R}^{n \times p}$  die Residuen des Modells mit Realisierungen  $r$  und als

$$\begin{aligned} r &= \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} = \left( r^{(1)}, \dots, r^{(p)} \right) = \begin{pmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & & \vdots \\ r_{n1} & \cdots & r_{np} \end{pmatrix} \\ &= \begin{pmatrix} y_{11} & \cdots & y_{1p} \\ \vdots & & \vdots \\ y_{n1} & \cdots & y_{np} \end{pmatrix} - \begin{pmatrix} 1 & x_{11} & \cdots & x_{1q} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nq} \end{pmatrix} \begin{pmatrix} \hat{\beta}_{01} & \cdots & \hat{\beta}_{0p} \\ \vdots & & \vdots \\ \hat{\beta}_{q1} & \cdots & \hat{\beta}_{qp} \end{pmatrix} \end{aligned}$$

definiert. Dann ist  $R_{ik}(Y^{(k)}, X, \beta^{(k)}) = |Y_{ik} - X_i^\top \beta^{(k)}|$  das sogenannte absolute Residuum des  $i$ -ten theoretischen Wertes der  $k$ -ten Variable. Folglich sind  $R_{(1k)}(Y_{ik}, X, \beta^{(k)}) \leq R_{(2k)}(Y^{(k)}, X, \beta^{(k)}) \leq \dots \leq R_{(nk)}(Y^{(k)}, X, \beta^{(k)})$  die geordneten absoluten Residuen.

Die Regressions-LTS-Schätzung  $\hat{\beta}_h(Y^{(k)}, X)$  bezüglich  $h$  basierend auf  $Y^{(k)}$ ,  $k = 1, \dots, p$  und  $X$  ist definiert als

$$\hat{\beta}_h^{(k)}(Y^{(k)}, X) \in \arg \min_{\beta^{(k)}} \sum_{n=1}^h R_{(nk)}(Y^{(k)}, X, \beta^{(k)})^2,$$

wobei  $h$  die Anzahl der Residuen ist.  $h$  ergibt sich hierbei aus  $h = \lfloor (n/2) \rfloor + \lfloor ((p+1)/2) \rfloor$  (Rousseeuw; 1985, S.388-433).

### 3.1.5 Schätzung von $\Sigma$

Ein unverzerrter Schätzer für  $\Sigma = \text{cov}(Y_i)$  ist gegeben durch

$$S_e = \frac{(Y - X\hat{B})^\top (Y - X\hat{B})}{n - q - 1}.$$

$(Y - X\hat{B})^\top (Y - X\hat{B})$  ist dabei eine Matrix mit  $p$  Zeilen und  $p$  Spalten (Rencher und Christensen; 2012, S.342).

### 3.1.6 Modellierung der Rissbreiten

Bei der Modellierung der beiden Risse  $R_{1(t)}^s$  und  $R_{2(t)}^s$  werden diese durch die erklärenden Variablen  $R_{(t-1)}^s, TB4_{(t)}^s, TS4_{(t)}^s, TBM_{(t)}^s$  und  $TSM_{(t)}^s$  beschrieben. Für die gleitenden Fenster über 28 Tage werden die Daten für  $t \in K + 1, K + 2, \dots, K + 28$  mit  $K = 7, 8, \dots, 883 - 28$  benutzt, da sich der Messzeitraum über insgesamt 883 Tage erstreckt. Für  $K = 7$  ergibt sich somit  $t = 8$ . Die Variablen  $TBM_{(t)}^s$  und  $TSM_{(t)}^s$  geben für  $t = 8$  folglich die gemittelte Temperatur der Tage 1 bis

7 an. Insgesamt gibt es für jede Stunde 848 verschiedene Zeitfenster mit jeweils 28 Tagen. Bei der konkreten Modellierung durch ein Multivariates Lineares Modell, wie soeben beschrieben, erhält man das folgende Modell. Es ergibt sich  $p = 2$ , da die Breiten zweier Risse modelliert werden sollen,  $n = 28$ , da ein Zeitfenster sich über 28 Tage erstreckt und es somit 28 Beobachtungen einer Variable gibt und  $q = 5$ , da die Rissbreiten durch die 5 Kovariablen  $R_{(t-1)}^s, TB4_{(t)}^s, TS4_{(t)}^s, TBM_{(t)}^s$  und  $TSM_{(t)}^s$  beschrieben werden sollen.

Daraus ergibt sich

$$Y = \begin{pmatrix} R_{1(K+1)}^s & R_{2(K+1)}^s \\ \vdots & \vdots \\ R_{1(K+28)}^s & R_{2(K+28)}^s \end{pmatrix} \in \mathbb{R}^{28 \times 2}$$

als Matrix der Kovariablen,

$$X = \begin{pmatrix} 1 & R_{(K+1-1)}^s & TB4_{(K+1)}^s & TS4_{(K+1)}^s & TBM_{(K+1)}^s & TSM_{(K+1)}^s \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & R_{(K+28-1)}^s & TB4_{(K+28)}^s & TS4_{(K+28)}^s & TBM_{(K+28)}^s & TSM_{(K+28)}^s \end{pmatrix} \in \mathbb{R}^{28 \times 6},$$

$$B = \begin{pmatrix} \beta_{01} & \beta_{02} \\ \vdots & \vdots \\ \beta_{51} & \beta_{52} \end{pmatrix} \in \mathbb{R}^{5 \times 2}$$

und

$$\mathcal{E} = \begin{pmatrix} \mathcal{E}_{11} & \mathcal{E}_{n_{12}} \\ \vdots & \vdots \\ \mathcal{E}_{(28)1} & \mathcal{E}_{(28)2} \end{pmatrix} \in \mathbb{R}^{28 \times 2}.$$

### 3.2 Gütekriterien Multivariater Linearer Modelle

Gütekriterien für Multivariate Lineare Modelle geben an, wie gut sich ein vorhandenes Multivariates Lineares Modell an die beobachteten Daten anpasst. Dafür versuchen sie, den Anteil der durch die Kovariablen aufgeklärten Varianz der Zielvariable zu berechnen.

### 3.2.1 Wilk's $\Lambda$

Das Multivariate Gütekriterium Wilk's  $\Lambda$  kann zum Testen der Hypothese  $H_0 : B_1 = 0$  vs.  $H_1 : B_1 \neq 0$  verwendet werden. Die Matrix  $B_1$  enthält alle Zeilen der Matrix  $B$ ,

$$B = \begin{pmatrix} \beta_0 \\ B_1 \end{pmatrix} = \begin{pmatrix} \beta_{01} & \cdots & \beta_{0p} \\ \beta_{11} & \cdots & \beta_{1p} \\ \vdots & & \vdots \\ \beta_{q1} & \cdots & \beta_{qp} \end{pmatrix}$$

außer der ersten Zeile. Diese wird aus der Hypothese entfernt, da sonst impliziert wird, dass alle  $Y^{(k)}, k = 1, \dots, p$  einen Interzept von 0 haben. Aus der  $H_1$  Hypothese folgt, dass  $\beta_{jk} \neq 0$  für mind. ein  $j = 1, \dots, q; k = 1, \dots, p$  ist. Folglich wird  $H_0$  abgelehnt, sobald eine der Kovariablen einen signifikanten Einfluss auf die Zielvariable hat. Die Teststatistik  $\Lambda$  ist als

$$\Lambda = \frac{|E|}{|E + H|}$$

definiert, basierend auf der Fehler-Matrix  $E$  und der Hypothesen-Matrix  $H$ . Die Matrix  $E$  ist gegeben durch

$$E := (Y - X\hat{B})^\top (Y - X\hat{B}).$$

Dabei gilt

$$E = Y^\top Y - \hat{B}^\top X^\top Y \in \mathbb{R}^{p \times p}.$$

Dies folgt aus

$$\begin{aligned} & (Y - X\hat{B})^\top (Y - X\hat{B}) \\ &= (Y^\top - \hat{B}^\top X^\top)(Y - X\hat{B}) \\ &= Y^\top Y - Y^\top X\hat{B} - \hat{B}^\top X^\top Y + \hat{B}^\top X^\top X\hat{B} \\ &= Y^\top Y - Y^\top X(X^\top X)^{-1}X^\top Y - Y^\top X(X^\top X)^{-1}X^\top Y + Y^\top X(X^\top X)^{-1}X^\top X(X^\top X)^{-1}X^\top Y \\ &= Y^\top Y - 2Y^\top X(X^\top X)^{-1}X^\top Y + Y^\top X(X^\top X)^{-1}X^\top X(X^\top X)^{-1}X^\top Y \\ &= Y^\top Y - 2Y^\top X(X^\top X)^{-1}X^\top Y + Y^\top X(X^\top X)^{-1}X^\top Y \\ &= Y^\top Y - Y^\top X(X^\top X)^{-1}X^\top Y \\ &= Y^\top Y - \hat{B}^\top X^\top Y. \end{aligned}$$

Mit Hilfe von Projektionsmatritzen lässt sich  $E$  auch darstellen als

$$E = Y^\top (I - P)Y,$$

wobei  $P = X(X^\top X)^{-1}X^\top$  entspricht und  $I \in \mathbb{R}^{n \times n}$  eine Einheitsmatrix darstellt.  $I - P$  hat den Rang  $\text{rg}(I - P) = n - (q + 1)$ .

Die Matrix  $H$  ist als

$$H := Y^\top X(X^\top X)^{-1}X^\top Y - n\bar{Y}\bar{Y}^\top = \hat{B}^\top X^\top Y - n\bar{Y}\bar{Y}^\top \in \mathbb{R}^{p \times p}$$

definiert.  $\bar{Y}$  ergibt sich aus  $(\bar{Y}^{(1)}, \dots, \bar{Y}^{(p)})$ , wobei  $\bar{Y}^{(k)}$  als  $\bar{Y}^{(k)} = \frac{1}{n} \mathbf{1}_n^\top Y^{(k)} \in \mathbb{R}$  definiert ist. Als Darstellung mithilfe von Projektionsmatrizen entspricht  $H$

$$H = Y^\top (P - P_0)Y.$$

$P_0$  ist hierbei als  $P_0 = X_0(X_0^\top X_0)^{-1}X_0^\top$  definiert, wobei  $X_0 \in \mathbb{R}^{n \times (q+1)}$  die Matrix der Kovariablen des Modells angibt, in dem keine der Kovariablen einen signifikanten Einfluss hat. Dies trifft unter der Annahme von  $H_0$  zu.  $P - P_0$  hat hierbei den Rang  $\text{rg}(P - P_0) = q$ .

$E + H$  setzt sich aus

$$\begin{aligned} E + H &= (Y^\top Y - \hat{B}^\top X^\top Y) + (\hat{B}^\top X^\top Y - n\bar{Y}\bar{Y}^\top) \\ &= Y^\top Y - n\bar{Y}\bar{Y}^\top \in \mathbb{R}^{p \times p} \end{aligned}$$

zusammen. Mit Hilfe von Projektionsmatrizen lässt sich  $E + H$  als

$$E + H = Y^\top P Y$$

darstellen, wobei  $P$  den Rang  $\text{rg}(P) = n - 1$  hat.

Die Verteilung der Teststatistik  $\Lambda$  setzt sich aus dem Verhältnis der Determinanten von 2 unabhängigen Wishart-Verteilungen zusammen. Unter der Annahme, dass  $H_0$  gilt, liegt  $E$  und  $H$  die Wishart-Verteilung mit Parametern

$$E \sim W(n - \text{rg}(X), \Sigma, 0) \text{ und } H \sim W(\text{rg}(X) - \text{rg}(X_0), \Sigma, 0)$$

zugrunde. Die Teststatistik

$$\Lambda = \frac{|E|}{|E + H|}$$

kann durch die Multiplikation mit  $|E^{-1}|$  in die Form

$$\Lambda = \frac{|I|}{|I + H E^{-1}|} = |I + H E^{-1}|^{-1}$$

gebracht werden.  $I \in \mathbb{R}^{p \times p}$  ist dabei die Einheitsmatrix. Unter  $H_0$  hat die Teststatistik  $\Lambda$  eine Verteilung der Form

$$\Lambda \sim \Lambda(p, d, n - \text{rg}(X)),$$

wobei  $d = \text{rg}(X) - \text{rg}(X_0)$  ist. Die Teststatistik  $\Lambda$  nimmt Werte von  $0 \leq \Lambda \leq 1$  an. Je kleiner  $\Lambda$ , desto größer ist der Anteil der durch die Kovariablen erklärten Varianz der Zielvariablen.  $H_0$  wird abgelehnt, wenn die Teststatistik  $\Lambda$  kleiner gleich dem  $\alpha$ -Quantil der Verteilung  $\Lambda(p, d, n - \text{rg}(X))$  ist und somit  $\Lambda \leq \Lambda(p, d, n - \text{rg}(X))$  gilt. Ist  $H$  "groß" auf Grund von hohen Werten der  $\hat{\beta}'_{jk}s$ , dann wird erwartet, dass  $|E + H|$  bedeutend größer als  $|E|$  ist. Folge dessen wird die Teststatistik  $\Lambda$  klein und  $H_0$  wird abgelehnt. Mit einem "großen"  $H$  wird gemeint, dass die Einträge auf der Hauptdiagonalen von  $H$  große Werte annehmen. Unter  $H_0$  lässt sich die Verteilung von  $\Lambda$  durch eine F-Verteilung approximieren. Dazu seien

$$\begin{aligned} r &= \text{rg}(X), \\ d &= \text{rg}(X) - \text{rg}(X_0), \\ s &= \frac{pd}{2} + 1, \\ f &= (n - r) + d - \frac{1}{2}(d + p + 1), \end{aligned}$$

und

$$t = \begin{cases} [(p^2 d^2 - 4)/p^2 + d^2 - 5]^{\frac{1}{2}} & \text{falls } \min(p, d) \geq 2 \\ 1 & \text{falls } \min(p, d) = 1 \end{cases}.$$

Dann gilt

$$\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{ft - s}{pd} \sim F(pd, ft - s).$$

Bei einem Test zum Signifikanzniveau  $\alpha$  wird das  $(1 - \alpha)$ -Quantil der  $F(pd, ft - s)$ -Verteilung als kritischer Wert angesehen, da die Teststatistik  $\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \frac{ft - s}{pd} \sim F(pd, ft - s)$  groß wird, wenn  $\Lambda$  klein wird.

Nimmt das Gütekriterium Wilk's  $\Lambda$  kleine Werte an, so spricht dies für eine gute Modellgüte (Rencher und Christensen; 2012, S.344/344), (Christensen; 2019, S.366-372).

### 3.2.2 Weitere Multivariate Gütekriterien

Um ein weiteres Gütemaß zwischen den  $y$ 's und den  $x$ 's zu definieren, wird zunächst die Matrix  $R^2$  benötigt. Diese setzt sich aus

$$R^2 = (Y^\top Y - n\bar{Y}\bar{Y}^\top)^{-1}(\hat{B}^\top X^\top Y - n\bar{Y}\bar{Y}^\top) = (E + H)^{-1}H \in \mathbb{R}^{p \times p}$$

zusammen. Daraus lassen sich die multivariaten Gütekriterien  $r_t^2$  und  $r_d^2$  definieren, welche einen Wertebereich von  $0 \leq r_t^2 \leq 1$  und  $0 \leq r_d^2 \leq 1$  haben. Sie sind definiert als

$$r_t^2 = \frac{\text{tr}(R^2)}{p} = \frac{1}{p} \text{tr}((E + H)^{-1}H)$$

und

$$r_d^2 = |R^2| = \frac{|H|}{|E + H|}.$$

Hierbei gilt, je größer der Wert der Gütekriterien, desto größer ist der Anteil der durch die Kovariablen erklärten Varianz der Zielvariablen. Allerdings weisen beide Gütemaße die Eigenschaft auf, dass sie den wahren Zusammenhang zwischen den  $Y$ 's und  $X$ 's unterschätzen. Insbesondere gilt dies für  $r_d^2$ .

Nehmen die Gütekriterien  $r_t^2$  und  $r_d^2$  große Werte an, dann spricht dies für eine gute Güte des Modells (Rencher; 1998, S.289).

### 3.2.3 Minimum-Volume-Ellipsoid-Schätzung

Die Minimum-Volumen-Ellipsoid-Schätzung, kurz MVE-Schätzung, ist ein robustes Maß für die Schätzung der Güte. Sie beruht auf der Streuungsschätzung der Residuen  $R \in \mathbb{R}^{n \times p}$  des Multivariaten Linearen Modells.

Die empirische Kovarianzmatrix  $S_r$  der Residuen mit Realisierungen  $r$  ist die  $p \times p$ -Matrix der empirischen Kovarianzen von allen  $r_{i1}, \dots, r_{ip}$  in jedem Vektor  $r_i, i = 1, \dots, n$ . Somit entspricht  $S_r$

$$S_r = \text{cov}(r_i) = \begin{pmatrix} s_{11} & \cdots & s_{1p} \\ \vdots & & \vdots \\ s_{p1} & \cdots & s_{pp} \end{pmatrix}.$$

Die Kovarianzmatrix stellt zwar eine Verallgemeinerung der Varianz dar, ist jedoch als Maß für die Streuung der Daten ungeeignet, da sie eine Matrix ist. Als positive eindimensionale Größe kann das Volumen der Ellipse, das durch die Kovarianzmatrix  $S_r$  und einem Lokationsvektor  $\mu \in \mathbb{R}^p$  gegeben ist, als Maß für die Streuung verwendet werden. Das Volumen einer Ellipse

$$\mathcal{E}(\Sigma) := \{z_0 \in \mathbb{R}^p; (z_0 - \mu)^\top \Sigma^{-1} (z_0 - \mu) \leq c\}$$

ist proportional zu  $|\Sigma|$ . Also ist das Volumen von

$$\mathcal{E}(S_r) := \{z_0 \in \mathbb{R}^p; (z_0 - \mu)^\top S_r^{-1} (z_0 - \mu) \leq c\}$$

proportional zur Determinante  $|S_r|$  der Kovarianzmatrix  $S_r$ .  $c \in \mathbb{R}^+ < \infty$  ist dabei eine obere Schranke des Volumens der Ellipse.

Für  $R \in \mathbb{R}^{n \times p}$  ist die MVE-Schätzung für die Kovarianzmatrix  $\Sigma$  gegeben durch

$$\hat{\Sigma}_{MVE}(r) \in \arg \min \left\{ |(\Sigma)|, \Sigma \in \mathbb{R}^{p \times p} \text{ mit } \sum_{i=1}^p 1_{\mathcal{E}(\Sigma)}(r_n) \geq \frac{p+n+1}{2} \right\}.$$

Die Größe  $|\hat{\Sigma}_{MVE}(r)|$  beschreibt die Streuung von  $\Sigma$  und den Zusammenhang der  $p$  Residuen  $r^{(1)}, \dots, r^{(p)}$ . Somit ist die MVE-Schätzung ein robustes Gütemaß für die Modellanpassung.

Geringe Werte der MVE-Schätzung sprechen folglich für eine gute Modellgüte (Rousseeuw; 1985, S.388-433).

### 3.2.4 Minimum-Kovarianz-Determinant-Schätzung

Die Minimum-Kovarianz-Determinant-Schätzung, kurz MCD-Schätzung, ist ein weiteres robustes Gütemaß. Sei für  $h \in \{[(p+n+1)/2], \dots, n\}$

$$\mathcal{R}^h(r) := \{(r_{i_1}, \dots, r_{i_h})^\top; \{i_1, \dots, i_h\} \text{ ist eine } h\text{-elementige Teilmenge von } \{1, \dots, n\}\}$$

die Menge aller möglichen Beobachtungsmatrizen basierend auf  $h$  Vektoren aus den  $n$  Vektoren der Residuen  $r_1, \dots, r_n \in \mathbb{R}^p$ . Die MCD-Schätzung für die Kovarianzmatrix  $\Sigma$  ist gegeben durch

$$\hat{\Sigma}_{MCD}(r) = S_{r_*^h},$$

wobei

$$r_*^h \in \arg \min \{|(S_{r^h}; r^h \in \mathcal{R}^h(r))|\}$$

gilt. Die Größe  $|\hat{\Sigma}_{MCD}(r)|$  ist somit ein robustes Gütemaß für die Modellanpassung des Multivariaten Linearen Modells.

Nimmt die MCD-Schätzung kleine Werte an, dann spricht dies für eine gute Modellanpassung (Rousseeuw; 1985, S.388-433).

## 4 Auswertung

Folgende Abbildungen und Ergebnisse werden mit Hilfe der statistischen Software R erhalten (R Core Team; 2021).

### 4.1 Deskriptive Analyse

Um sich zunächst einen groben Überblick über die Datensituation zu verschaffen, werden die Daten zu Beginn der Auswertung deskriptiv analysiert. Hierbei wird untersucht, ob sich bereits anhand der deskriptiven Analyse der Rissbreiten sowie anhand der deskriptiven Analyse der Temperatur Auffälligkeiten feststellen lassen. Hierbei wird besonders auf auffällige jahreszeitliche Veränderungen geachtet.

Für die Analyse der Rissbreiten werden die Messungen der beiden Wegaufnehmer WWN2 und

WWN3 in folgender Abbildung 3 graphisch dargestellt. Somit wird eine Übersicht der Rissbreiten beider Risse über den gesamten Messzeitraum erhalten. Auf der  $x$  - Achse der Graphik ist der gesamte Messzeitraum vom 01.06.2016 bis zum 31.10.2018 dargestellt. Auf der  $y$  - Achse befindet sich die Rissbreite in mm.

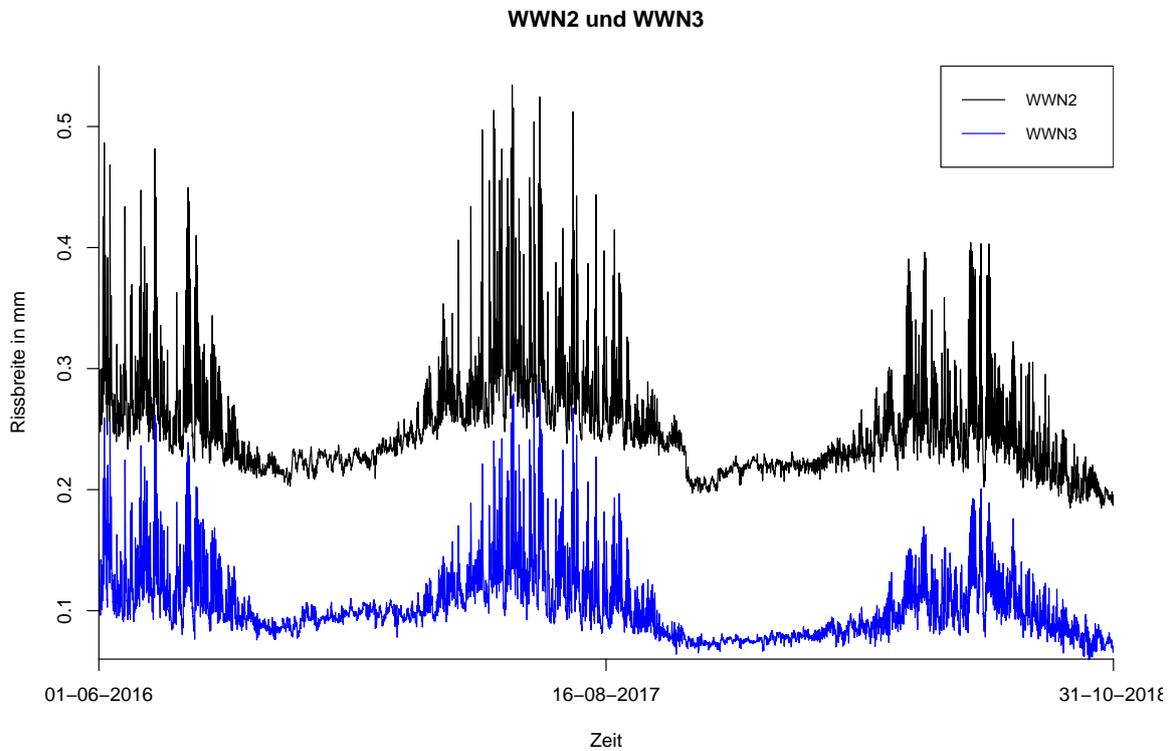


Abbildung 3: Messergebnisse des WWN2 und WWN3 über den gesamten Messzeitraum

Im Folgenden wird der Riss, der durch den WWN2 erfasst wird, als Riss 1 und der Riss, an dem WWN3 installiert ist, als Riss 2 bezeichnet. Des Weiteren stellt die schwarze Kurve in dieser Arbeit immer die Daten für Riss 1 und die blaue Kurve immer die Daten für Riss 2 dar. Zunächst fällt auf, dass sich die Schwankungen der Rissbreite von Riss 1 und Riss 2 ähnlich zueinander verhalten. Riss 1 ist ca. doppelt so groß wie Riss 2, jedoch ähneln sich die Stärke sowie der Zeitpunkt, zu dem die Schwankungen auftreten, sichtbar stark. Es fällt auf, dass die Rissbreite beider Risse über die Sommermonate stärker variiert als über die Wintermonate. Insbesondere im Sommer des Jahres 2017 schwanken die Rissbreiten stärker als zu den anderen Zeitpunkten. Die Schwankungen sind ebenfalls stärker als in den Sommermonaten der Jahre 2016 und 2018. Auch insgesamt sind die Risse im Sommer größer als im Winter. Hierbei lässt sich diese Auffälligkeit besonders im Sommer 2017 feststellen. Die stärkeren Schwankungen und die größeren Risse in den Sommermonaten lassen sich dadurch erklären, dass sowohl die Temperatur im Allgemeinen höher als auch die Temperaturunterschiede zwischen Tag und Nacht in diesen Monaten größer als im Winter sind.

Ob die Temperaturschwankungen auch bei der graphischen Analyse der Temperaturen unterhalb und oberhalb der Brücke zu beobachten sind, wird in folgender Abbildung 4 dargestellt. Dazu werden die Messungen der Temperatursensoren auf der Schattenseite und auf der Sonnenseite der Brücke betrachtet. Dabei wird auf der  $x$  - Achse der Graphiken der gesamte Messzeitraum dargestellt. Auf der  $y$  - Achse befindet sich die Temperatur in Grad Celsius.

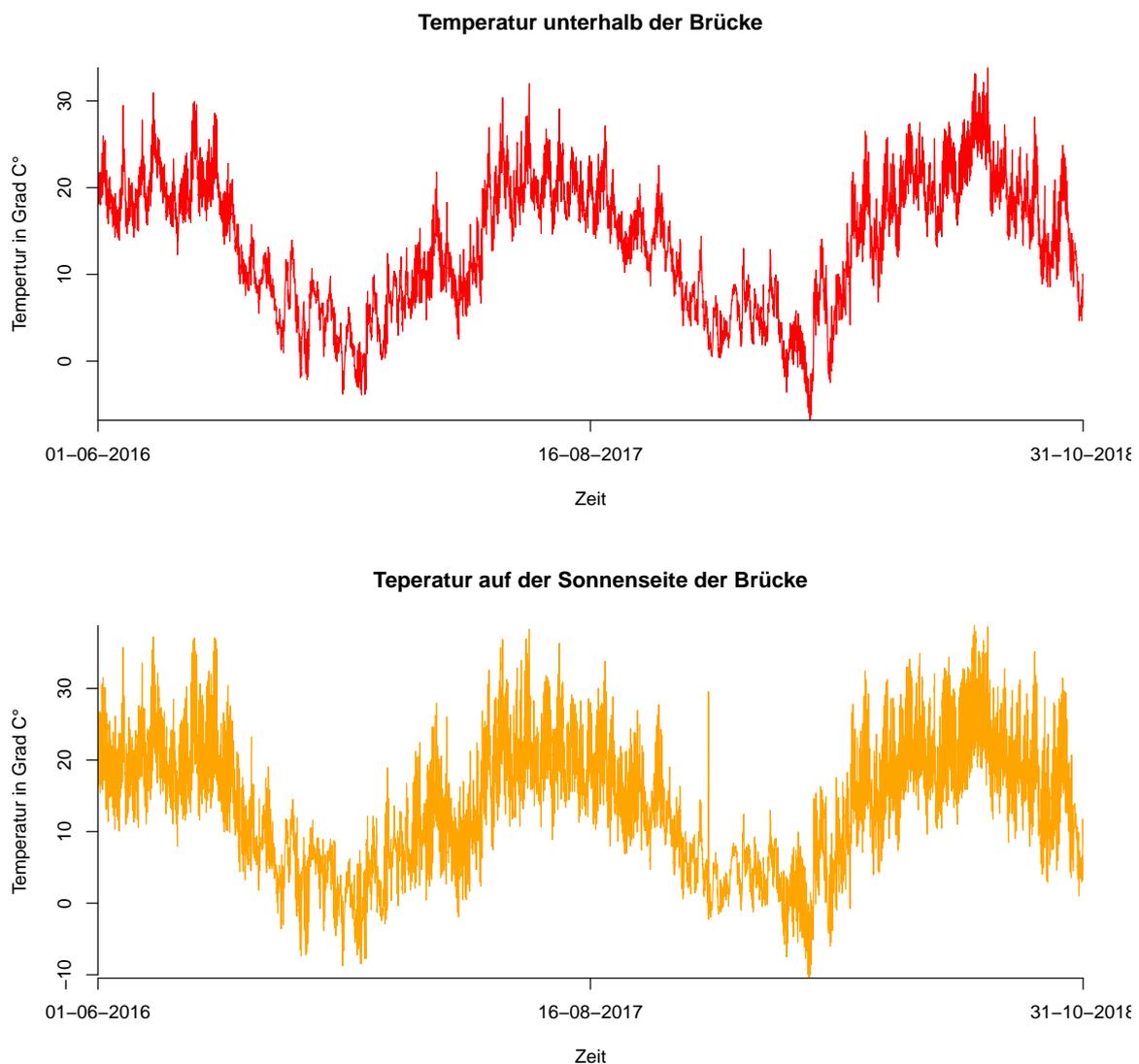


Abbildung 4: Messergebnisse des Temperatursensoren unterhalb und oberhalb der Brücke über gesamten Messzeitraum

Es ist zu erkennen, dass der Verlauf der Temperaturen oberhalb und unterhalb der Brücke sichtbar ähnlich zueinander ist. In den Sommermonaten wurden im Allgemeinen höhere Temperaturen gemessen als im Winter. Dabei ist kein Unterschied zwischen den Sommermonaten der Jahre 2016, 2017 und 2018 zu erkennen. Wie bereits durch die deskriptive Analyse der Risse vermutet, sind die Schwankungen der Temperaturkurven in den Sommermonaten stärker und häufiger als in

den Wintermonaten. Dies ist auf einen größeren Temperaturunterschied zwischen Tag und Nacht zurückzuführen. Vor allem auf der Oberseite der Brücke tritt dies verstärkt auf. Hier ist zu erkennen, dass die Schwankungen in einem deutlich geringeren Zeitabstand auftreten. Dies ist durch den starken Einfluss der Sonne auf die Temperatur und auf das Baumaterial zurückzuführen. Besonders in den Sommermonaten ist der Einfluss der Sonne auf die Temperatur groß. Somit lassen sich die größeren und häufigeren Schwankungen der Temperatur, insbesondere im Sommer, auf der Sonnenseite oberhalb der Brücke erklären. Da sowohl die Außentemperatur als auch die Sonne einen Einfluss auf das Baumaterial hat und dieses bei hoher Temperatur erhitzen, dehnen sich als Folge dessen die Risse aus. Bei sinkender Temperatur des Baumaterials ziehen sich die Risse wieder zusammen.

## 4.2 Anpassung der Multivariaten Linearen Modelle

Da die Modellanpassung des Multivariaten Linearen Modells über Zeitfenster von 28 Tagen für die Stunden getrennt von Interesse ist, werden zunächst die Modelle für jede Stunde einzeln aufgestellt und angepasst. Es werden die Schätzungen für die Koeffizienten der Modelle ermittelt, welche für die Kovariablen  $R_{(t-1)}^s$ ,  $TB4_{(t)}^s$ ,  $TS4_{(t)}^s$ ,  $TBM_{(t)}^s$  und  $TSM_{(t)}^s$  und den Interzept erhalten werden. Um einen groben Überblick über den Verlauf zu bekommen, werden die Koeffizienten für beide Risse über den gesamten Messzeitraum graphisch dargestellt. Ebenfalls lassen sich eventuelle Thesen über die Modellgüte anhand des Verlaufs der Koeffizienten aufstellen.

Folgende Abbildung 5 zeigt die geschätzten Koeffizienten über den Verlauf des Messzeitraums für die einzelnen Kovariablen für den Messzeitpunkt 06:30 Uhr. Auf der  $x$ - Achse ist der gesamte Messzeitraum dargestellt. Dabei ist anzumerken, dass als Zeitpunkt das mittlere Datum eines jeweiligen Zeitfensters benutzt worden ist. Auf der  $y$  - Achse befindet sich der Wert des jeweiligen geschätzten Koeffizienten.

# Koeffizienten der Stunde 06:30 Uhr

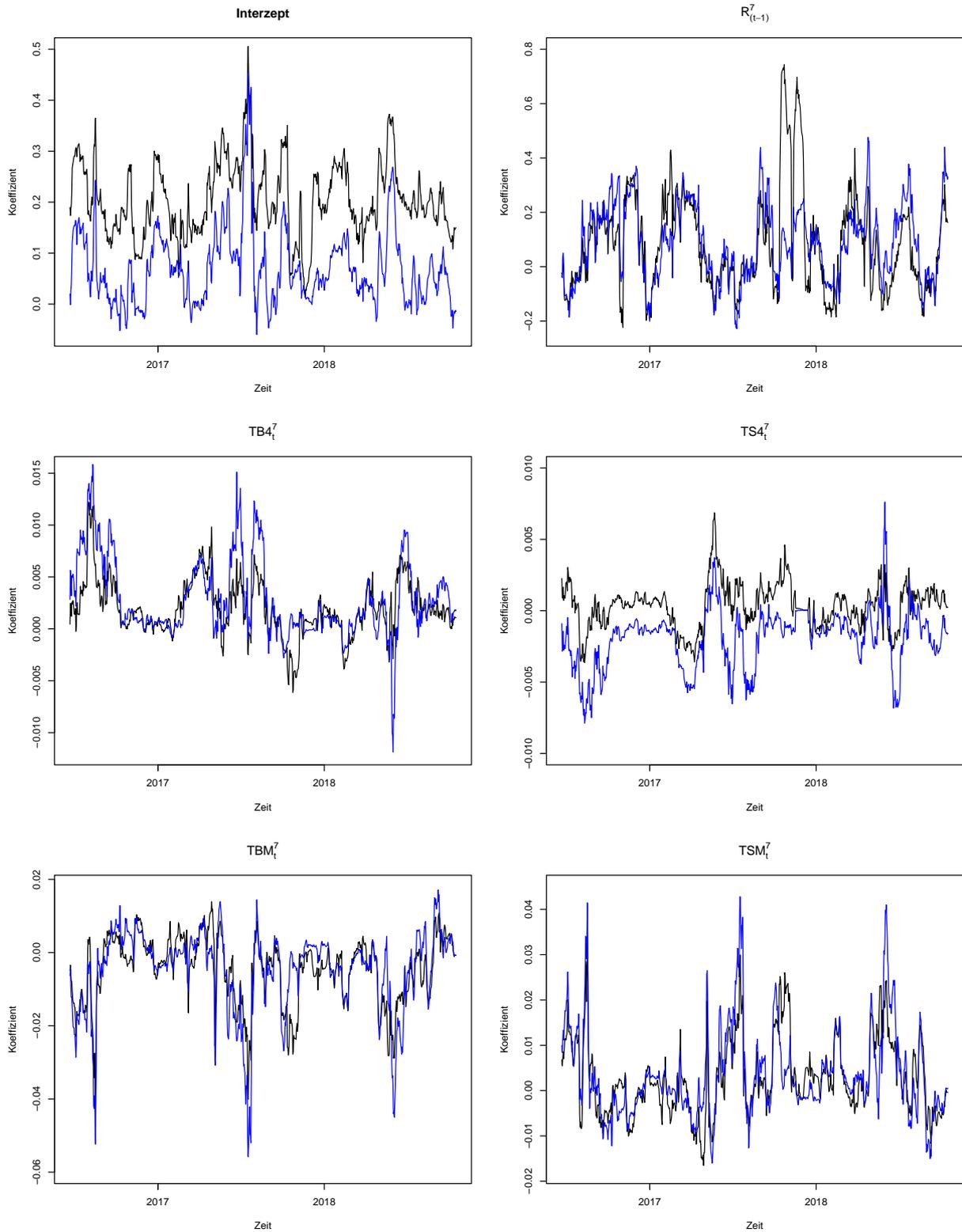


Abbildung 5: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 06:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

Zunächst ist zu beachten, dass sich die Koeffizienten in unterschiedlichen Größenordnungen bewegen. Folge dessen sind die  $x$  - Achsen Skalierungen der einzelnen Graphiken verschieden.

Bei Betrachtung der Abbildung ist zu erkennen, dass sich die Koeffizientenschätzungen beider Risse sichtbar ähnlich zueinander verhalten. Sowohl die Schwankungen über die Zeit hinweg als auch die Größenordnung der Schätzungen unterscheiden sich nur gering voneinander.

Als Ausnahme lässt sich der Interzept ansehen. Dessen Schwankungen sind im Verlauf des Messzeitraums für Riss 1 und Riss 2 zwar ähnlich zueinander, jedoch sind die Größenordnungen verschieden. Dies resultiert daraus, dass Riss 1 fast doppelt so groß wie Riss 2 ist, wie bei der deskriptiven Analyse deutlich geworden ist. Des Weiteren fällt auf, dass die Größenordnung des Koeffizienten der Kovariable  $R_{(t-1)}^s$  am größten ist. Die Variable  $R_{(t-1)}^s$  gibt das arithmetische Mittel der beiden Risse am Vortag an. Vergrößern sich die Risse am Vortag, dann ist auch eine Vergrößerung des Risses am darauf folgenden Tag zu erwarten.

Ebenfalls ist bei der Betrachtung der Abbildung zu beobachten, dass sich die Koeffizienten der Kovariablen  $TB4_{(t)}^s$  und  $TS4_{(t)}^s$  in der selben Größenordnung bewegen. Der Verlauf der Koeffizienten im Messzeitraum scheint beinahe spiegelverkehrt zu sein. Dabei meint spiegelverkehrt hierbei, dass die Koeffizienten um die  $y$  - Achse herum gespiegelt sind. Dies kann vermutlich durch eine hohe Korrelation der beiden Variablen erklärt werden. Steigt die Temperatur auf der Schattenseite unterhalb der Brücke an, dann steigt die Temperatur auch auf der Sonnenseite oberhalb der Brücke an. Bei Betrachtung der Koeffizienten der Kovariablen  $TBM_{(t)}^s$  und  $TSM_{(t)}^s$  ist ebenfalls zu beobachten, dass der Verlauf der Koeffizienten fast spiegelverkehrt zu sein scheint. Des Weiteren fällt sowohl bei den Kovariablen  $TB4_{(t)}^s$  und  $TS4_{(t)}^s$  als auch bei den Kovariablen  $TBM_{(t)}^s$  und  $TSM_{(t)}^s$  auf, dass die Koeffizienten in den Sommermonaten extremere Werte als in den Wintermonaten annehmen. Es sind extremere Werte sowohl im positiven als auch im negativen  $x$  - Achsenbereich zu beobachten. Auch die Schwankungen der Koeffizientenkurven sind im Sommer stärker und häufiger als im Winter. Zudem ist zu erkennen, dass die Einflussrichtung der Koeffizientenschätzungen der Kovariablen im Winter und im Sommer verschieden ist.

Anhand der Graphiken der Koeffizienten über den Messzeitraum hinweg lässt sich vermuten, dass es einen Unterschied bei der Güte der Modellanpassung in den verschiedenen Jahreszeiten gibt. Die starken und häufigen Schwankungen in den Sommermonaten lassen die These aufstellen, dass die Modellgüte in den Sommermonaten schlechter als in den anderen Jahreszeiten ist.

Auch bei der graphischen Analyse der Koeffizienten der restlichen Stunden fällt Ähnliches auf, sodass die These der unterschiedlichen Modellgüte in den Sommer - und Wintermonaten verfestigt werden kann. Die Abbildungen A13, A14, ..., A35 der graphischen Darstellungen der weiteren Stunden befinden sich im Anhang.

### 4.3 Modellgüte mittels Wilk's $\Lambda$

Um die Modellanpassung des Multivariaten Linearen Modells für die verschiedenen Stunden über den gesamten Messzeitraum zu untersuchen, wird die Modellanpassung zunächst anhand des Gütekriteriums Wilk's -  $\Lambda$  herausgearbeitet. Dazu wird dieses für jedes Zeitfenster von 28 Tagen über den gesamten Messzeitraum für jede Stunde getrennt bestimmt. Um die Modellgüte über die Zeit hinweg zu analysieren, wird das Gütekriterium für jede Stunde separat über den Messzeitraum graphisch dargestellt. Die folgende Abbildung 6 stellt das Gütekriterium Wilk's  $\Lambda$  über den gesamten Messzeitraum für die Messzeitpunkte 00:30 Uhr, 02:30 Uhr, ... , 22:30 Uhr dar. Auf der  $x$  - Achse jeder Graphik befindet sich der Messzeitraum und auf der  $y$  - Achse wird die Güte dargestellt.

## Wilk's Lambda

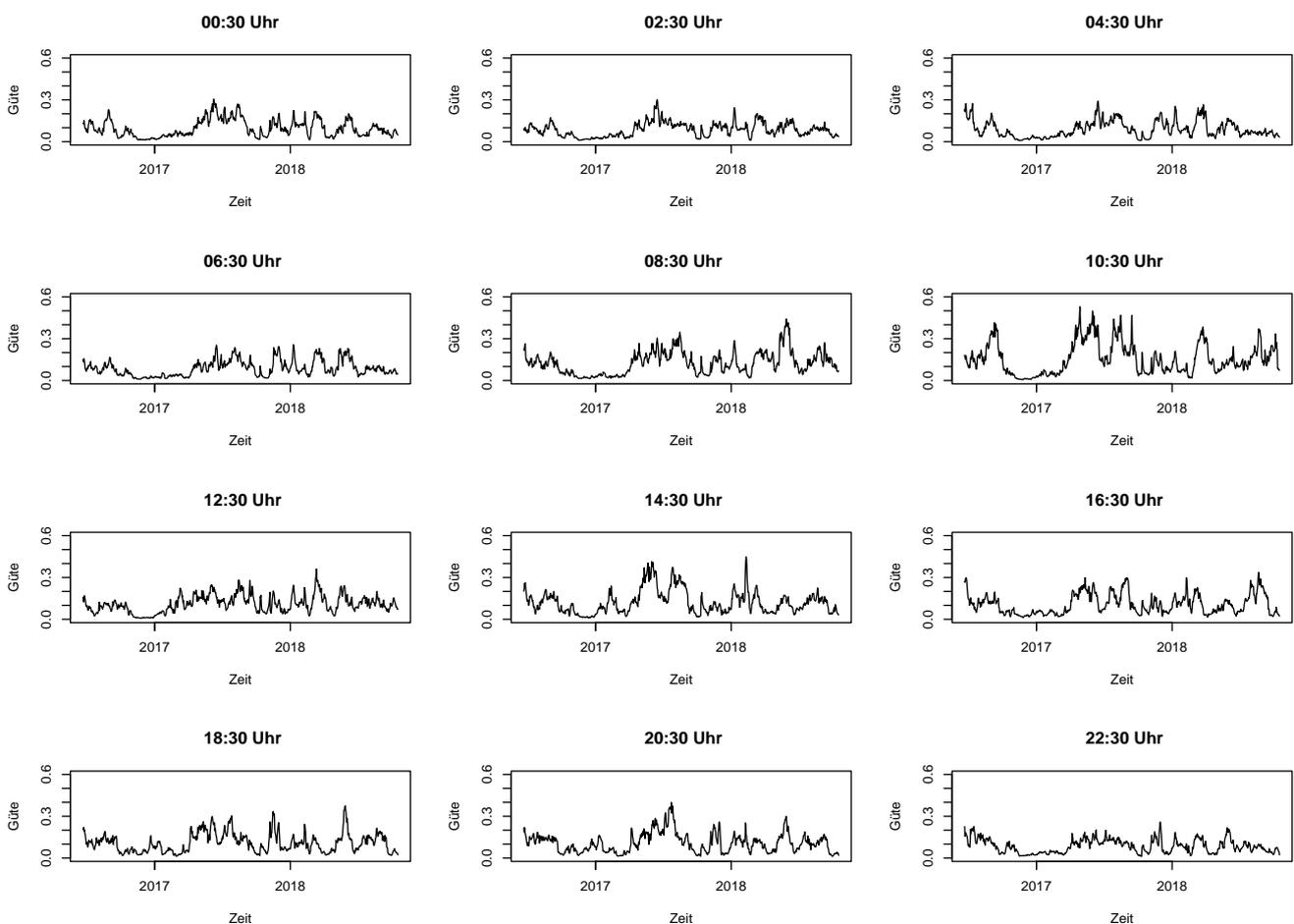


Abbildung 6: Gütekriterium Wilk's  $\Lambda$  zur Ermittlung der Modellanpassung eines Multivariaten Linearen Modells zur Vorhersage von Rissen im Baumaterial einer Brücke

Bei Betrachtung der Abbildung fällt auf, dass Wilk's  $\Lambda$  in den Wintermonaten niedrigere Werte als in den Sommermonaten annimmt. Kleine Werte sprechen hierbei für eine gute Modellanpassung. Besonders im Winter 2017 ist dies zutreffend. In den Sommermonaten nimmt das Gütekriterium teilweise Werte von bis zu 0,5 an. Dies bedeutet, dass 50% der Varianz der Zielvariable der Rissbreiten nicht durch das Modell erklärt werden können. In den Wintermonaten sind dies lediglich weniger als 5% der Varianz, die nicht erklärt werden können. Ebenfalls lässt sich aus der Graphik entnehmen, dass das Gütekriterium für die nächtlichen Messzeitpunkte 00:30 Uhr, 02:30 Uhr, 04:30 Uhr, 06:30 Uhr und 22:30 Uhr niedrigere Werte als für die Mittagsstunden annimmt. Somit scheint die Modellanpassung zu den Uhrzeiten, bei denen die Tage ihre Tieftemperaturen annehmen, besser zu sein. Hinzu kommt, dass die Temperaturschwankungen zwischen den verschiedenen Tagen nachts nicht so stark variiert wie mittags. Vor allem für die Wintermonate ist dies zutreffend, wie bereits in der deskriptiven Analyse der Temperatur deutlich wurde. Für die Messzeitpunkte 01:30 Uhr, 03:30 Uhr, ... , 23:30 Uhr lässt sich ein gleiches Schema der Schwankungen der Koeffizienten über den Messzeitraum feststellen. Die Abbildung A36 dazu befindet sich im Anhang. Somit befestigt Wilk's  $\Lambda$  die These, dass die Modellanpassung im Sommer schlechter als im Winter ist. Zusätzlich deutet dieses Gütekriterium auf eine schlechtere Modellgüte während der Mittagsstunden hin.

#### 4.4 Modellgüte mittels $rt$ und $rd$

Um die Modellanpassung anhand weiterer multivariater Gütemaße zu beurteilen, werden die beiden Gütekriterien  $rt$  und  $rd$  im Folgenden über den Messzeitraum für die Stunden getrennt dargestellt. Daraufhin werden sie sowohl miteinander als auch mit dem Gütekriterium Wilk's  $\Lambda$  verglichen. Dabei ist zu beachten, dass bei den Kriterien  $rt$  und  $rd$  große Werte für eine gute Güte sprechen, anders als bei Wilk's  $\Lambda$ . Folgende Abbildungen 7 und 8 stellen  $rt$  und  $rd$  für die Messzeitpunkte 00:30 Uhr, 02:30 Uhr, ... , 22:30 Uhr über den gesamten Messzeitraum getrennt dar, welcher auf der  $x$  - Achse abgebildet ist. Auf der  $y$  - Achse befindet sich die Güte.

# Gütekriterium $r_t$

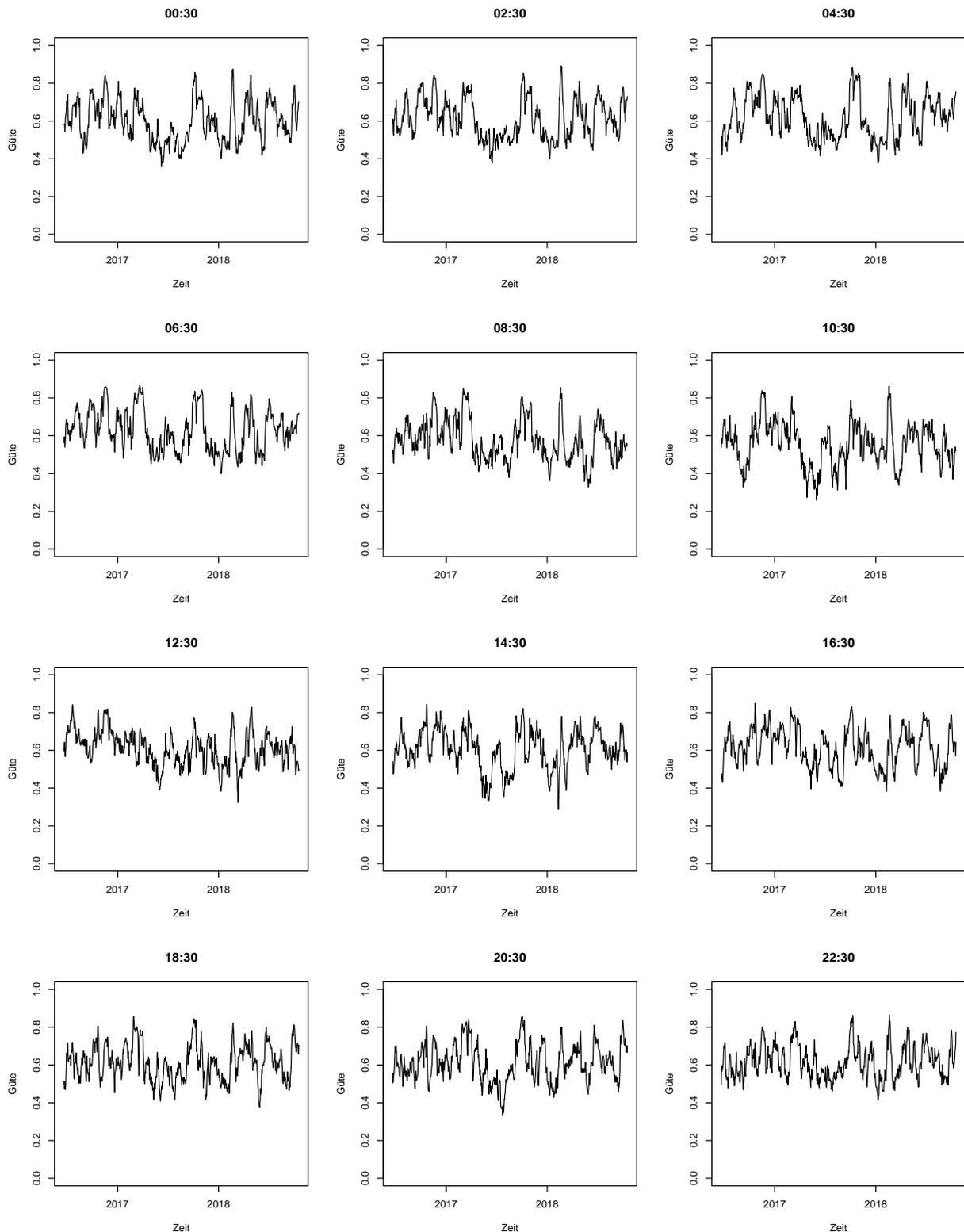


Abbildung 7: Gütekriterium  $r_t$  zur Ermittlung der Modellanpassung eines Multivariaten Linearen Modells zur Vorhersage von Rissen im Baumaterial einer Brücke

# Gütekriterium rd

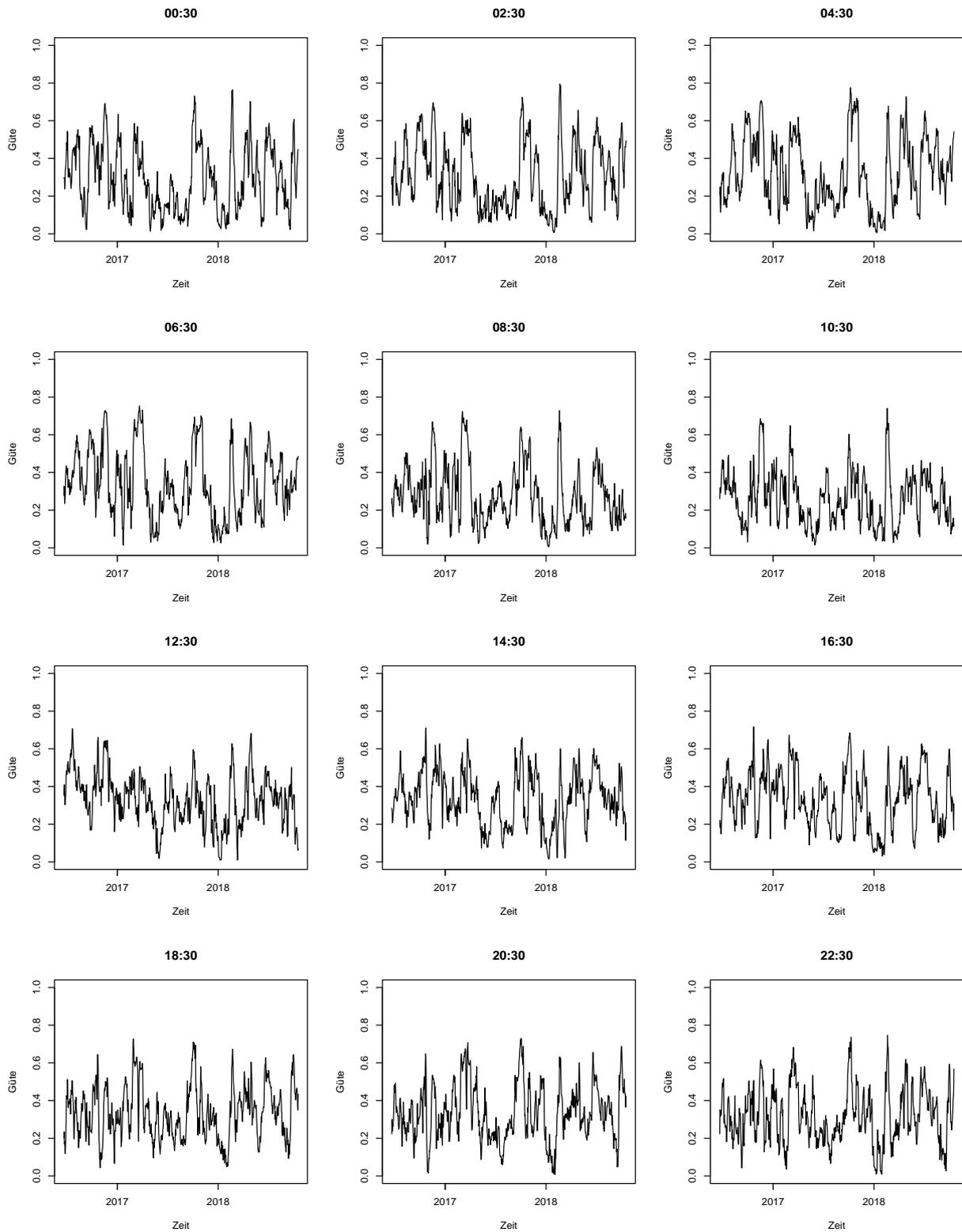


Abbildung 8: Gütekriterium rd zur Ermittlung der Modellanpassung eines Multivariaten Linearen Modells zur Vorhersage von Rissen im Baumaterial einer Brücke

Bei Betrachtung der Abbildung 7 ist zu sehen, dass das Gütekriterium  $rt$  niedrigere Werte in den Sommermonaten als in den restlichen Monaten annimmt. Dies bestätigt die These der schlechten Modellgüte in diesem Zeitraum. In den anderen Jahreszeiten sind die Schwankungen des Kriteriums etwas stärker, jedoch werden hier höhere Werte als im Sommer angenommen, was für eine gute Anpassung des Modells spricht. Gleiches ist bei Betrachtung des Gütekriteriums  $rd$  in der Abbildung 8 festzustellen. Auch hier nimmt das Kriterium im Sommer niedrigere Werte an. Beim Vergleich der beiden Kriterien  $rt$  und  $rd$  fällt jedoch auf, dass  $rd$  im Allgemeinen niedrigere Werte als  $rt$  annimmt und somit einen schlechteren Güte des Modells angibt. Dies lässt sich dadurch erklären, dass  $rd$  den wahren Zusammenhang zwischen der Zielvariable der Rissbreiten und den Kovariablen unterschätzt.

Anders als bei Wilk's  $\Lambda$  ist bei den Gütekriterien  $rt$  und  $rd$  kein Unterschied zwischen den Tageszeiten festzustellen. Somit kann hier die These nicht verfestigt werden, dass die Modellgüte zu den Mittagsstunden schlechter ist. Auch bei den Messzeitpunkten 01:30 Uhr, 03:30 Uhr, ... , 23:30 Uhr lässt sich ein gleiches Schema, sowohl bezüglich der Jahreszeiten als auch bezüglich der Tageszeiten, erkennen. Die Abbildungen A37 und A38 der Graphiken dieser Messzeitpunkte befinden sich im Anhang.

## 4.5 Anpassung der robusten Multivariaten Linearen Modelle

Um festzustellen, ob extreme Messwerte die Modellanpassung beeinflussen und die Modellgüte verschlechtern, werden die Koeffizienten im Folgenden für die Multivariaten Linearen Modelle robust geschätzt. Für die robuste Schätzung wird die LTS-Schätzung verwendet.

Durch die deskriptive Analyse der Rissbreiten zu Beginn der Auswertung wurde bereits deutlich, dass vor allem in den Sommermonaten extreme Werte gemessen wurden. Die nicht robusten Gütekriterien haben bereits gezeigt, dass die Modellanpassung des nicht robusten Multivariaten Linearen Modells im Sommer schlechter als im Winter ist. Ob ein robustes Modell, welches die extremen Werte nicht berücksichtigt, dies ändert, wird im Folgenden herausgearbeitet. Dazu werden zunächst die robusten Modelle für die verschiedenen Stunden getrennt über Zeitfenster von 28 Tagen angepasst. Die resultierenden Koeffizientenschätzungen für den Interzept und die Kovariablen werden über den gesamten Messzeitraum graphisch dargestellt. Folgende Abbildung 9 zeigt die robusten Koeffizientenschätzungen für den Messzeitpunkt 06:30 Uhr.

# 06:30 Uhr

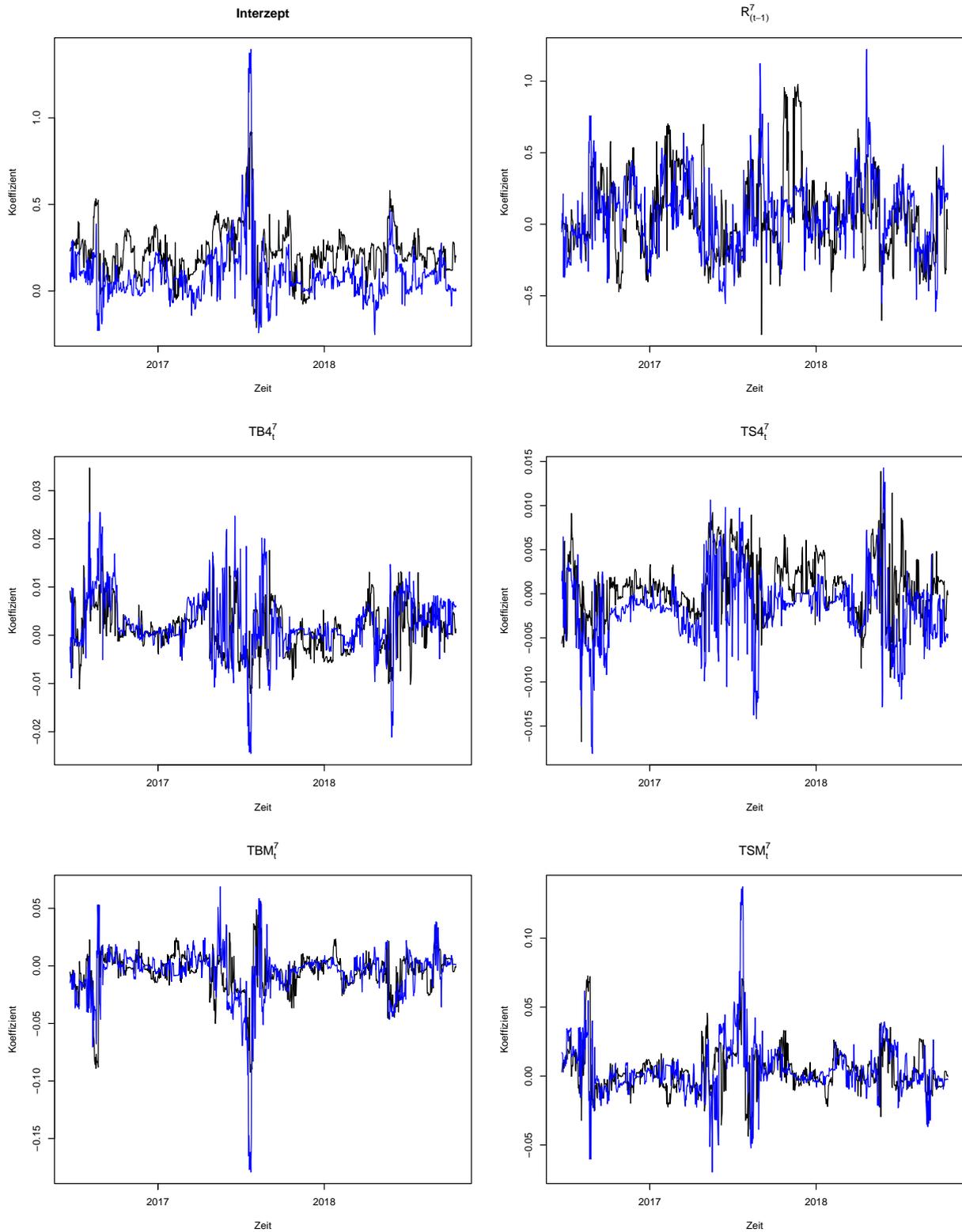


Abbildung 9: robuste Koeffizientenschätzungen des Multivariaten Linearen Modells für den Messzeitpunkt 06:30 Uhr über den gesamten Messzeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

Zunächst fällt auf, dass die Koeffizienten für Riss 1 und Riss 2 über den Messzeitraum hinweg sichtbar ähnlich zueinander sind. Wie auch bei den Koeffizientenschätzungen des nicht robusten Multivariaten Linearen Modells sind hierbei ebenfalls Unterschiede zwischen den Sommer - und Wintermonaten feststellbar. Im Sommer schwanken die Koeffizienten, vor allem für die Variablen  $TB4_{(t)}^s$ ,  $TS4_{(t)}^s$ ,  $TBM_{(t)}^s$  und  $TSM_{(t)}^s$ , deutlich mehr als im Winter. Darüber hinaus weisen die Temperatur angehenden Variablen einige extreme Werte der Koeffizienten im Sommer auf. Über den Winter hingegen schwanken die Koeffizienten kaum. Somit lässt sich auch hier eine schlechtere Modellgüte für die Sommermonate vermuten.

Die Größenordnungen der robusten Koeffizientenschätzungen sind zu den nicht robusten Schätzungen ähnlich. Dies gilt auch für den ähnlichen Verlauf der Koeffizienten für Riss 1 und Riss 2.

Da es auch hierbei keine großen Unterschiede zwischen den Schätzungen für die verschiedenen Stunden gibt, befinden sich die Graphiken A39,..., A61 mit den Koeffizienten über den Messzeitraum für die weiteren Messzeitpunkte im Anhang.

## 4.6 Modellgüte basierend auf dem MVE-Schätzer

Um die Modellanpassung des Multivariaten Linearen Modells mit robusten Koeffizientenschätzungen herauszuarbeiten, wird die Modellgüte anhand eines robusten Gütekriteriums, basierend auf der MVE-Schätzung, dargestellt. Dazu wird die Modellgüte für jede Stunde getrennt für die Zeitfenster über 28 Tage bestimmt. Je größere Werte das Gütekriterium annimmt, desto schlechter ist die Anpassung des Modells. Die folgende Abbildung 10 stellt dies für die Messzeitpunkte 00:30 Uhr, 02:30 Uhr, ... , 22:30 Uhr dar. Auf der  $x$  - Achse wird dabei der gesamte Messzeitraum abgebildet und auf der  $y$  - Achse befindet sich die Güte.

# robustes Gütekriterium basierend auf der MVE-Schätzung

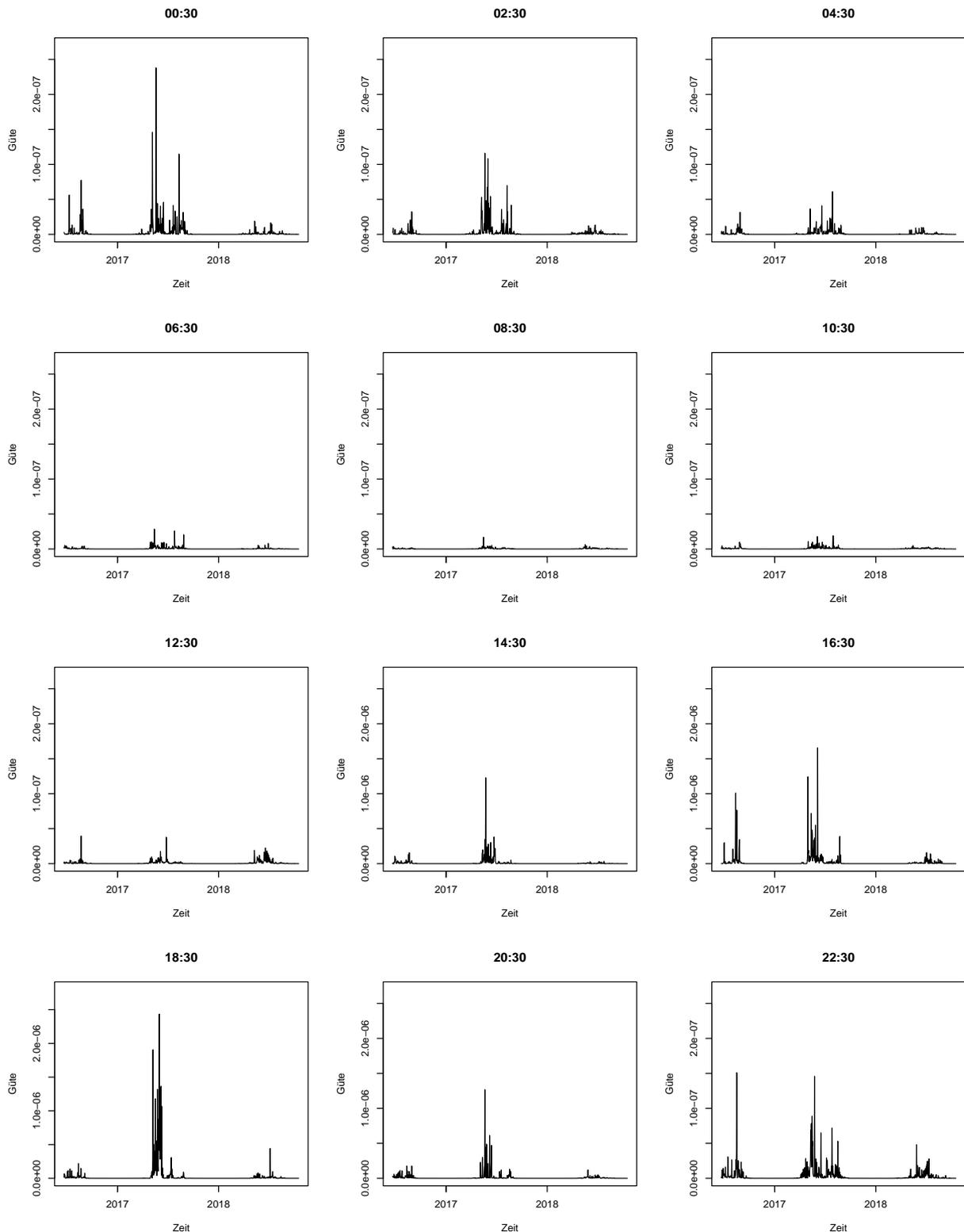


Abbildung 10: Gütekriterium basierend auf der MVE-Schätzung zur Ermittlung der Modellanpassung eines robusten Multivariaten Linearen Modells zur Vorhersage von Rissen im Baumaterial einer Brücke

Zunächst ist zu beachten, dass die  $x$  - Achsen Skalierung bei den Messzeitpunkten 14:30 Uhr, 16:30 Uhr, 18:30 Uhr und 20:30 Uhr um den Faktor 10 größer ist als bei den restlichen Messzeitpunkten. Somit ist zu erkennen, dass zu diesen Stunden extremere Werte als zu den restlichen dargestellten Stunden vorliegen. Dies impliziert eine schlechtere Modellgüte zu diesen Tageszeiten. Ähnlich wie bei der Modellgüte des Modells mit nicht robusten Schätzungen liegen die Stunden mit der schlechteren Anpassung im Mittags - bis frühen Abendbereich.

Eine schlechte Modellgüte wird zudem in den Sommermonaten erreicht, besonders im Sommer 2017. Hier ist die Anpassung des Modells schlechter als in den restlichen Monaten, da hier die Temperaturschwankungen zwischen den Tagen stärker sind. Somit kann nur ein geringer Anteil der Varianz der Rissbreiten durch das Modell erklärt werden. Gleiches war bereits bei der Güte des nicht robusten Modells zu erkennen. Im Gegensatz dazu fällt hier jedoch auf, dass die Unterschiede zwischen Sommer und Winter noch extremer sind und im Winter hierbei eine sehr gute Modellanpassung vorliegt. Der Sommer 2017 fällt hierbei besonders stark durch die hohen Werte des Gütekriteriums auf. Die Werte in diesem Zeitraum sind deutlich extremer als die Werte in den Sommermonaten der anderen beiden Jahre, sodass die Modellanpassung im Sommer 2017 auffällig schlecht ist.

Bei den restlichen Stunden ist ein ähnliches Schema zu beobachten. Die Abbildung A62 für diese Messzeitpunkte befindet sich im Anhang.

## 4.7 Modellgüte basierend auf dem MCD-Schätzer

Zusätzlich dazu wird die Modellgüte des Multivariaten Linearen Modells mit robusten Koeffizientenschätzungen basierend auf der MCD-Schätzung untersucht. Auch hierfür wird die Modellgüte für jede Stunde getrennt für die Zeitfenster über 28 Tage bestimmt und über den gesamten Messzeitraum geplottet. Hierbei gilt ebenfalls, dass größere Werte für eine schlechtere Güte sprechen. Die folgende Abbildung 11 stellt dies für die Messzeitpunkte 00:30 Uhr, 02:20 Uhr, ... , 22:20 Uhr dar. Auf der  $x$  - Achse wird dabei der gesamte Messzeitraum abgebildet und auf der  $y$  - Achse befindet sich die Güte.

# robustes Gütekriterium basierend auf der MCD-Schätzung

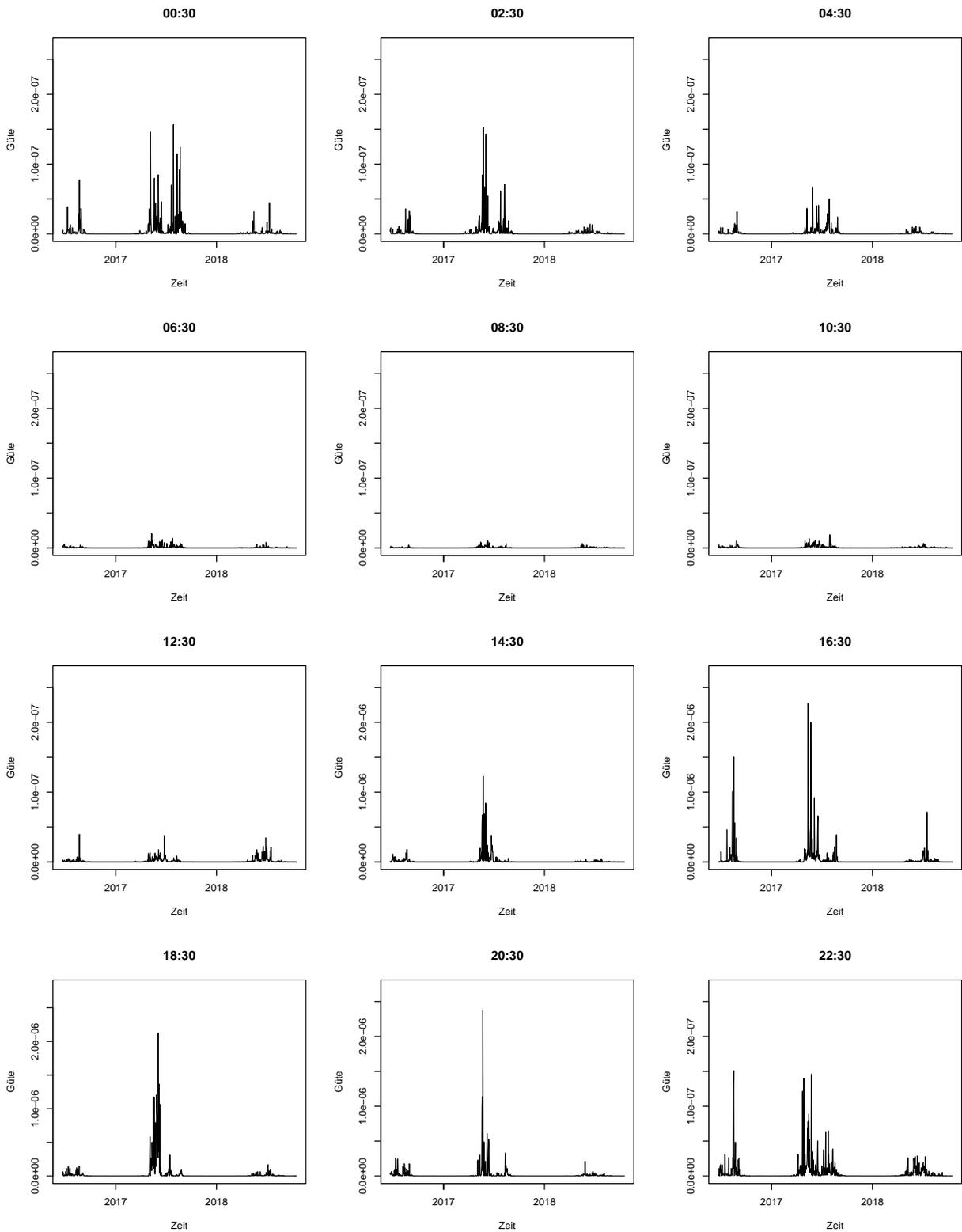


Abbildung 11: Gütekriterium basierend auf der MCD-Schätzung zur Ermittlung der Modellanpassung eines robusten Multivariaten Linearen Modells zur Vorhersage von Rissen im Baumaterial einer Brücke

Ebenfalls ist hier bei der Abbildung zu beachten, dass die Skalierung der  $x$ -Achse bei den Messzeitpunkten 14:30 Uhr, 16:30 Uhr, 18:30 Uhr und 20:30 Uhr um den Faktor 10 größer ist als bei den restlichen Messzeitpunkten. Folglich liegen bei diesen Zeitpunkten extremere Werte der Güte vor. Dies deutet auf eine schlechtere Modellgüte zu den Mittagsstunden und frühen Abendstunden als zu den restlichen Tageszeiten hin. Des Weiteren ist auffällig, dass das Gütekriterium zu allen dargestellten Stunden deutliche höhere Werte in den Sommermonaten annimmt. Speziell im Sommer 2017 werden besonders hohe Werte angenommen, was für eine auffällig schlechte Modellanpassung in diesem Zeitraum spricht.

Ein ähnliches Schema ist bei den anderen, bislang nicht dargestellten, Stunden zu beobachten. Die Abbildung A63 mit den Werten der Güte für diese Stunden befindet sich im Anhang.

Somit zeigen die beiden robusten Gütekriterien basierend auf der MVE-Schätzung und auf der MCD-Schätzung ähnliche Ergebnisse auf.

## 4.8 Betrachtung des Sommer 2017

Da alle Gütekriterien in den Sommermonaten des Jahres 2017 besonders auffällig waren, indem die Modellgüte in diesem Zeitraum besonders schlecht war, werden die Sommermonate des Jahres 2017 im Folgenden genauer untersucht. Bei der zu Beginn der Auswertung durchgeführten deskriptiven Analyse der Rissbreiten in Abbildung 3 ist bereits aufgefallen, dass die Schwankungen der Risse in den Sommermonaten stärker als in den Wintermonaten sind. Um zu überprüfen, ob es weitere Auffälligkeiten des Verlaufs der Rissbreiten im Sommer 2017 gibt, wird dieser erneut deskriptiv untersucht. Nachstehende Abbildung 12 bildet die Rissbreiten der Risse 1 und 2 in dem Zeitraum vom 20.02.2017 bis zum 10.10.2017 ab. Dabei ist auf der  $x$ -Achse der Messzeitraum und auf der  $y$ -Achse die Rissbreite in Millimetern dargestellt.

WWN2 und WWN3 im Sommer 2017

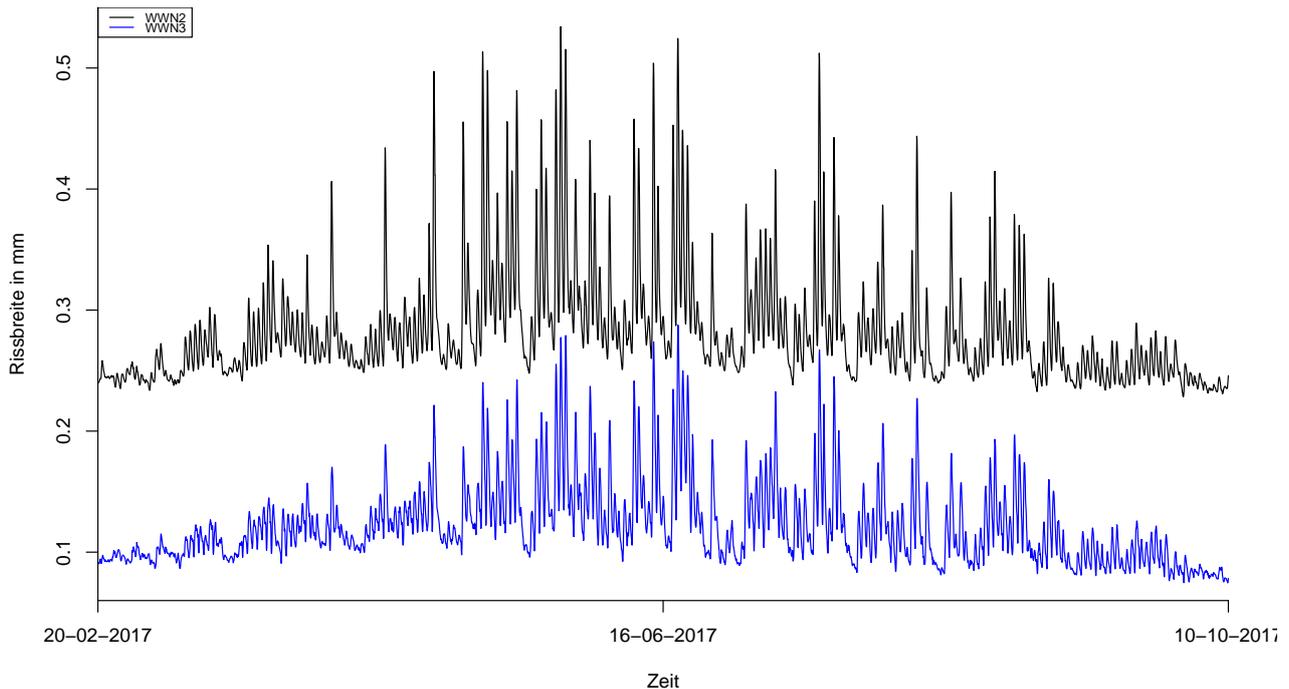


Abbildung 12: Messergebnisse des WWN2 und WWN3 über den Zeitraum vom 20.02.2017 bis zum 10.10.2017

Bei Betrachtung der Abbildung fällt auf, dass die Schwankungen beider Risse sehr stark sind. Insbesondere in den frühen Sommermonaten ist dies der Fall. Für Riss 1 wird die maximale Rissbreite in diesem Zeitraum am 26.05.2017 um 17:30 Uhr angenommen und für Riss 2 wird die maximale Rissbreite am 19.06.2017 um 18:30 Uhr erreicht. Diese Rissbreiten entsprechen gleichzeitig den maximalen Werten für beide Risse im gesamten Messzeitraum.

Neben den starken Schwankungen und großen Rissen im Allgemeinen in diesem Zeitraum gibt es jedoch keine anderen Auffälligkeiten des Verlaufs der Rissbreiten.

Folglich scheint die schlechte Modellgüte in den Sommermonaten, insbesondere im Sommer 2017, lediglich Folge der starken Schwankungen der Rissbreiten zu sein. Hinzu kommt der Einfluss der extremen Werte in diesem Zeitraum, wobei der Einfluss dieser durch die robusten Gütemaße verringert wird.

## 5 Zusammenfassung

Ziel der Arbeit war es, die Güte des Multivariaten Linearen Modells von Messdaten eines Monitorings einer Brücke in Bochum zu untersuchen. Durch den Einfluss von verschiedenen Faktoren, wie unter anderem der Sonneneinstrahlung und der Temperatur, schwanken die Rissbreiten un-

terschiedlich stark über den Messzeitraum hinweg. Bei der Analyse wurden die beiden größten Risse betrachtet. Bei der deskriptiven Betrachtung der Rissbreiten dieser beiden Risse ließen sich bereits Auffälligkeiten in bestimmten Zeiträumen erkennen. In den Sommermonaten waren sowohl besonders große als auch stark schwankende Rissbreiten sichtbar.

Ähnliche Auffälligkeiten ließen sich bei der deskriptiven Analyse der Temperatur oberhalb der Brücke auf der Sonnenseite und unterhalb der Brücke auf der Schattenseite feststellen. Zusätzlich zu der höheren Temperatur im Sommer sind ebenfalls die Temperaturschwankungen in diesem Zeitraum stärker und häufiger als in den Wintermonaten. Dies trifft vor allem auf die Sonnenseite oberhalb der Brücke zu.

Bei der Anpassung des Multivariaten Linearen Modells an die Messdaten wurden zunächst die Koeffizientenschätzungen des Modells betrachtet. Hierbei hat sich herausgestellt, dass vor allem die Rissbreite der Risse am Vortag einen starken Einfluss auf die Rissbreite am nachfolgenden Tag hat. Generell wurden auch hier stärkere Schwankungen der Koeffizientenschätzungen des Modells in den Sommermonaten festgestellt, sodass sich bereits eine schlechtere Modellgüte in diesem Zeitraum vermuten ließ.

Bei der Untersuchung der Güte mittels der Gütekriterien Wilk's- $\Lambda$ ,  $r_t$  und  $r_d$  hat sich diese These bestätigt. Die Modellgüte ist in den Sommermonaten schlechter als in den übrigen Monaten des Jahres. Insbesondere für den Sommer 2017 ist die Güte auffällig schlecht. Ebenfalls wurde herausgearbeitet, dass die Modellgüte zu den Mittagsstunden schlechter als zu den restlichen Tageszeiten ist. Jedoch sind die Schwankungen der Gütekriterien im Allgemeinen stark, sodass sich teilweise auf kein eindeutiges Ergebnis schließen lässt.

Da die insbesondere in den Sommermonaten vorkommenden extremen Messwerte die Analyse verzerren können, wurde das Multivariate Lineare Modell zusätzlich mittels robuster Schätzungen für die Koeffizienten aufgestellt. Bei der Analyse dieser Schätzungen hat sich herausgestellt, dass die Schwankungen der robusten Koeffizientenschätzungen ebenfalls in den Sommermonaten extremer als in dem übrigen Zeitraum sind.

Daraufhin wurde die Modellgüte anhand von robusten Schätzern herausgearbeitet. Hierbei wurden die robusten Gütemaße basierend auf der MVE-Schätzung und auf der MCD-Schätzung betrachtet. Anders als die nicht robusten Gütemaße schwanken die robusten Gütemaße insgesamt weniger stark über den Messzeitraum hinweg. Insbesondere in den Herbst-, Winter- und Frühlingsmonaten sprechen die Gütekriterien für eine sehr gute Anpassung des Modells. Auffällig ist hierbei jedoch erneut die schlechte Güte in den Sommermonaten, insbesondere im Sommer 2017. Die robusten Gütemaße zeigen den Unterschied der Güte zwischen Sommer und den restlichen Jahreszeiten deutlicher auf, wohingegen bei den nicht robusten Gütekriterien nur ein geringerer Unterschied der

Güte erkennbar war. Ebenfalls wurde bei den robusten Gütekriterien festgestellt, dass die Modelanpassung zu den Mittagsstunden schlechter als zu den restlichen Tageszeiten ist.

Zusammenfassend lässt sich sagen, dass die Modellgüte insbesondere in den Sommermonaten, speziell im Sommer 2017, aber auch zu den Mittagsstunden schlechter als zu den restlichen Zeiten ist.

Die robusten Gütekriterien zeigen dies besonders stark.

## Literaturverzeichnis

Christensen, R. (2019). *Advanced Linear Modeling - Statistical Learning and Dependent Data*, Springer Nature, Singapore.

Müller, C. (2022).

**URL:** <https://www.statistik.tu-dortmund.de/mueller1.html>

R Core Team (2021). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

**URL:** <https://www.R-project.org/>

Rencher, A. C. (1998). *Multivariate Statistical Interference and Applications*, John Wiley & Sons, New York.

Rencher, A. C. und Christensen, W. F. (2012). *Methods of Multivariate Analysis -*, John Wiley & Sons, New York.

Rousseeuw, P. (1985). *Multivariate estimation with high breakdown point*, Springer Netherlands, Dordrecht: Reidel.

Schürmann, D. (2018). Rissentwicklung im Spannbeton einer Brücke unter Temperatureinfluss.

**URL:** <https://www.statistik.tu-dortmund.de/1255.html>

Thunich, O. (2017). Eliminierung des Temperatureffekts bei Brückenmonitoringdaten.

**URL:** <https://www.statistik.tu-dortmund.de/1255.html>

# Anhang

# Koeffizienten der Stunde 00:30 Uhr

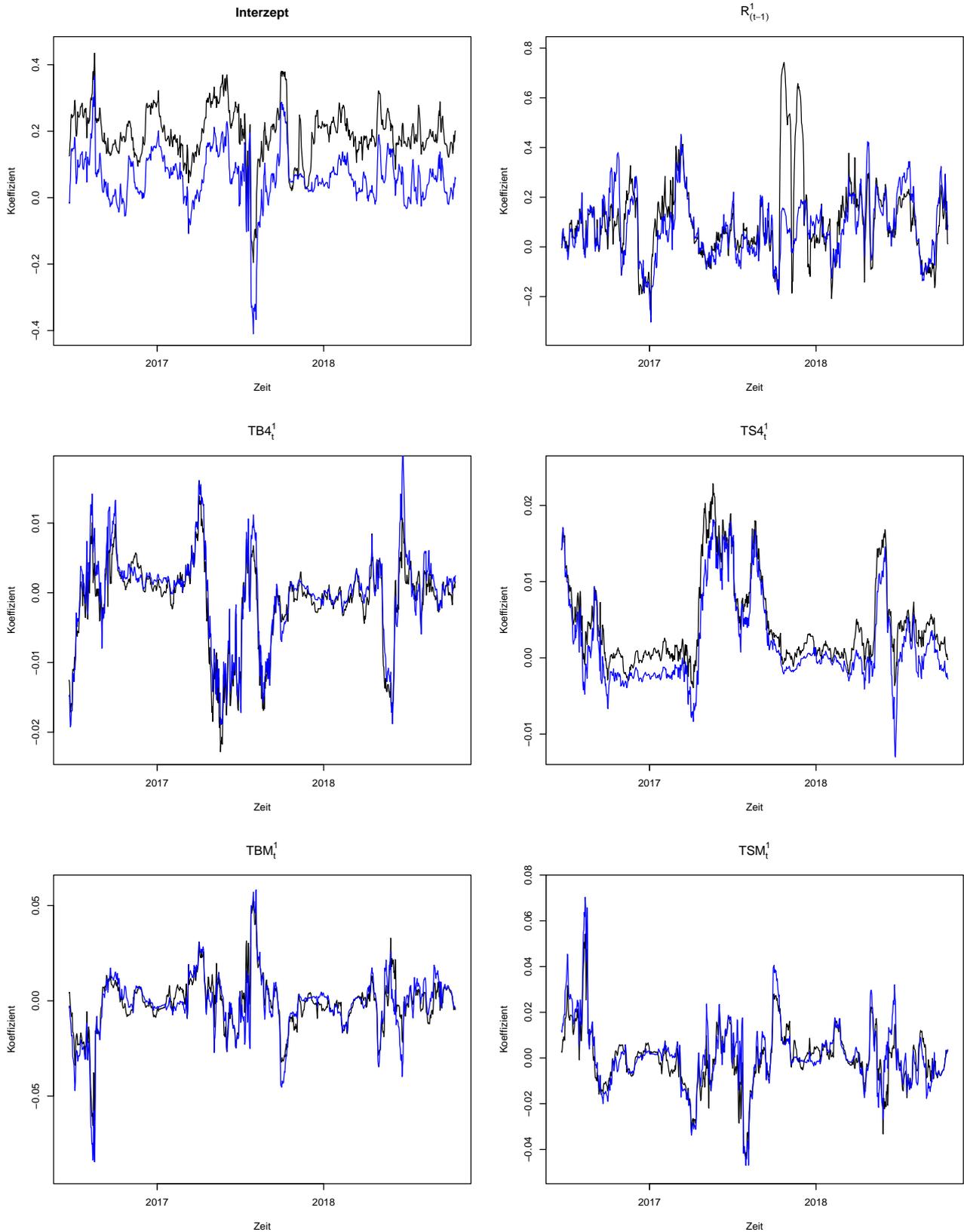


Abbildung 13: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 00:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 01:30 Uhr

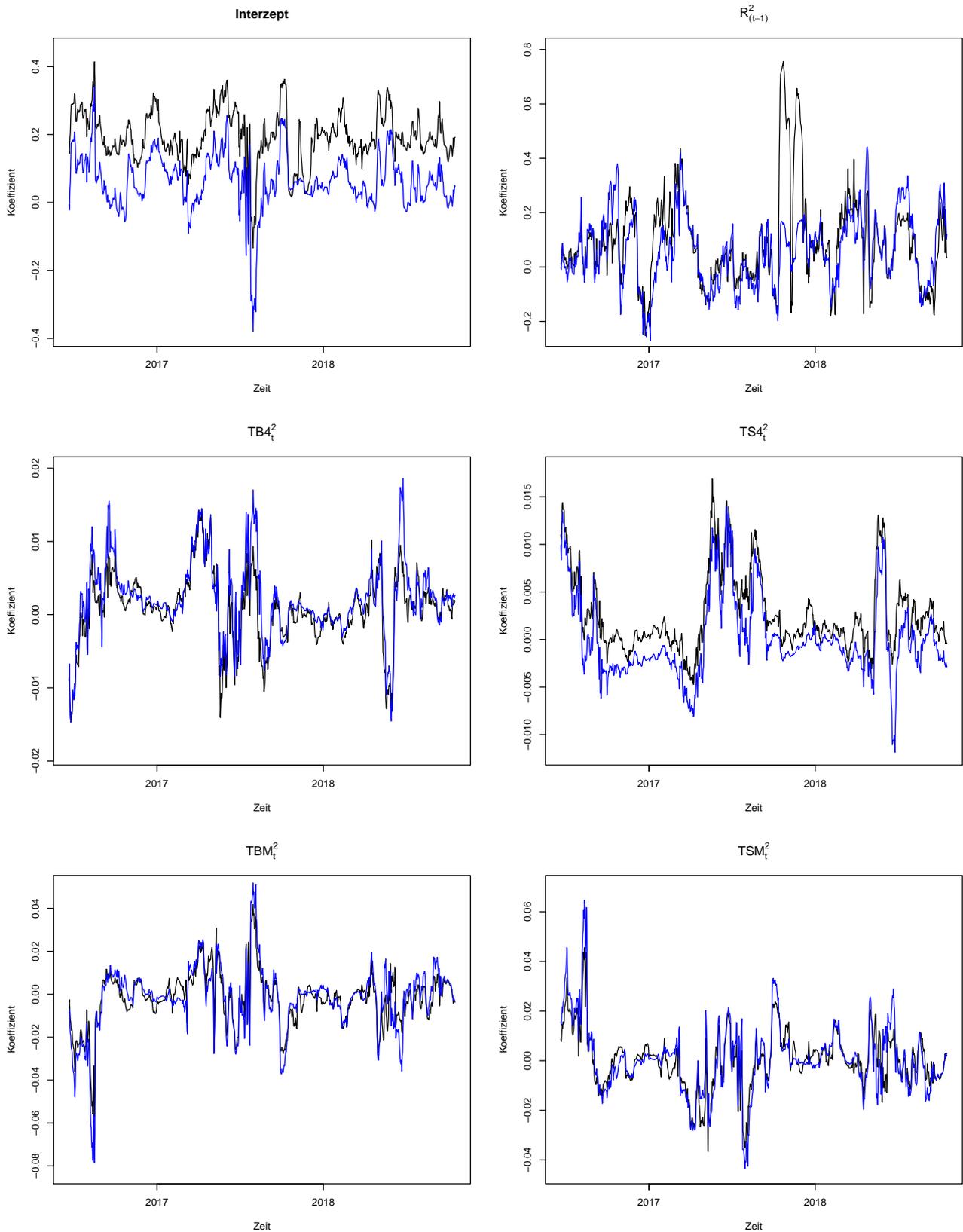


Abbildung 14: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 01:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 02:30 Uhr

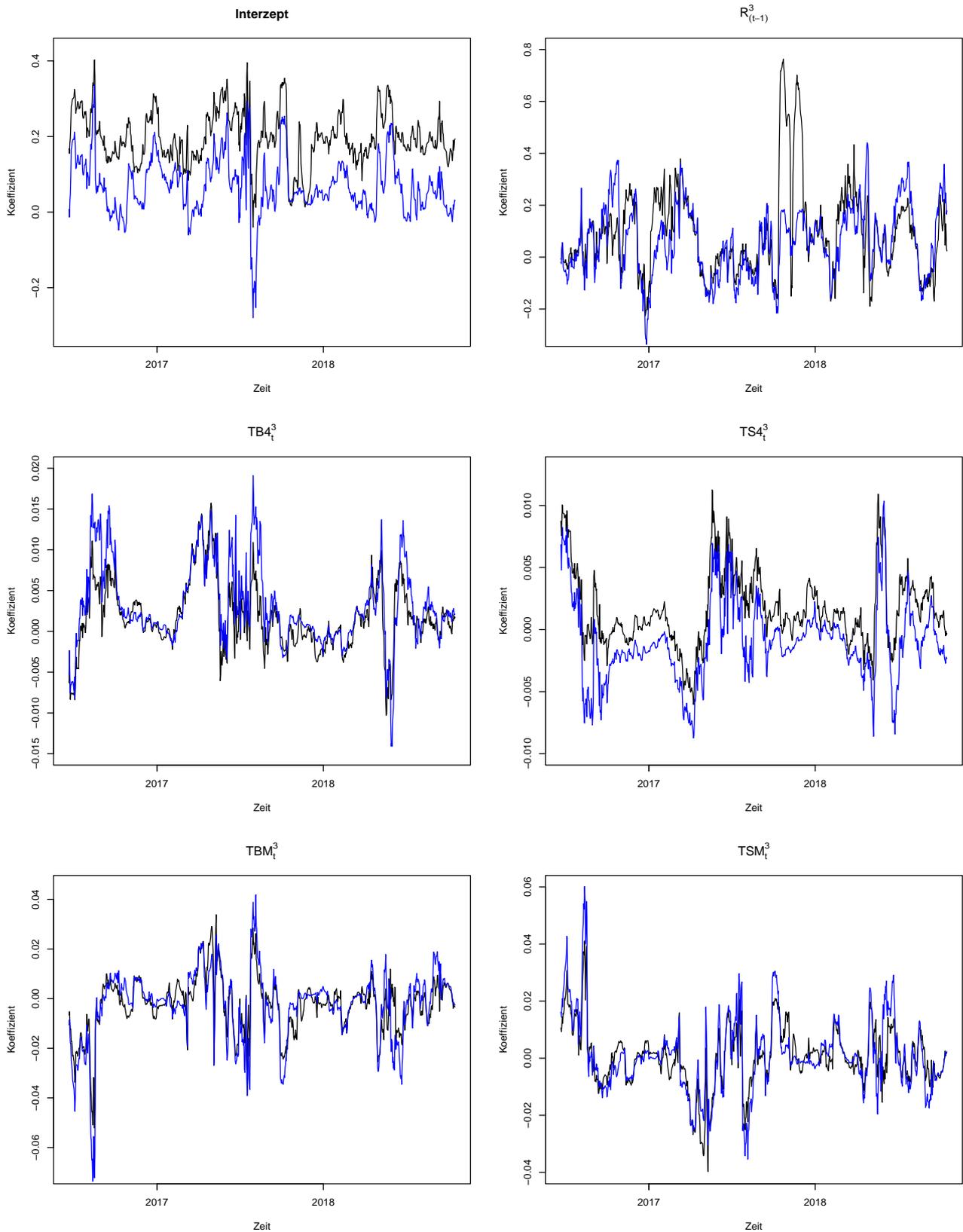


Abbildung 15: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 02:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 03:30 Uhr

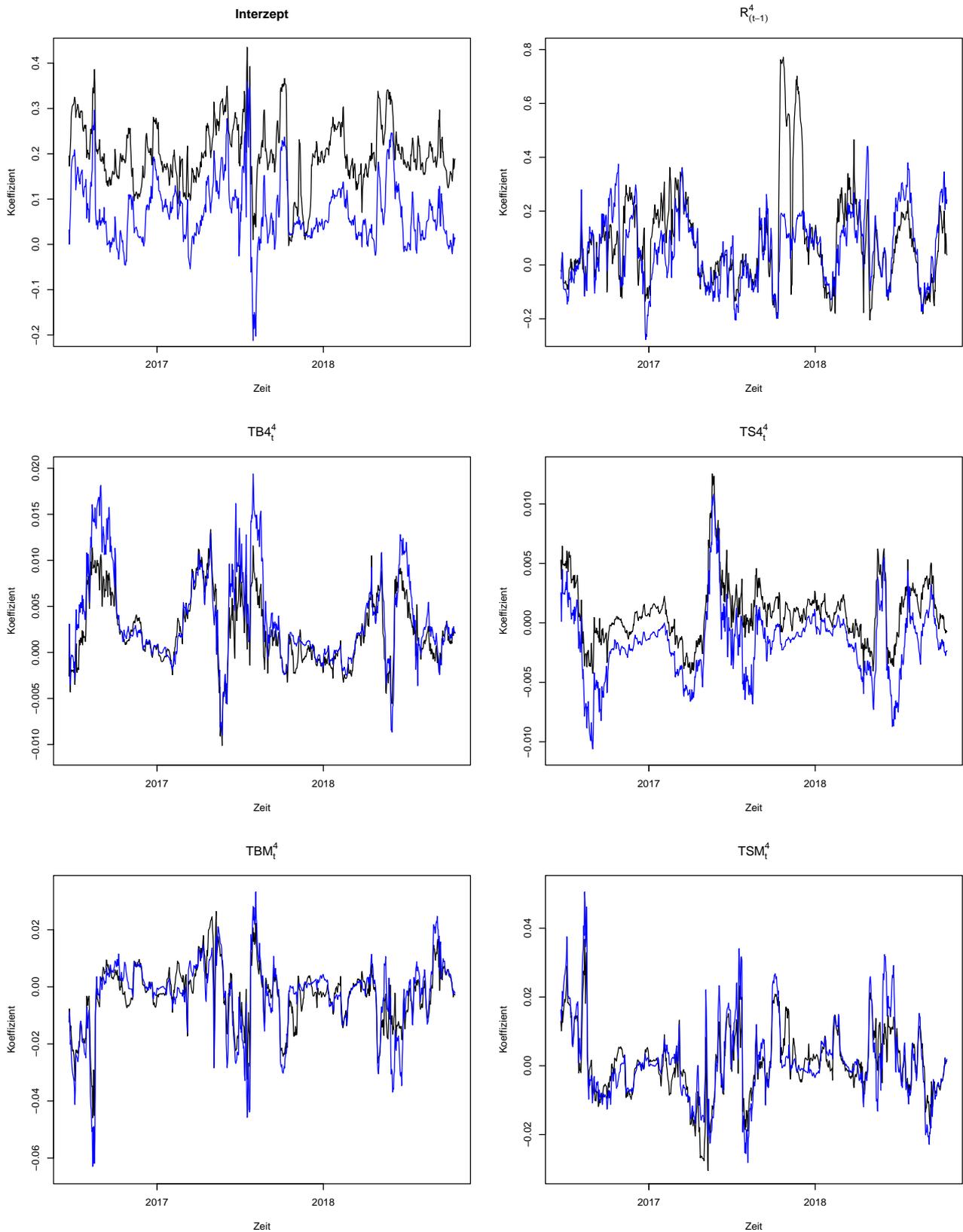


Abbildung 16: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 03:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 04:30 Uhr

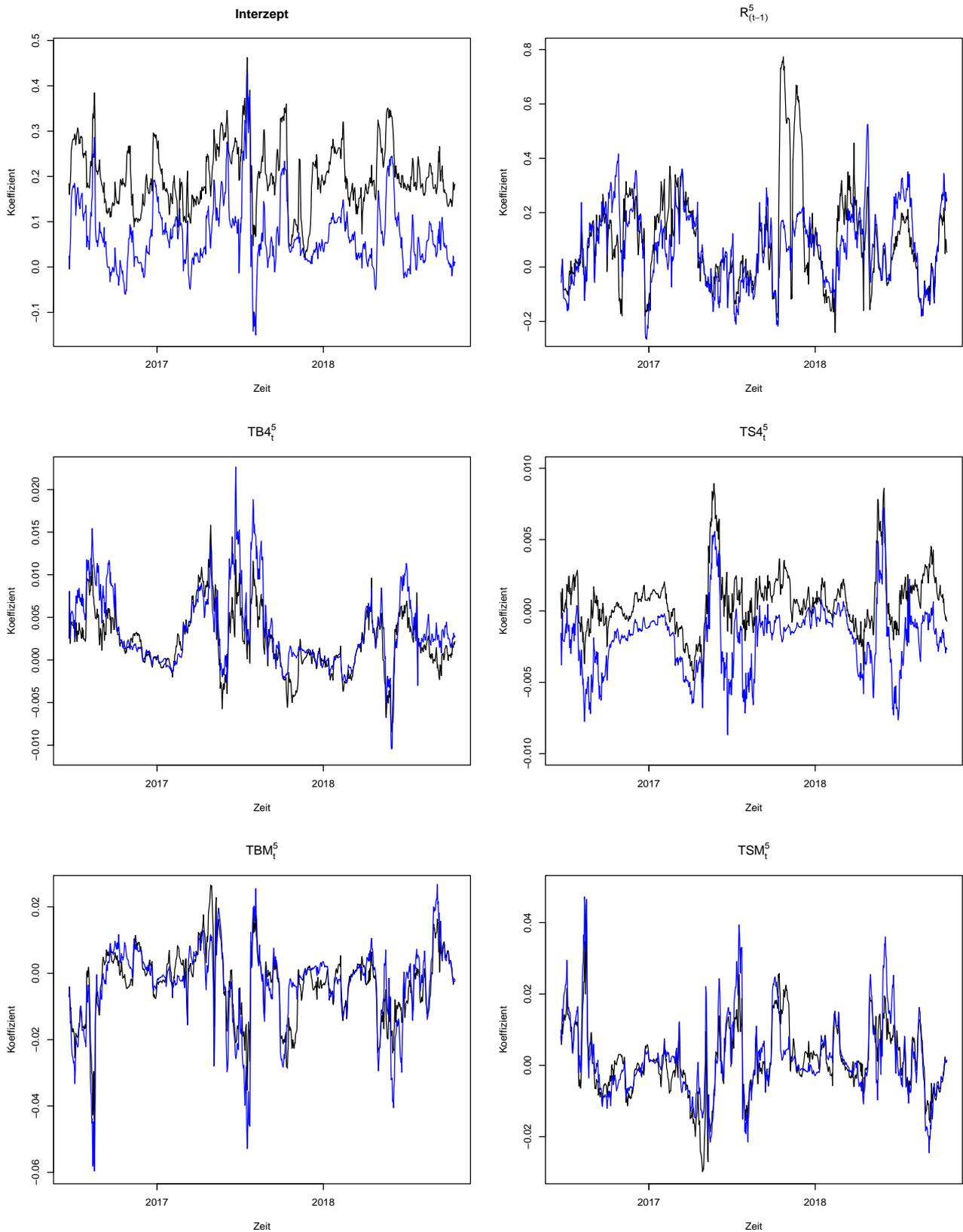


Abbildung 17: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 04:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 05:30 Uhr

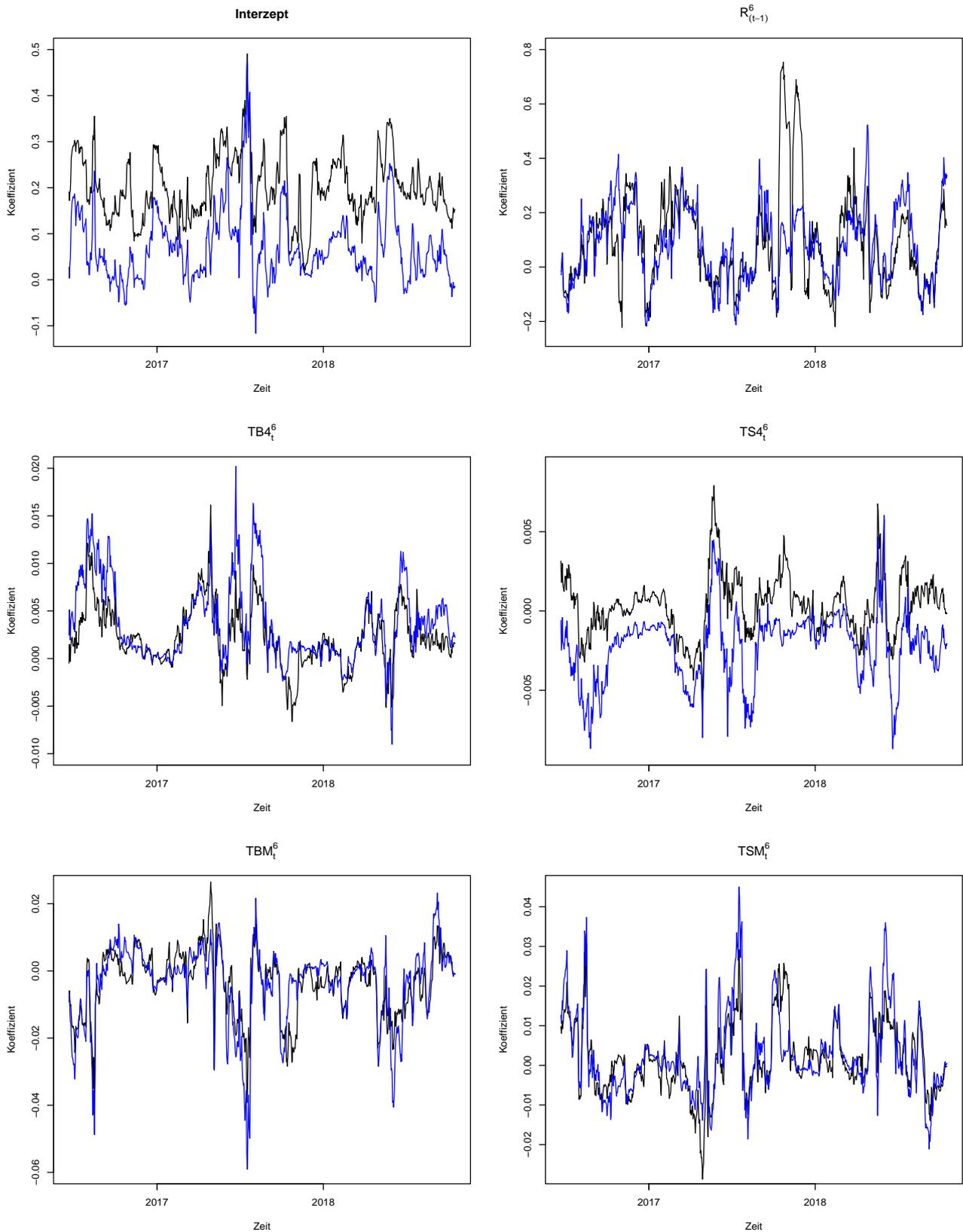


Abbildung 18: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 05:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 07:30 Uhr

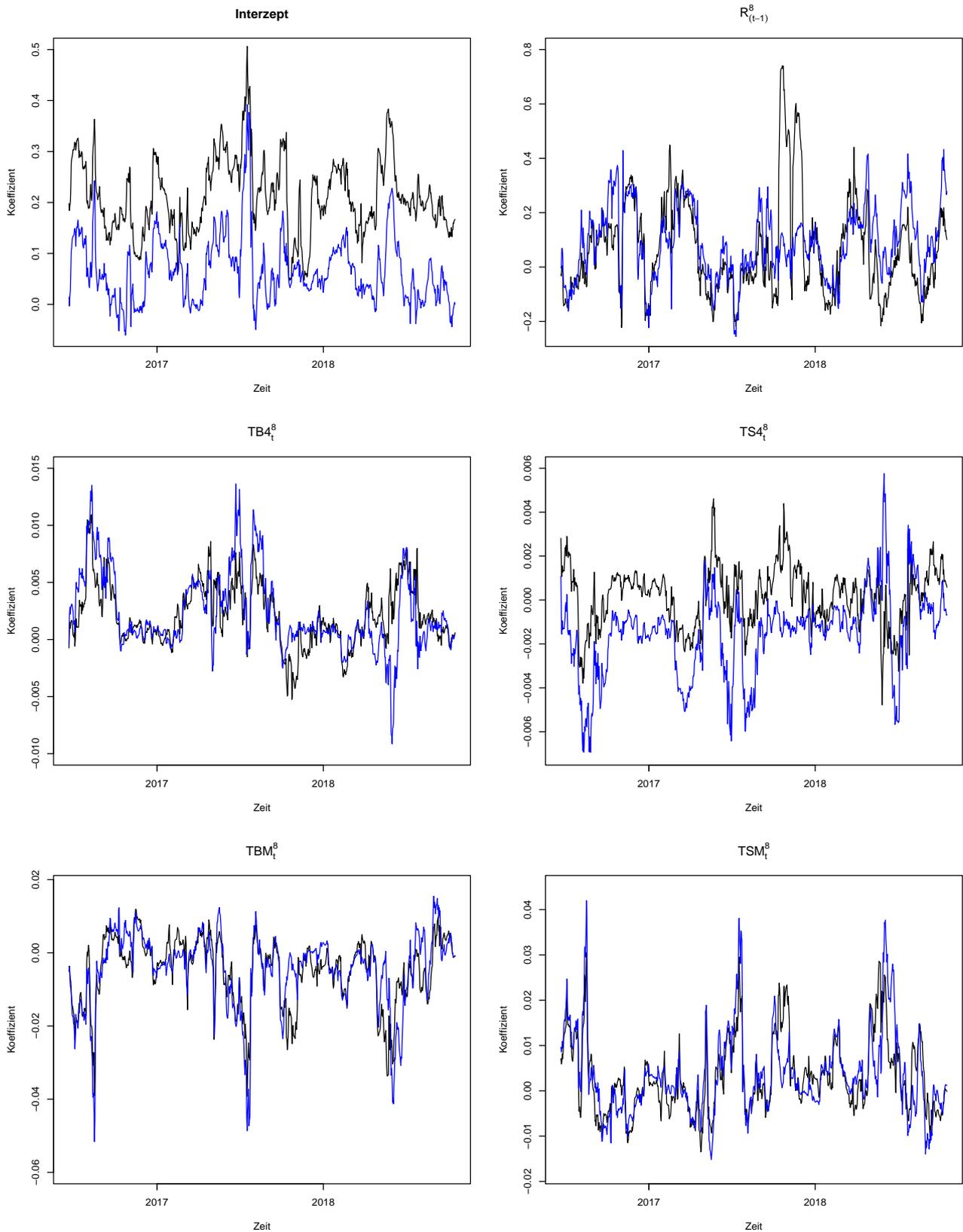


Abbildung 19: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 07:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 08:30 Uhr

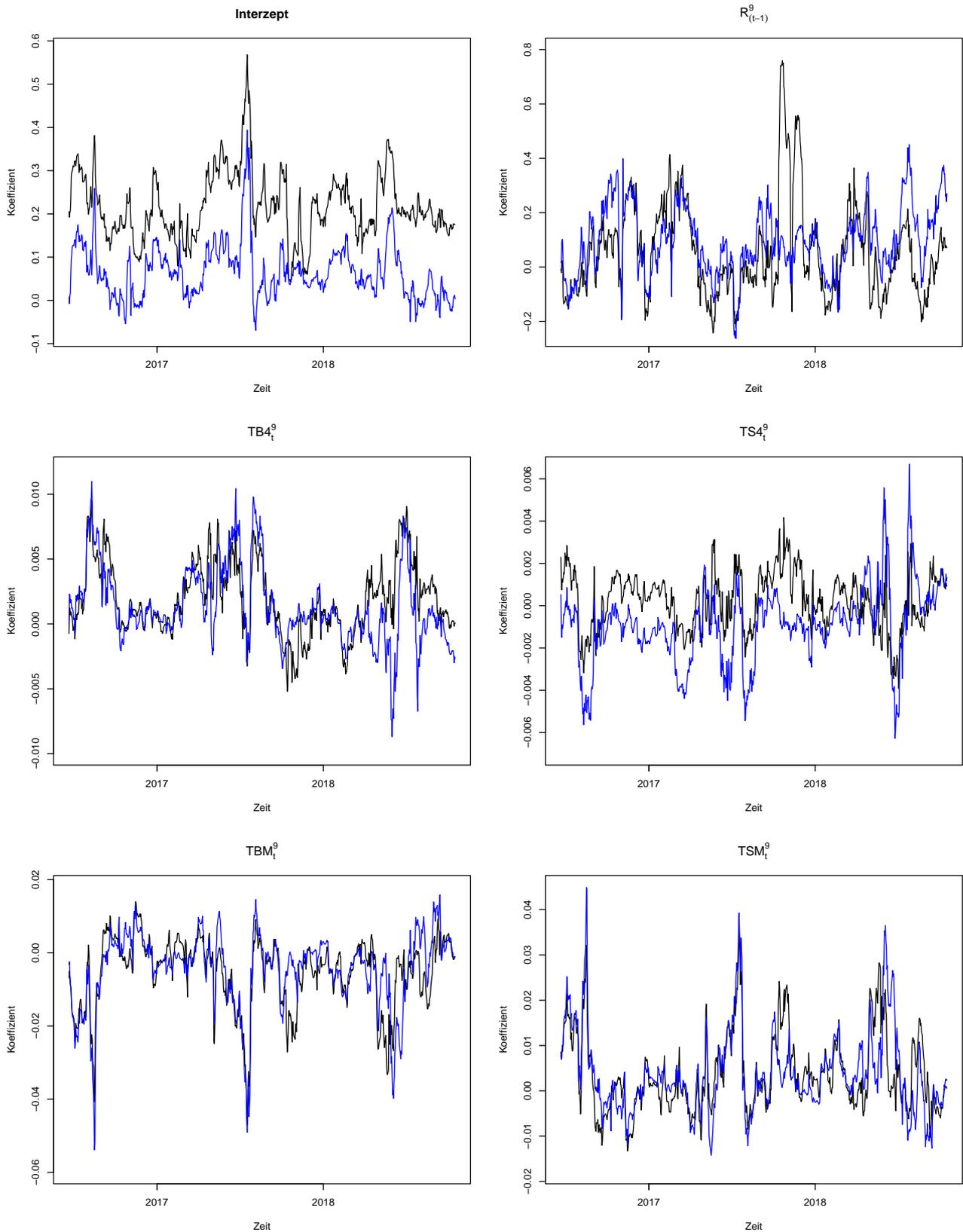


Abbildung 20: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 08:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 09:30 Uhr

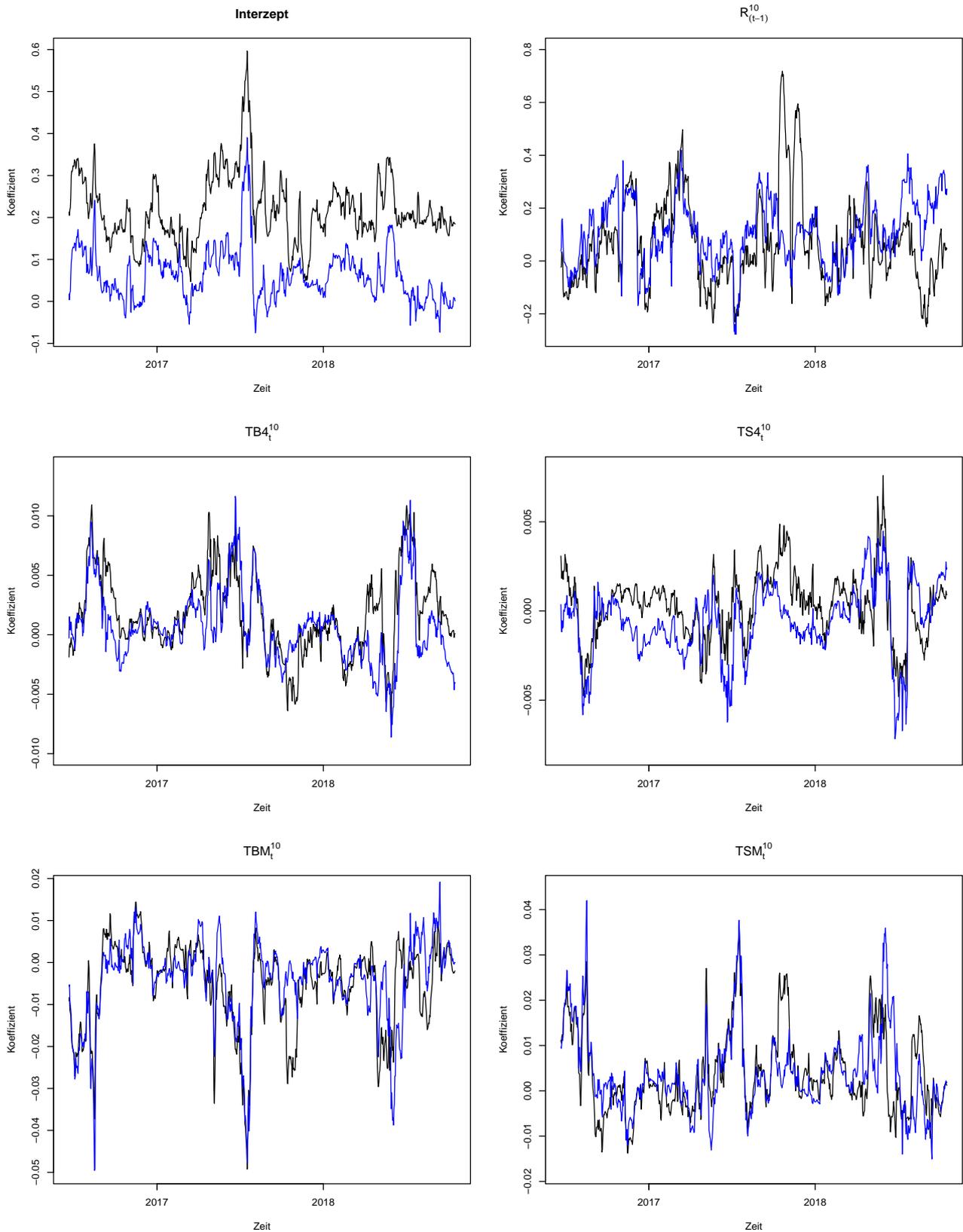


Abbildung 21: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 09:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 10:30 Uhr

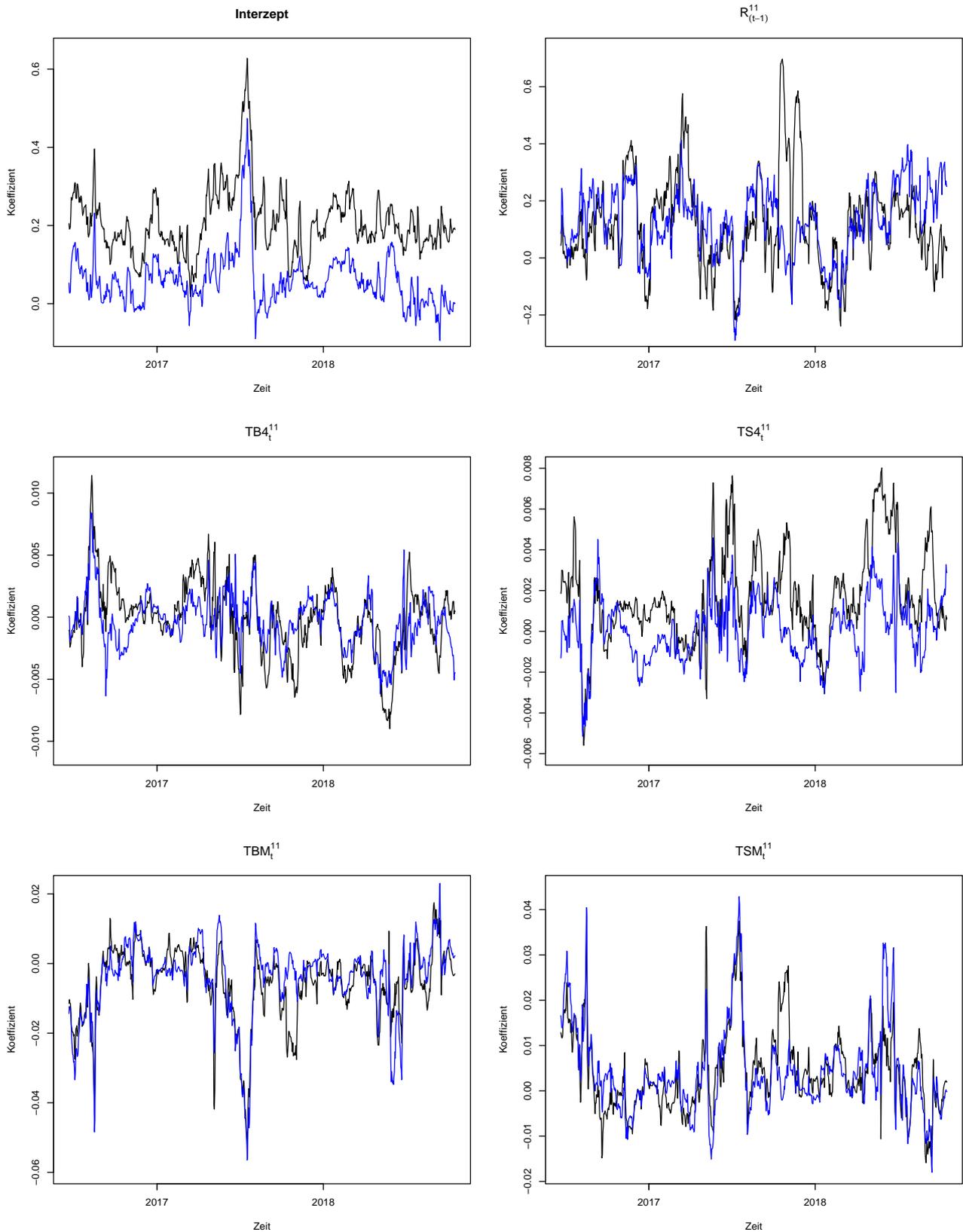


Abbildung 22: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 10:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 11:30 Uhr

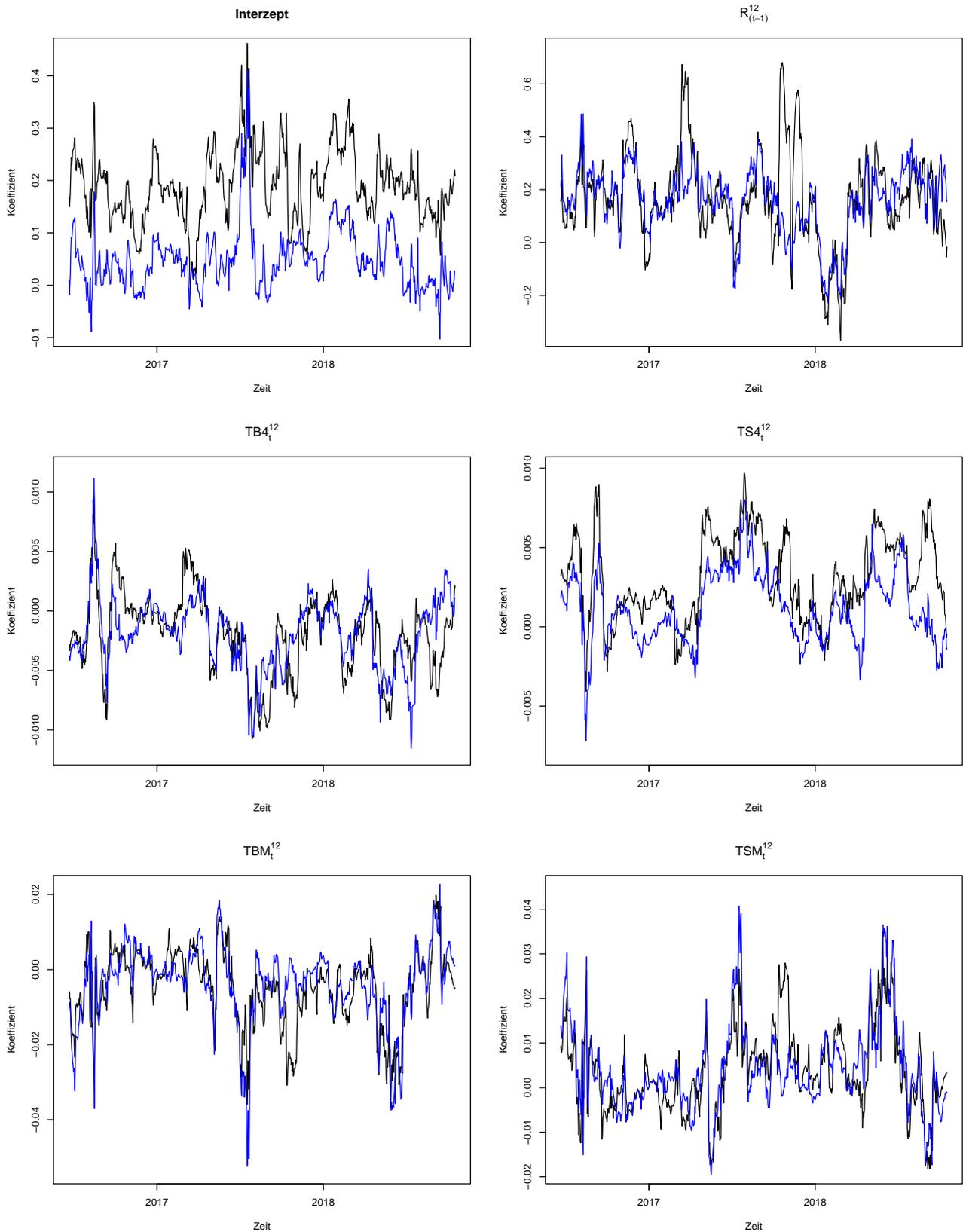


Abbildung 23: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 11:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 12:30 Uhr

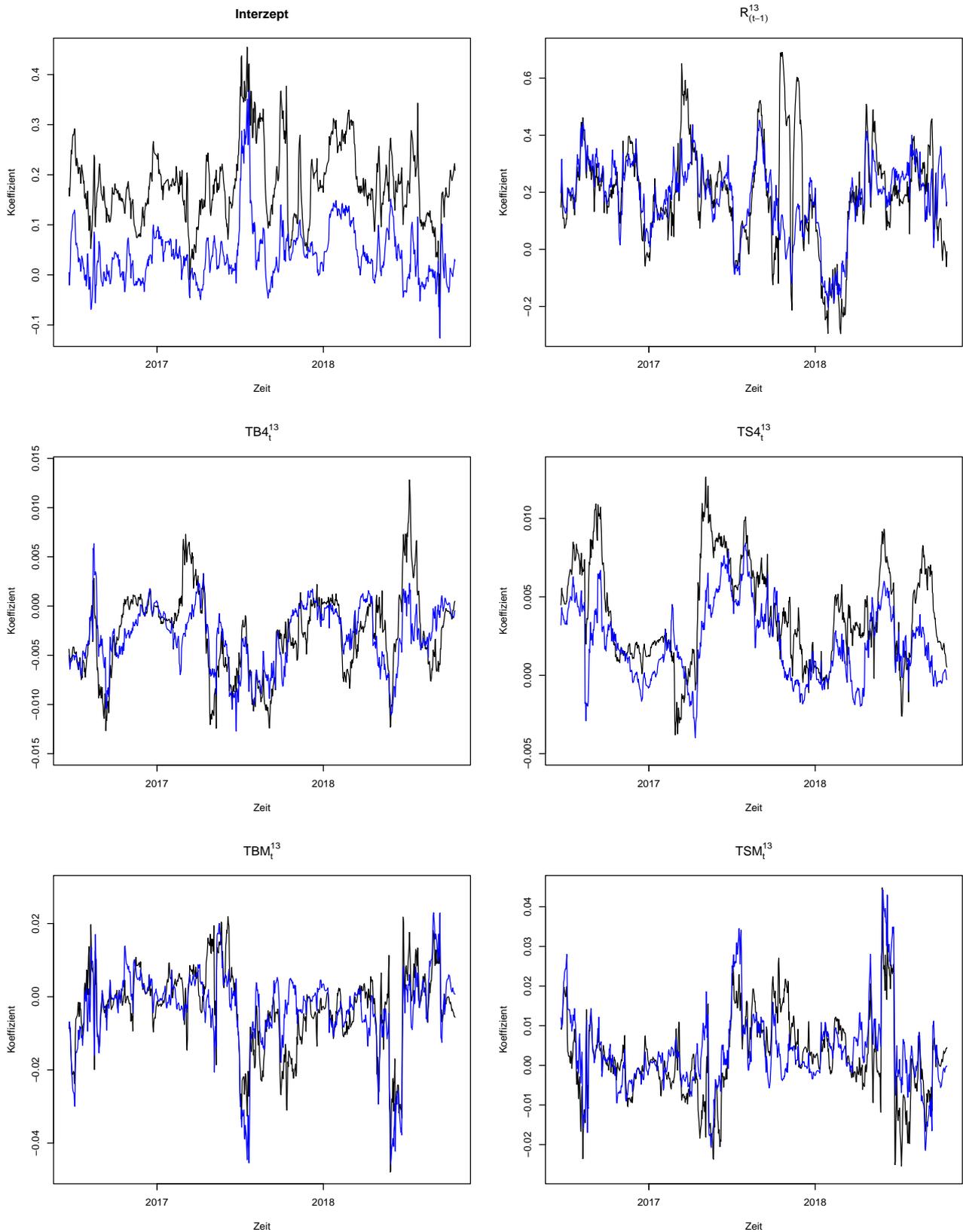


Abbildung 24: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 12:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 13:30 Uhr

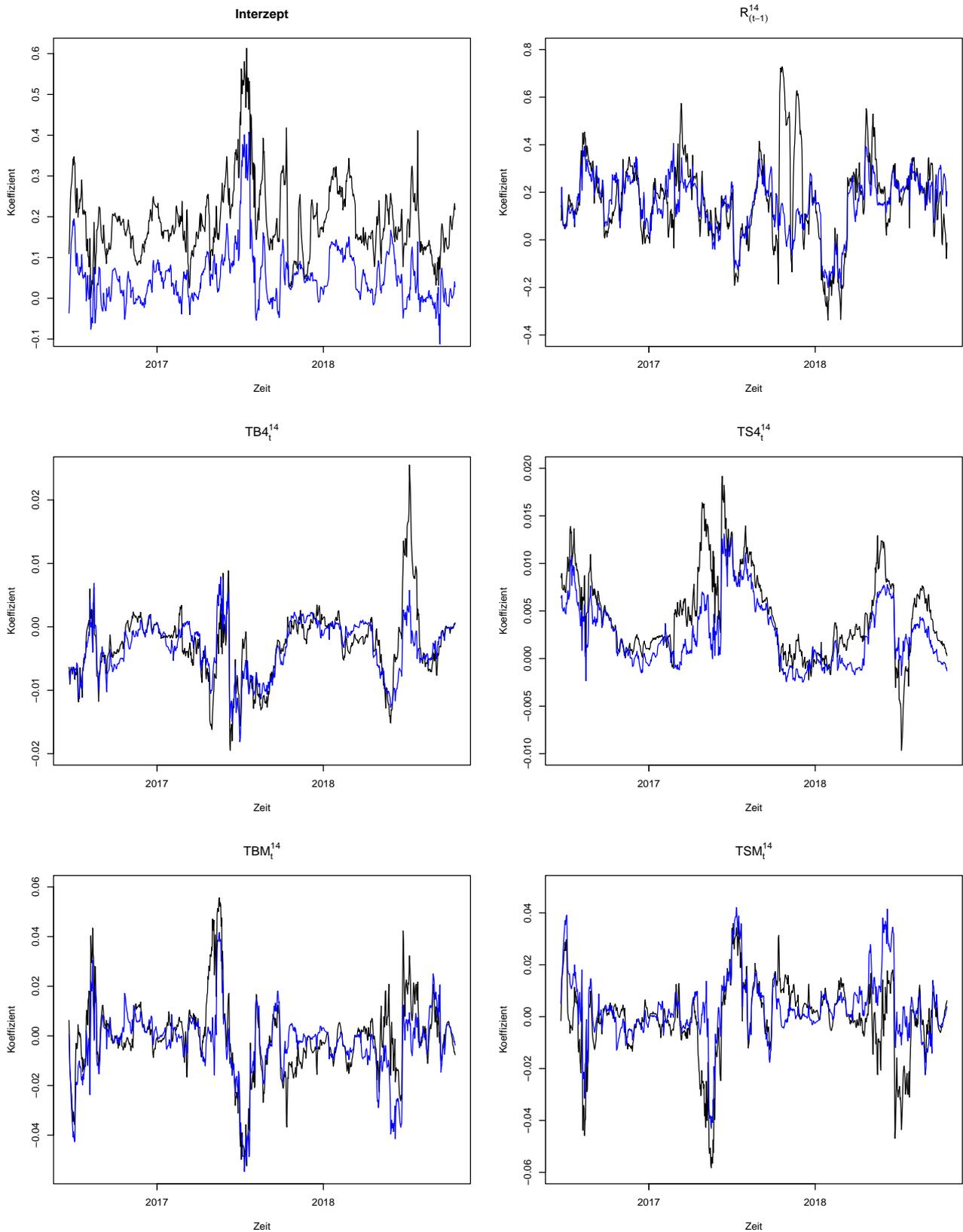


Abbildung 25: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 13:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 14:30 Uhr

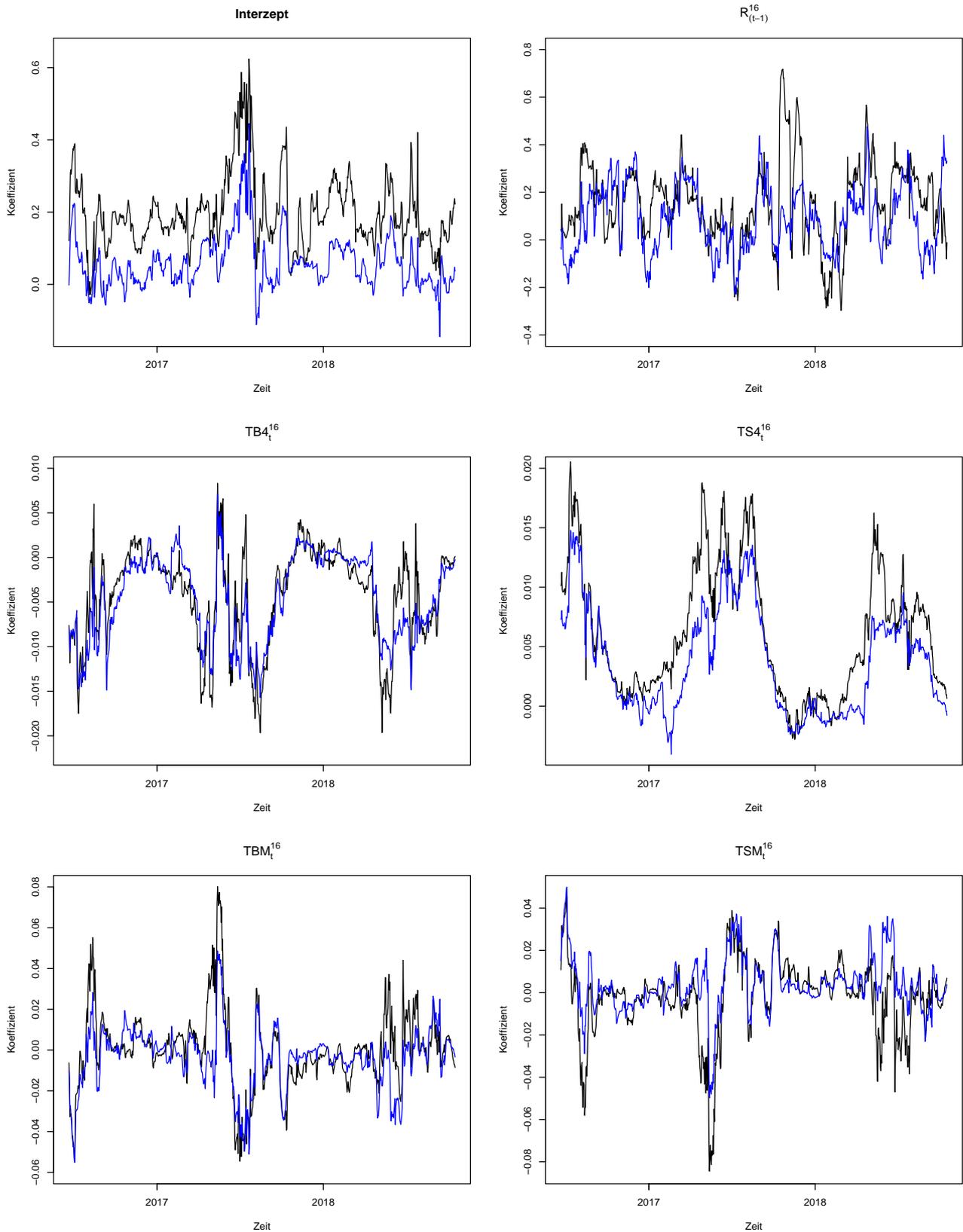


Abbildung 26: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 14:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 15:30 Uhr

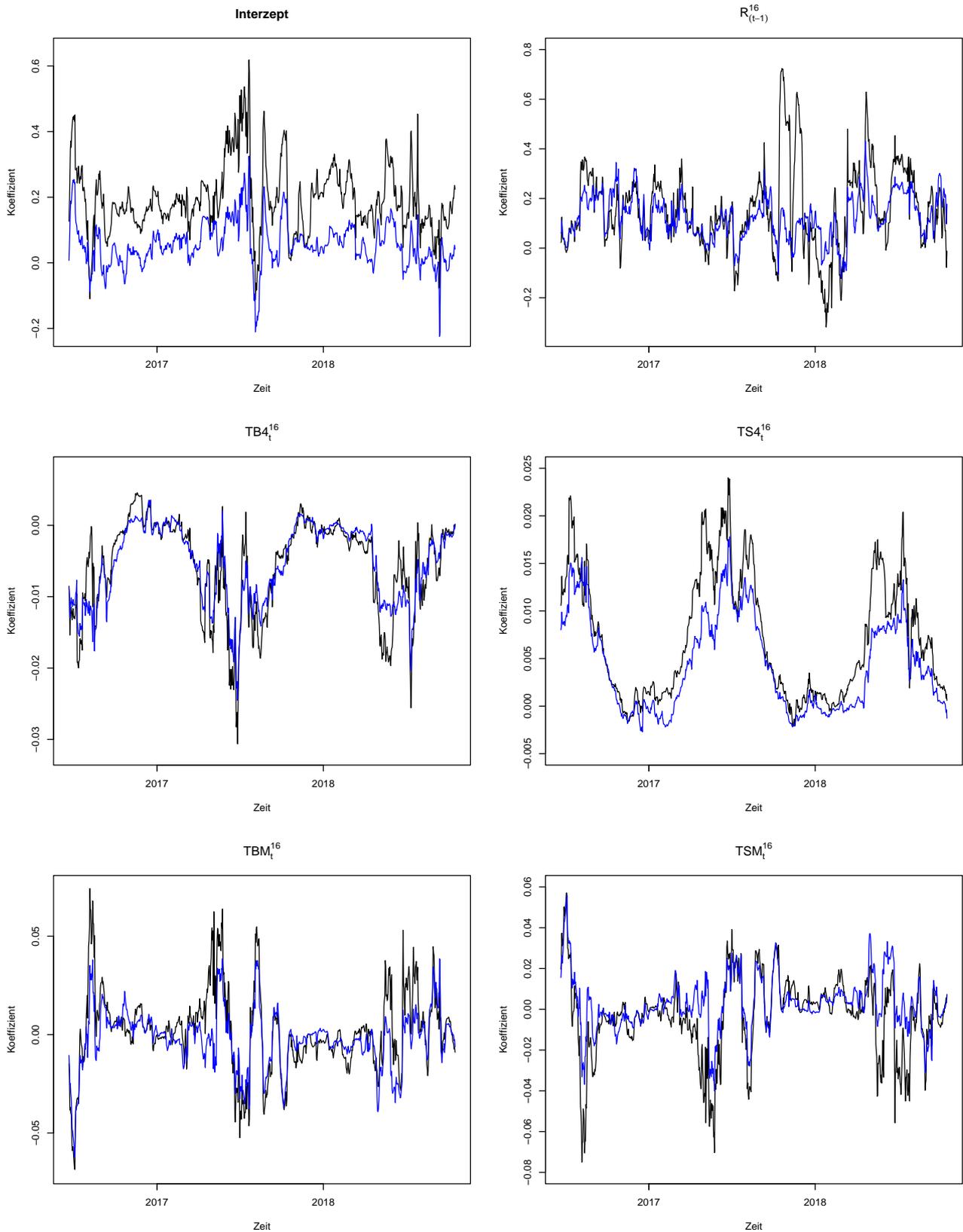


Abbildung 27: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 15:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 16:30 Uhr

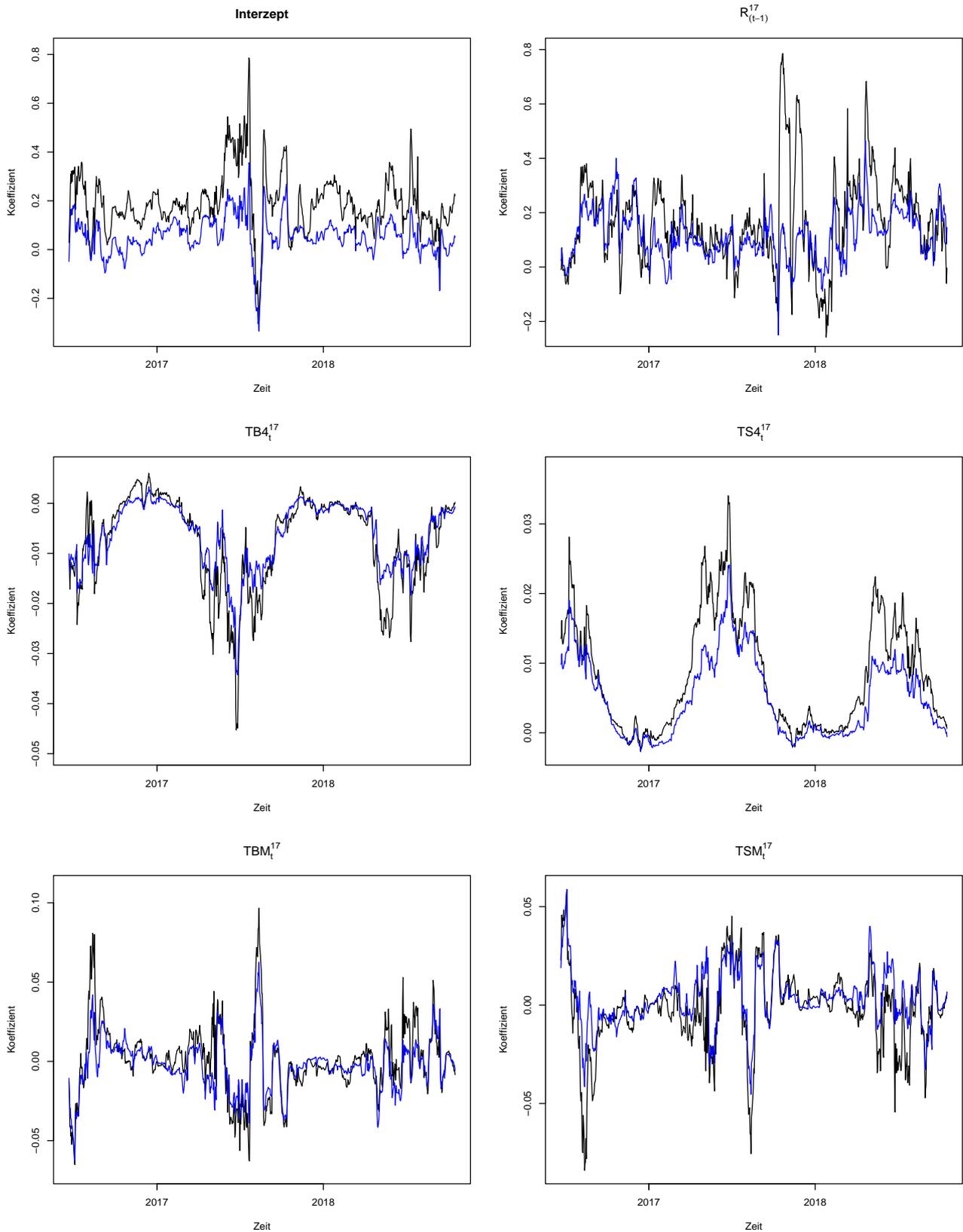


Abbildung 28: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 16:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 17:30 Uhr

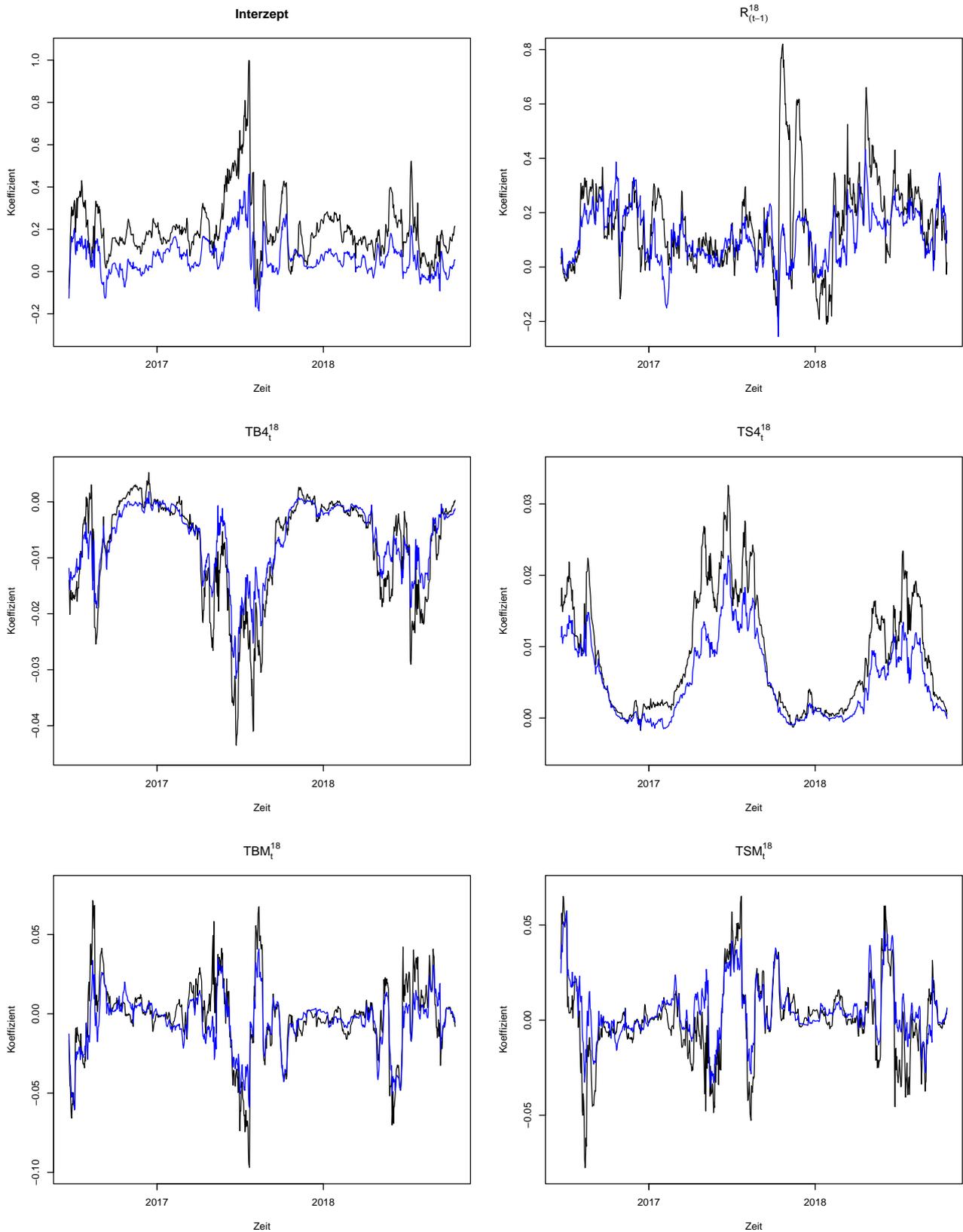


Abbildung 29: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 17:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 18:30 Uhr

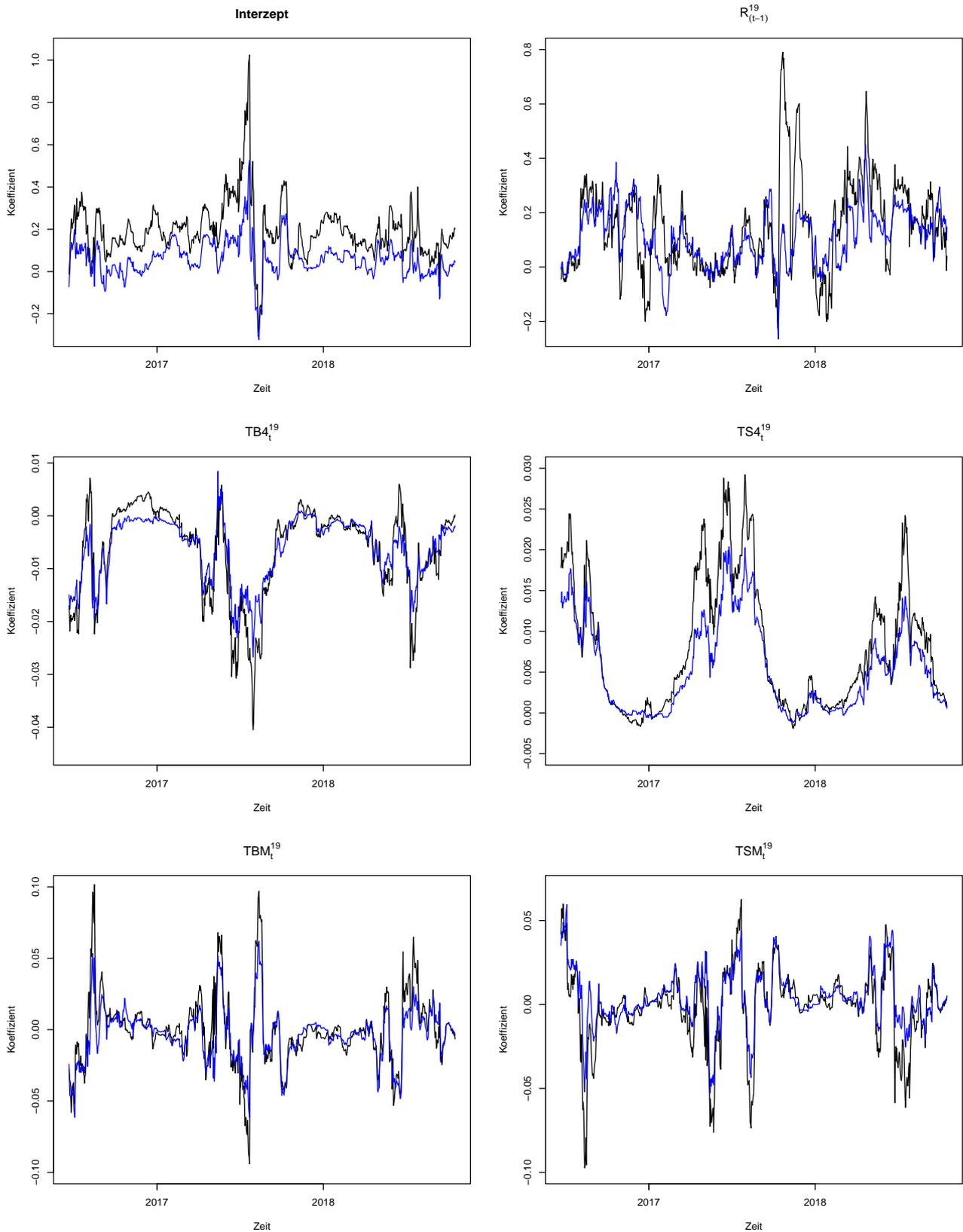


Abbildung 30: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 18:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 19:30 Uhr

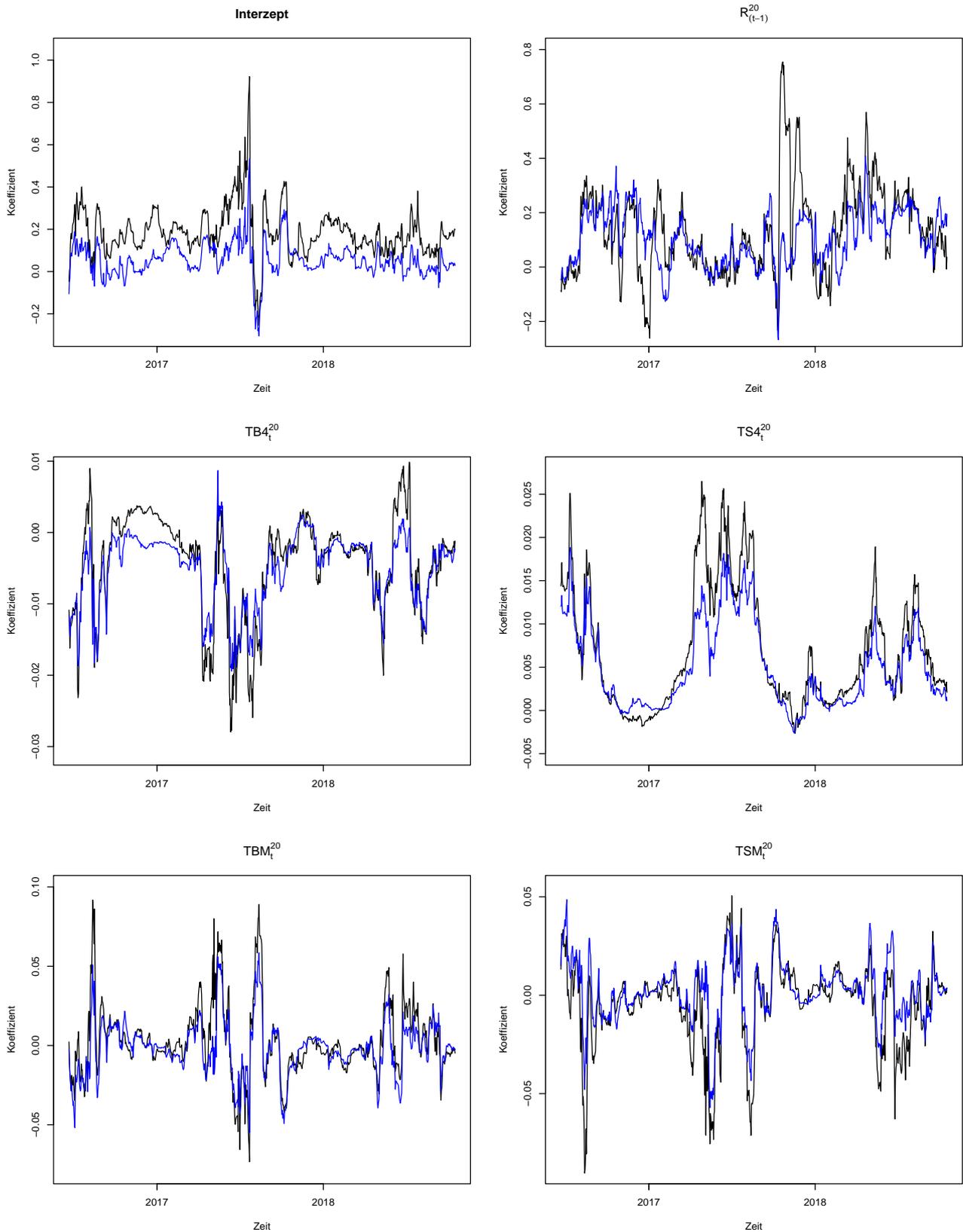


Abbildung 31: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 19:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 20:30 Uhr

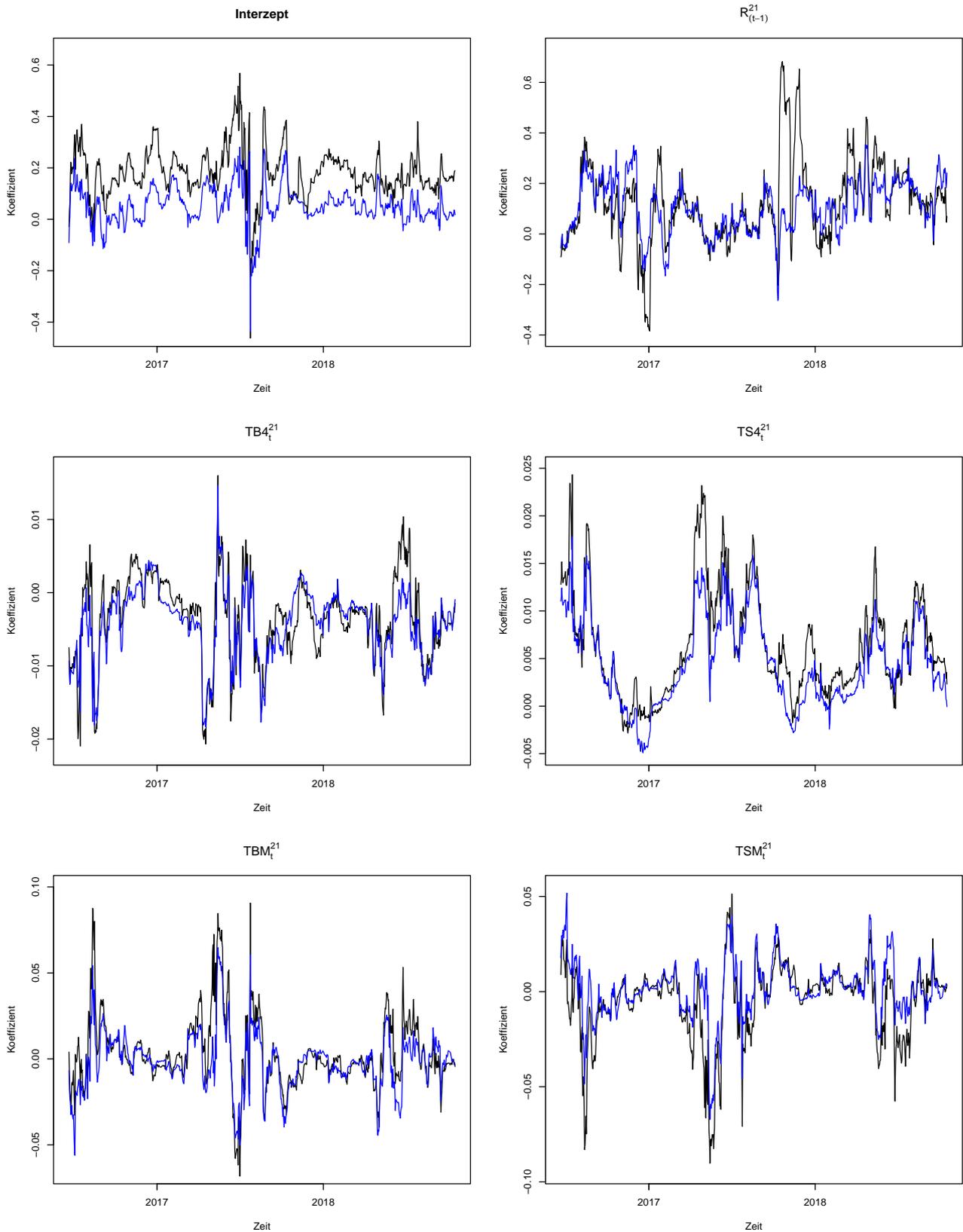


Abbildung 32: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 20:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 21:30 Uhr

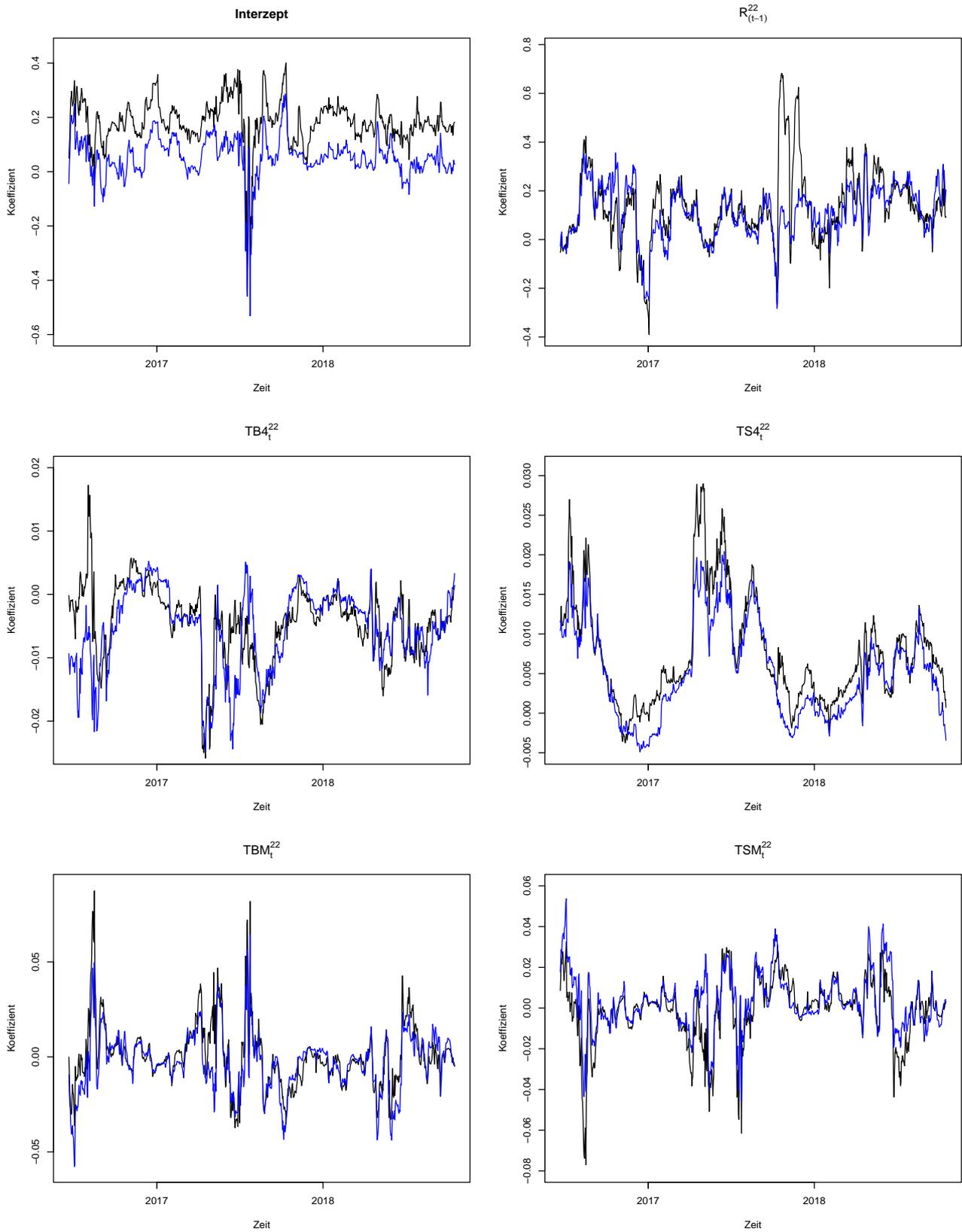


Abbildung 33: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 21:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 22:30 Uhr

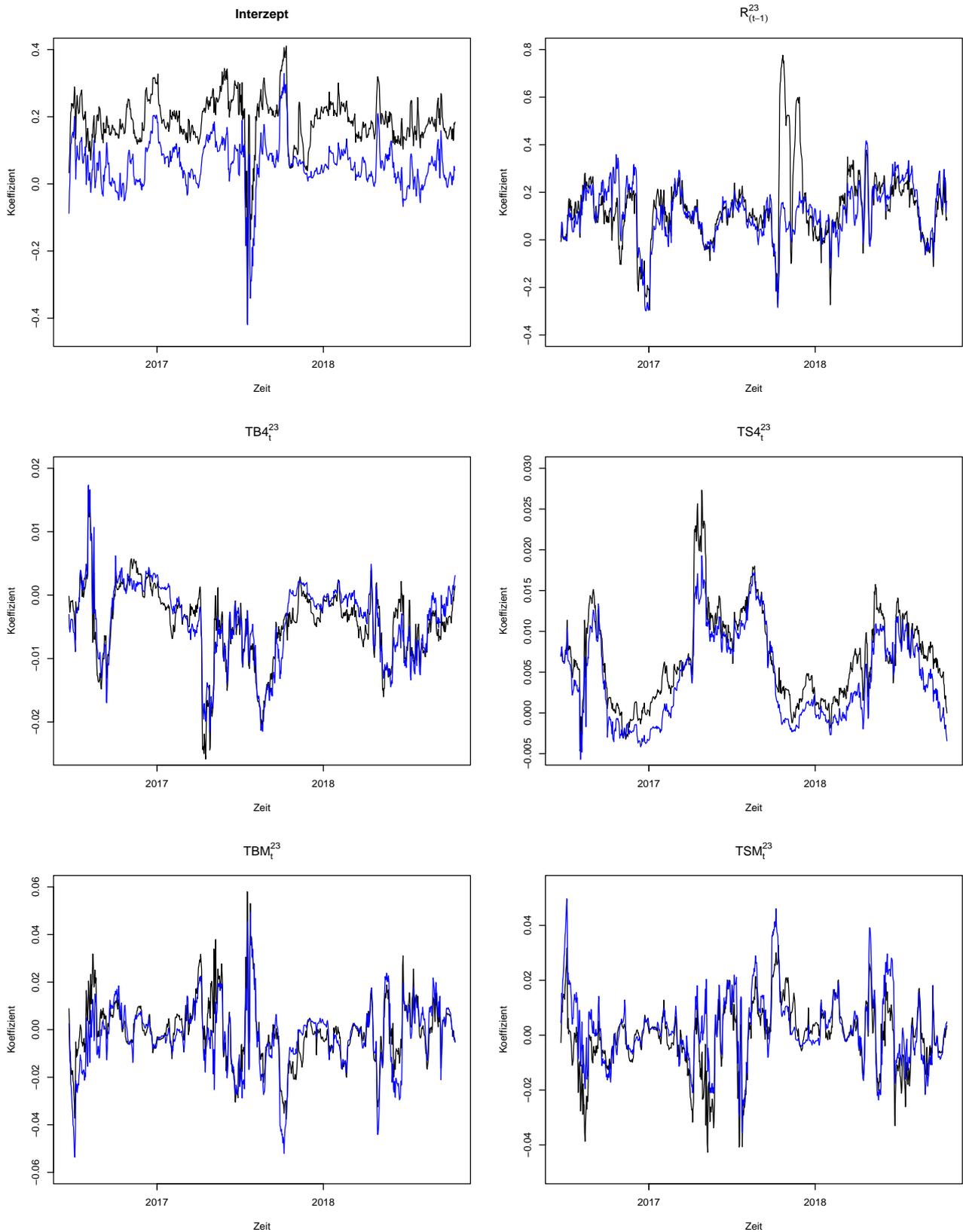


Abbildung 34: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 22:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Koeffizienten der Stunde 23:30 Uhr

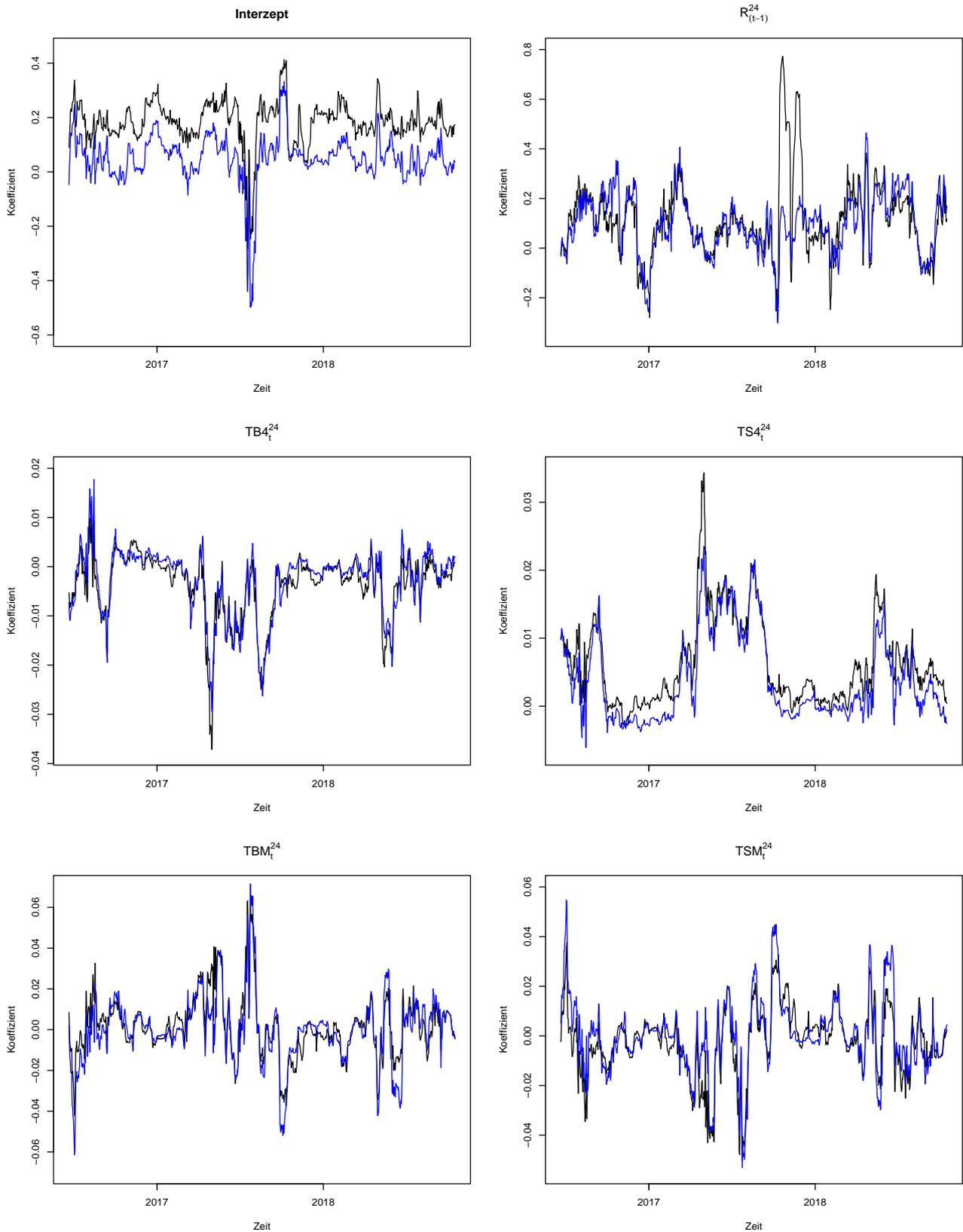


Abbildung 35: geschätzten Koeffizienten des Multivariaten Linearen Modells für den täglichen Messzeitpunkt 23:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# Wilk's Lambda

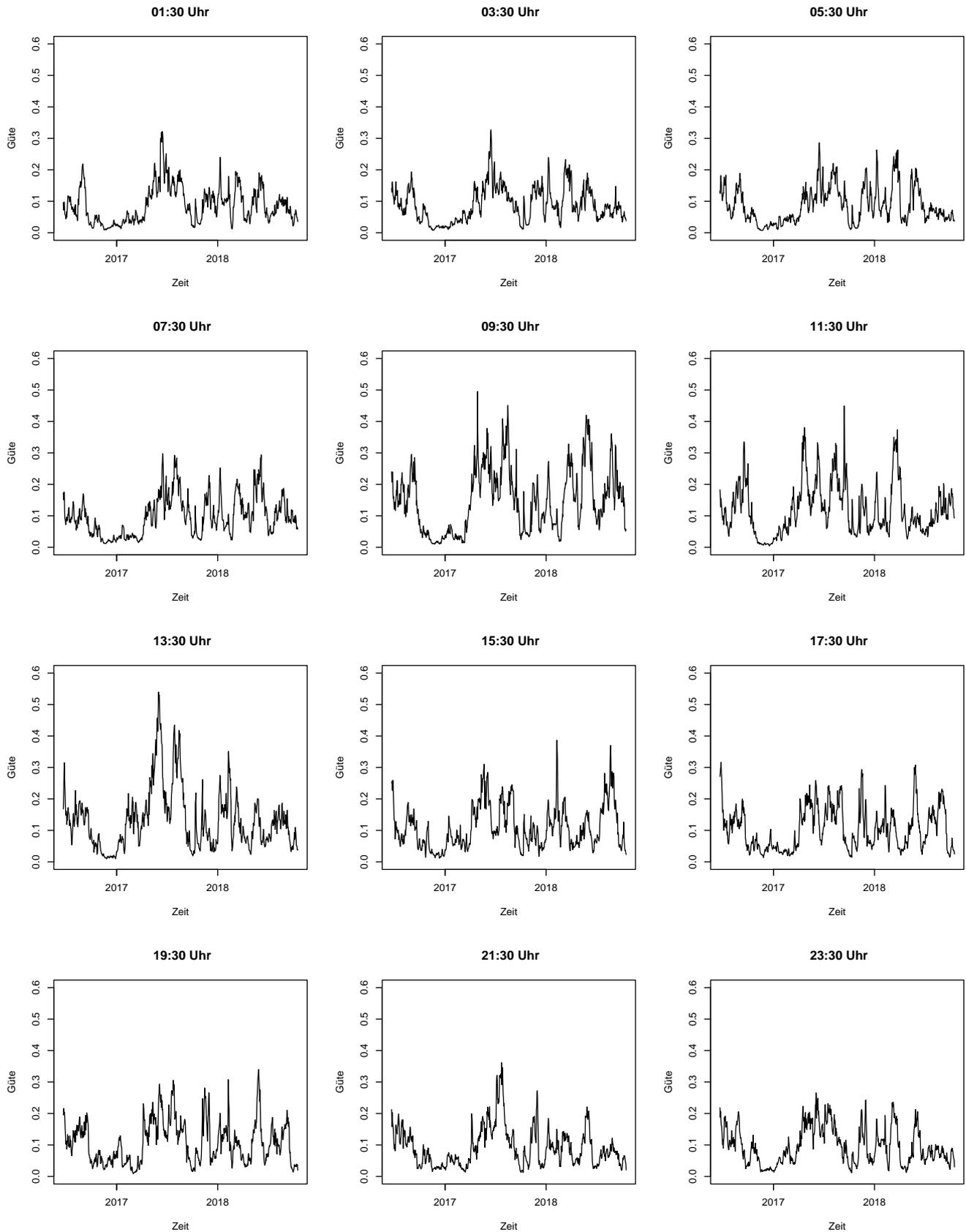


Abbildung 36: Gütekriterium Wilk's  $\Lambda$  zur Ermittlung der Modellanpassung eines Multivariaten Linearen Modells zur Analyse von Rissen im Baumaterial einer Brücke

# Gütekriterium $rt$

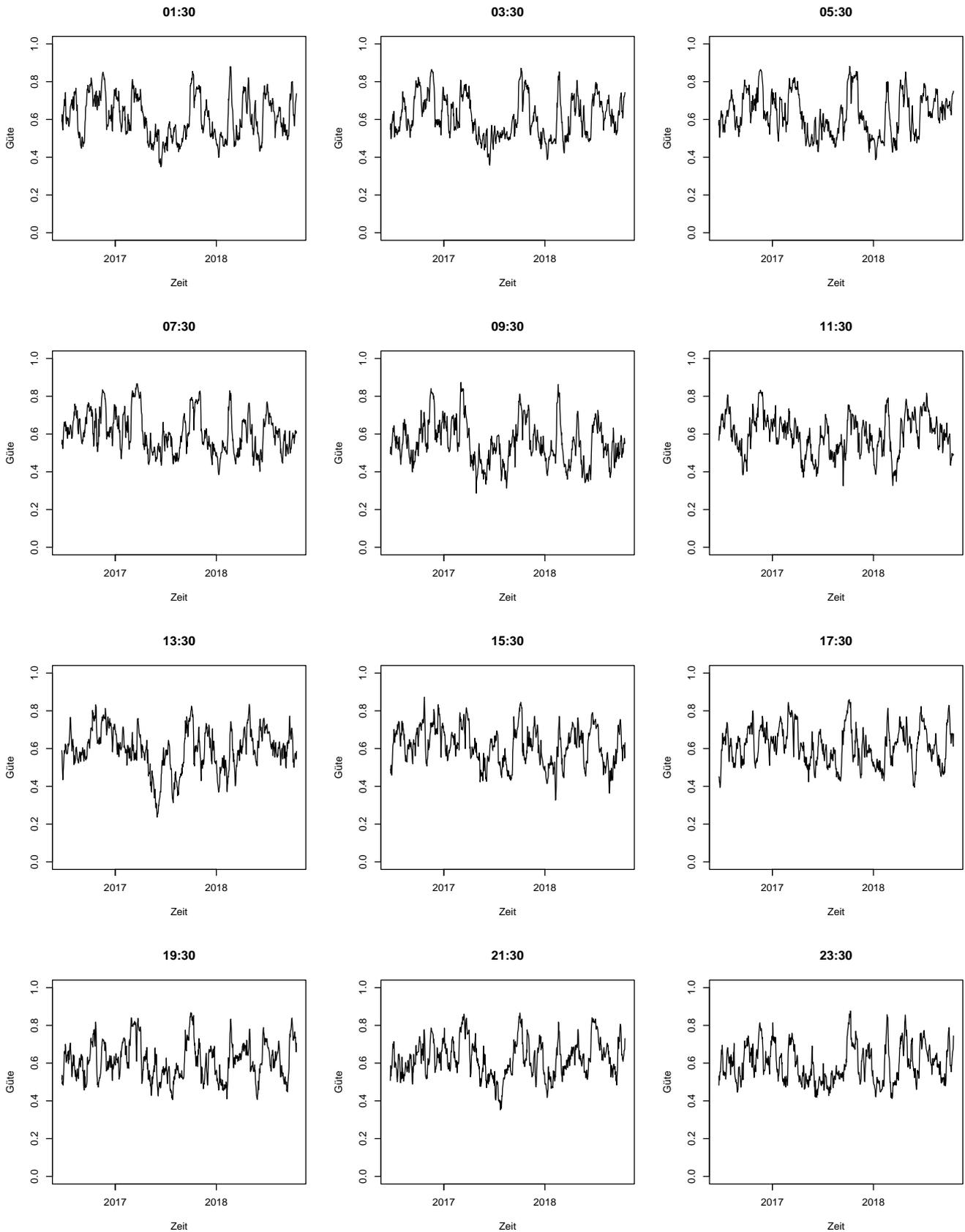


Abbildung 37: Gütekriterium  $rt$  zur Ermittlung der Modellanpassung eines Multivariaten Linearen Modells zur Analyse von Rissen im Baumaterial einer Brücke

# Gütekriterium $rd$

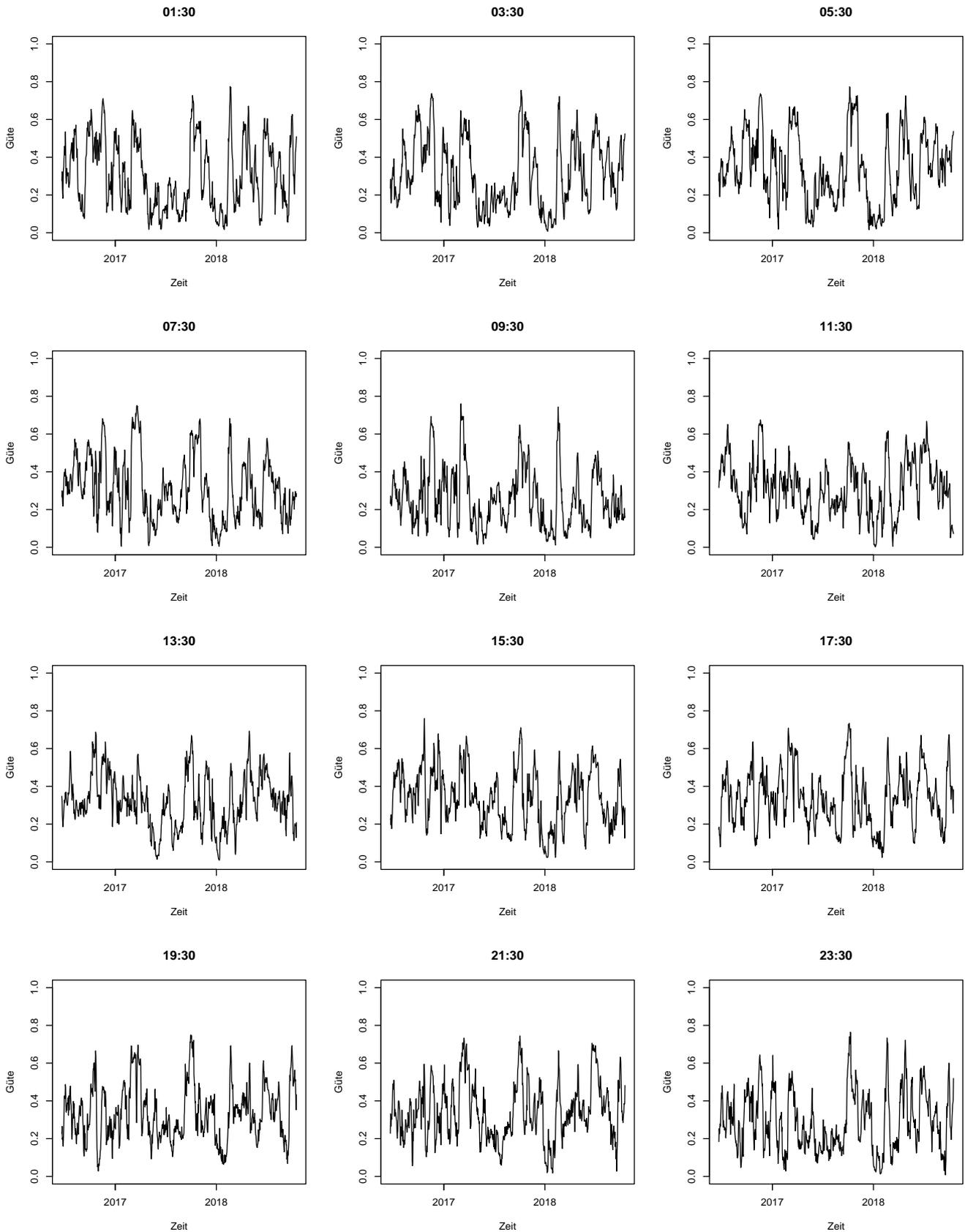


Abbildung 38: Gütekriterium  $rd$  zur Ermittlung der Modellanpassung eines Multivariaten Linearen Modells zur Analyse von Rissen im Baumaterial einer Brücke

# 00:30 Uhr

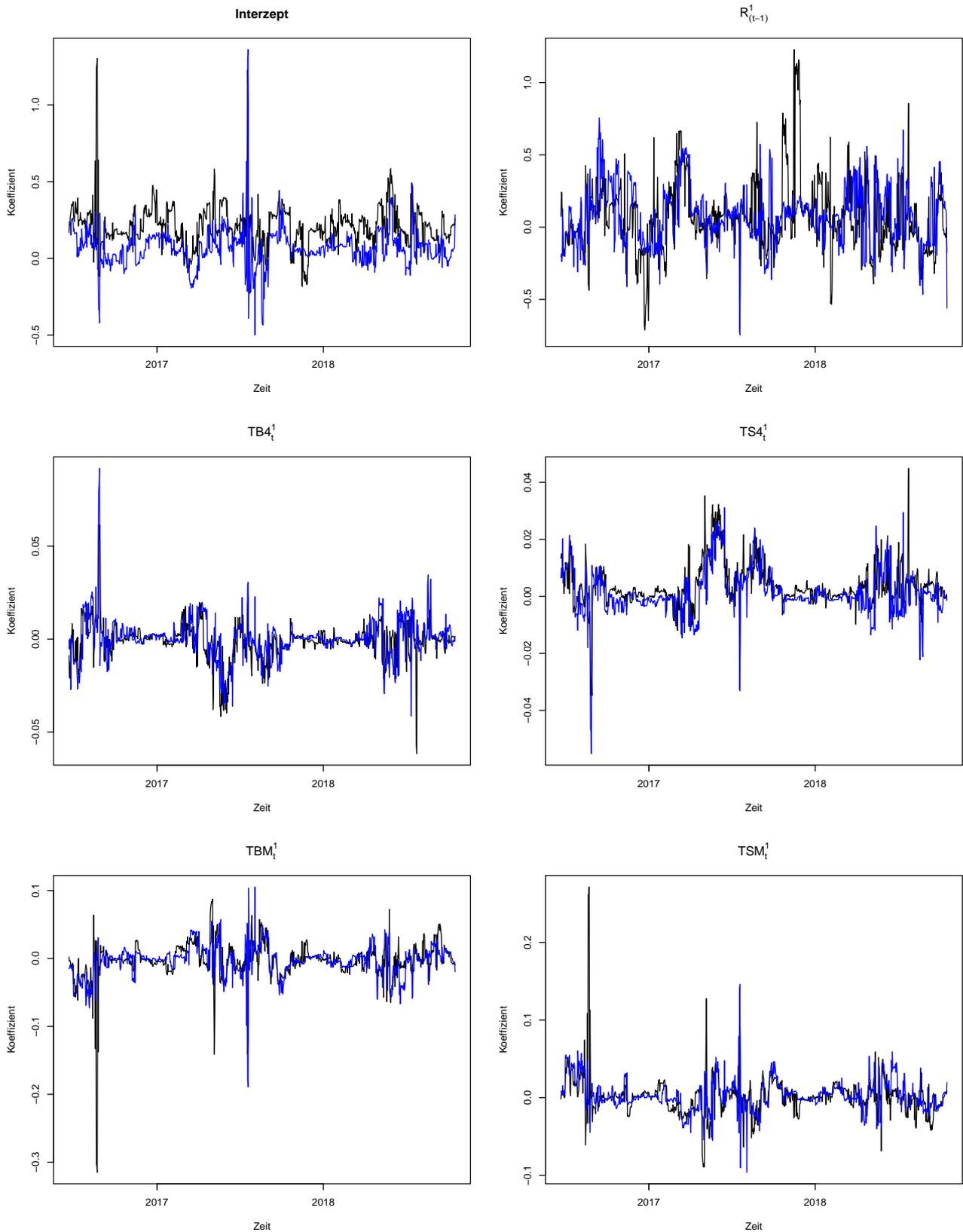


Abbildung 39: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 00:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 01:30 Uhr

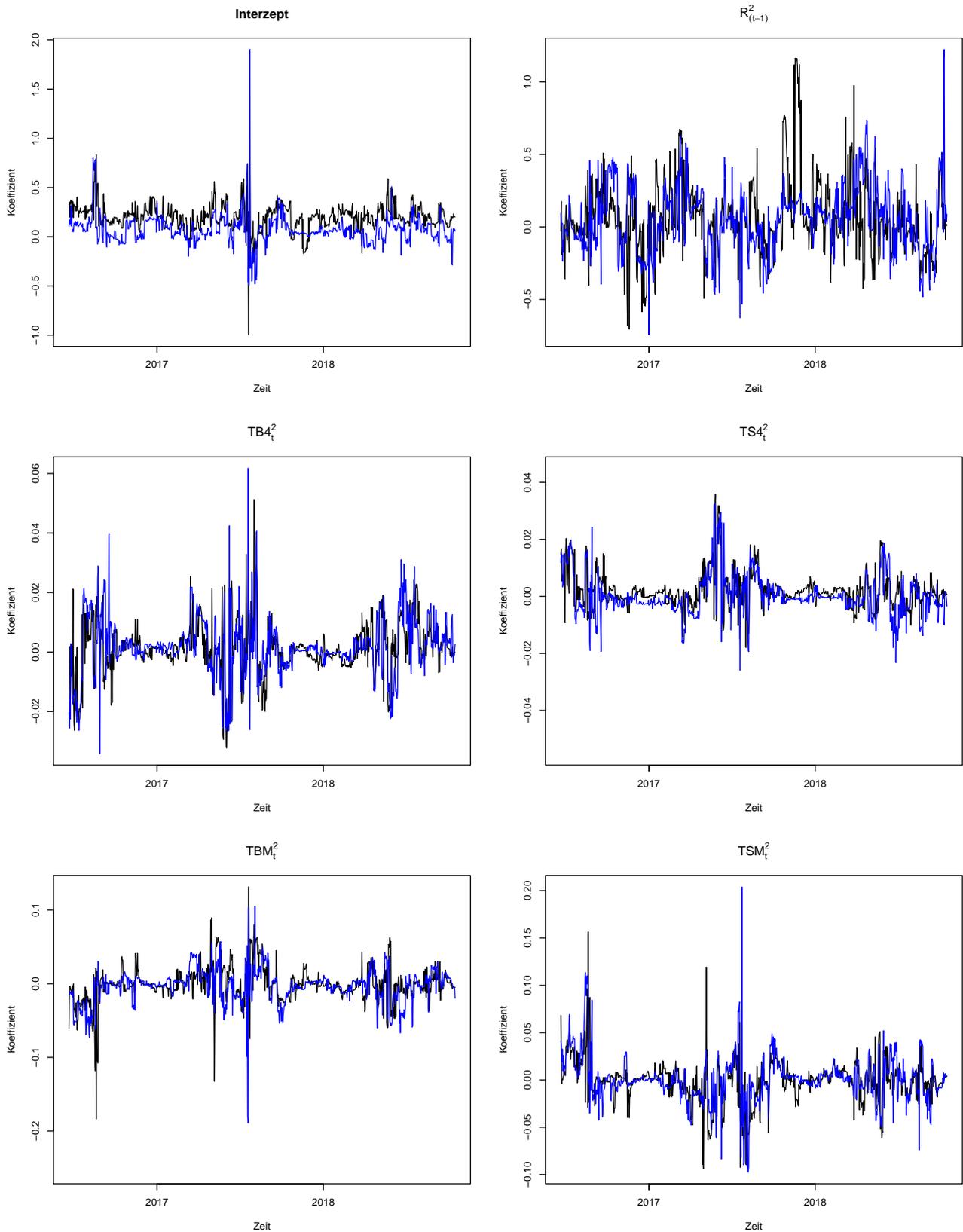


Abbildung 40: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 01:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 02:30 Uhr

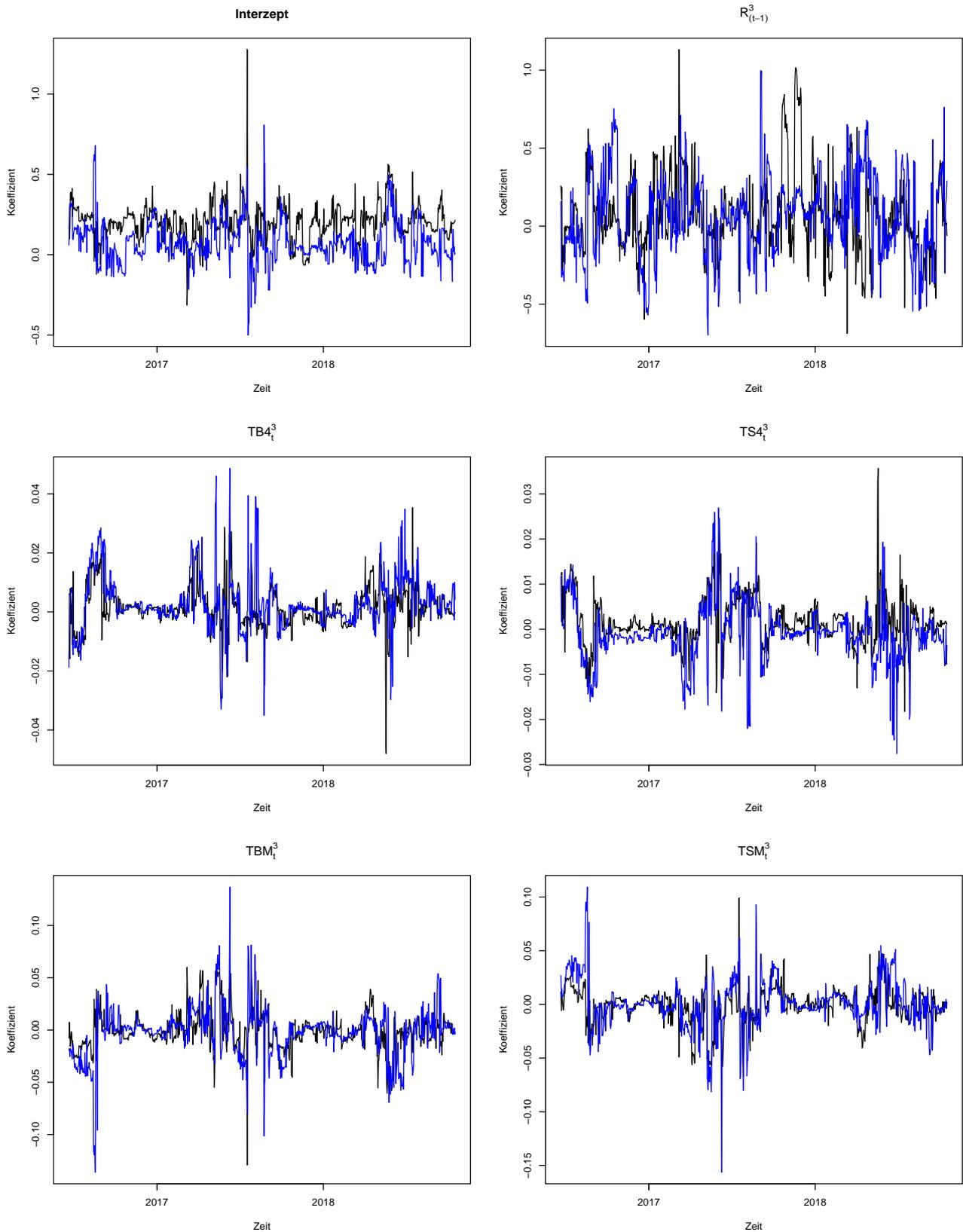


Abbildung 41: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 02:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 03:30 Uhr

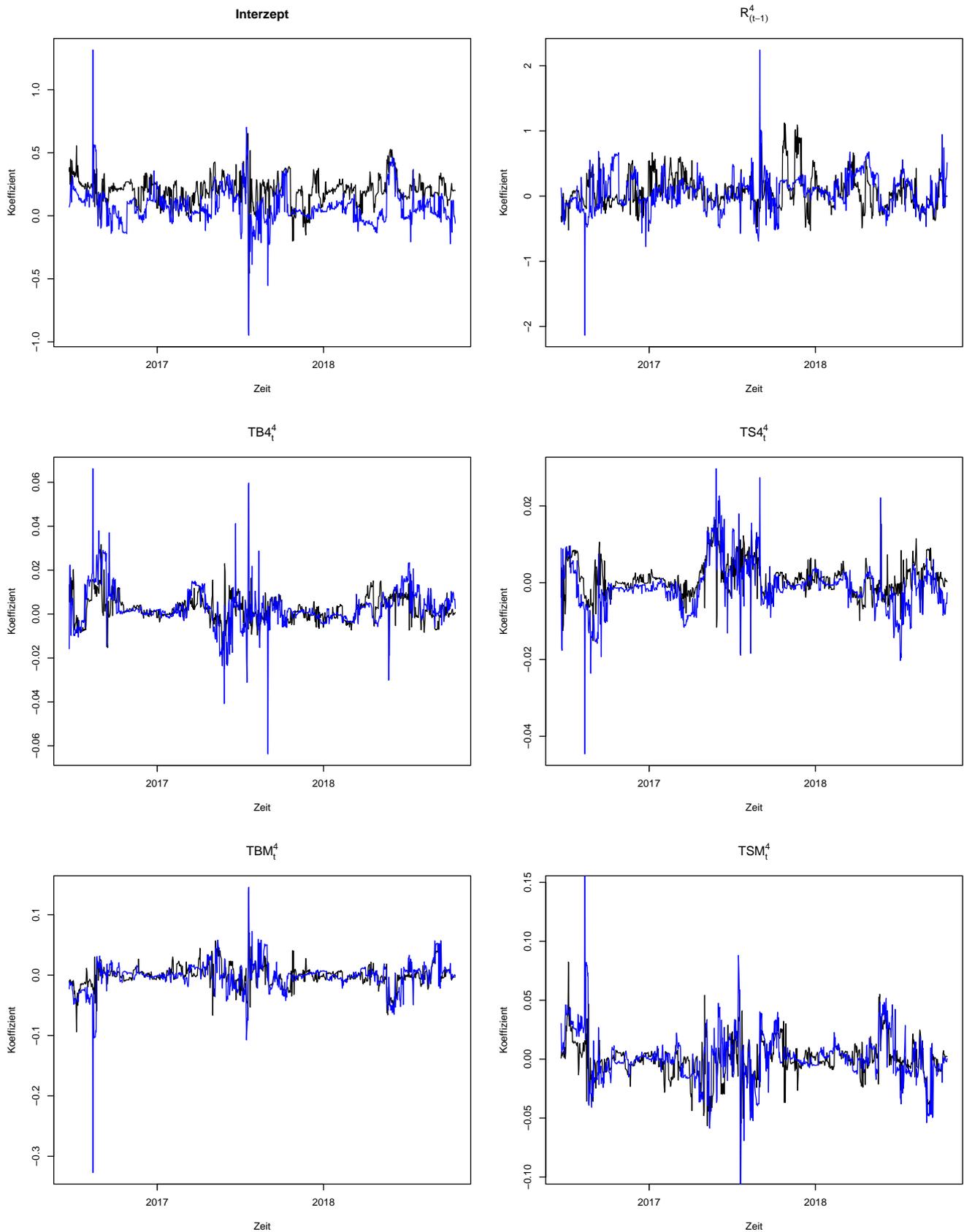


Abbildung 42: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 03:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 04:30 Uhr

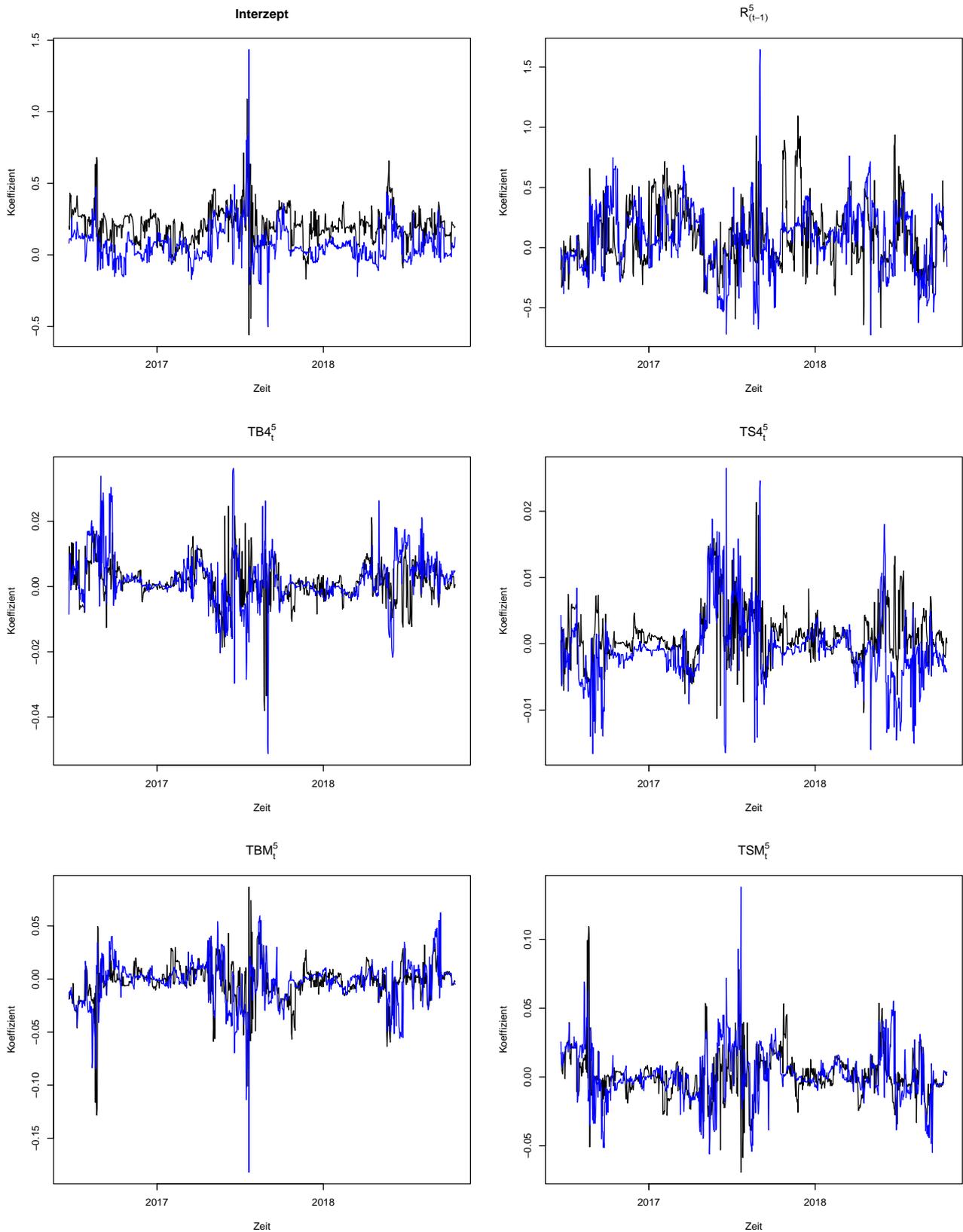


Abbildung 43: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 04:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 05:30 Uhr

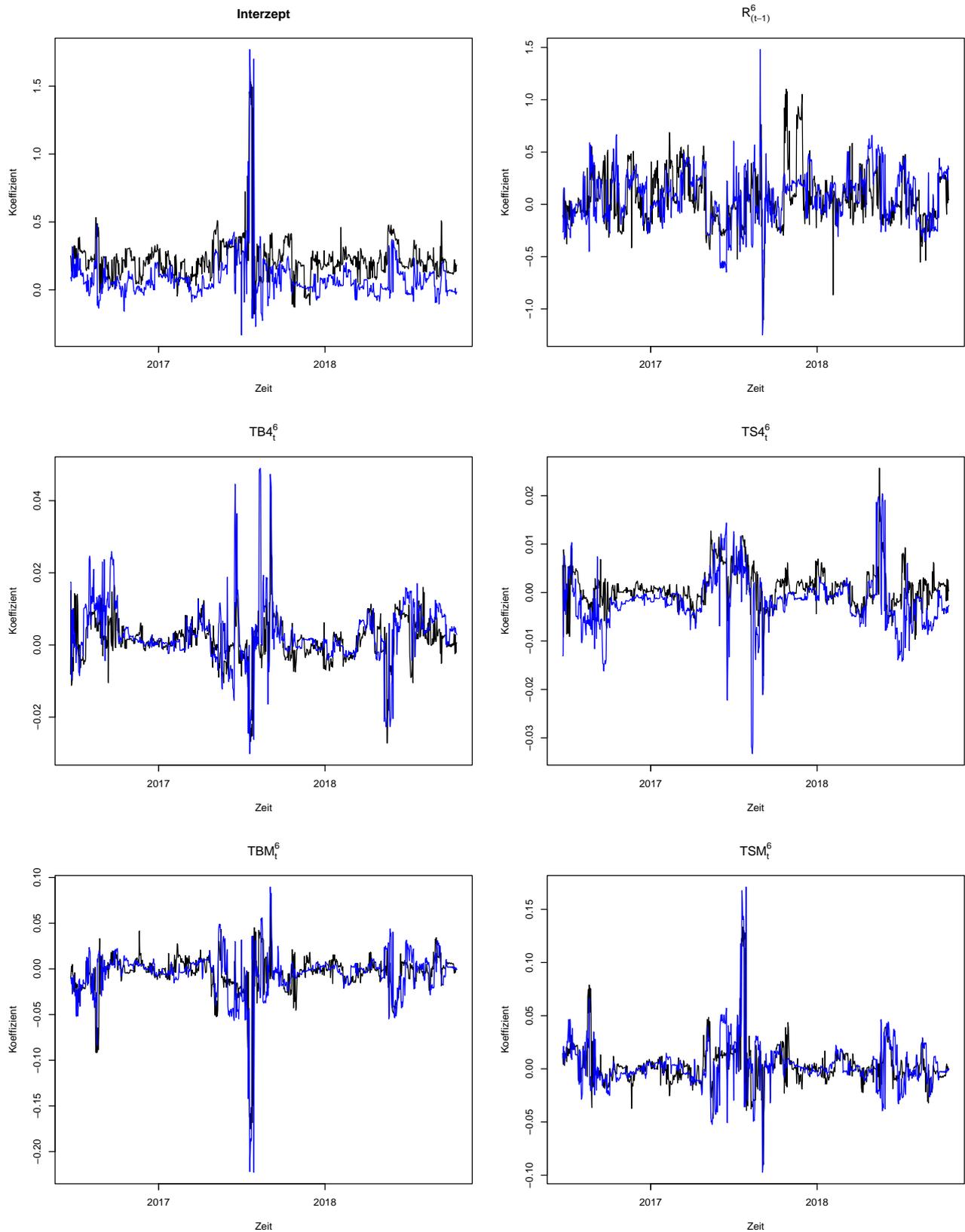


Abbildung 44: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 05:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 07:30 Uhr

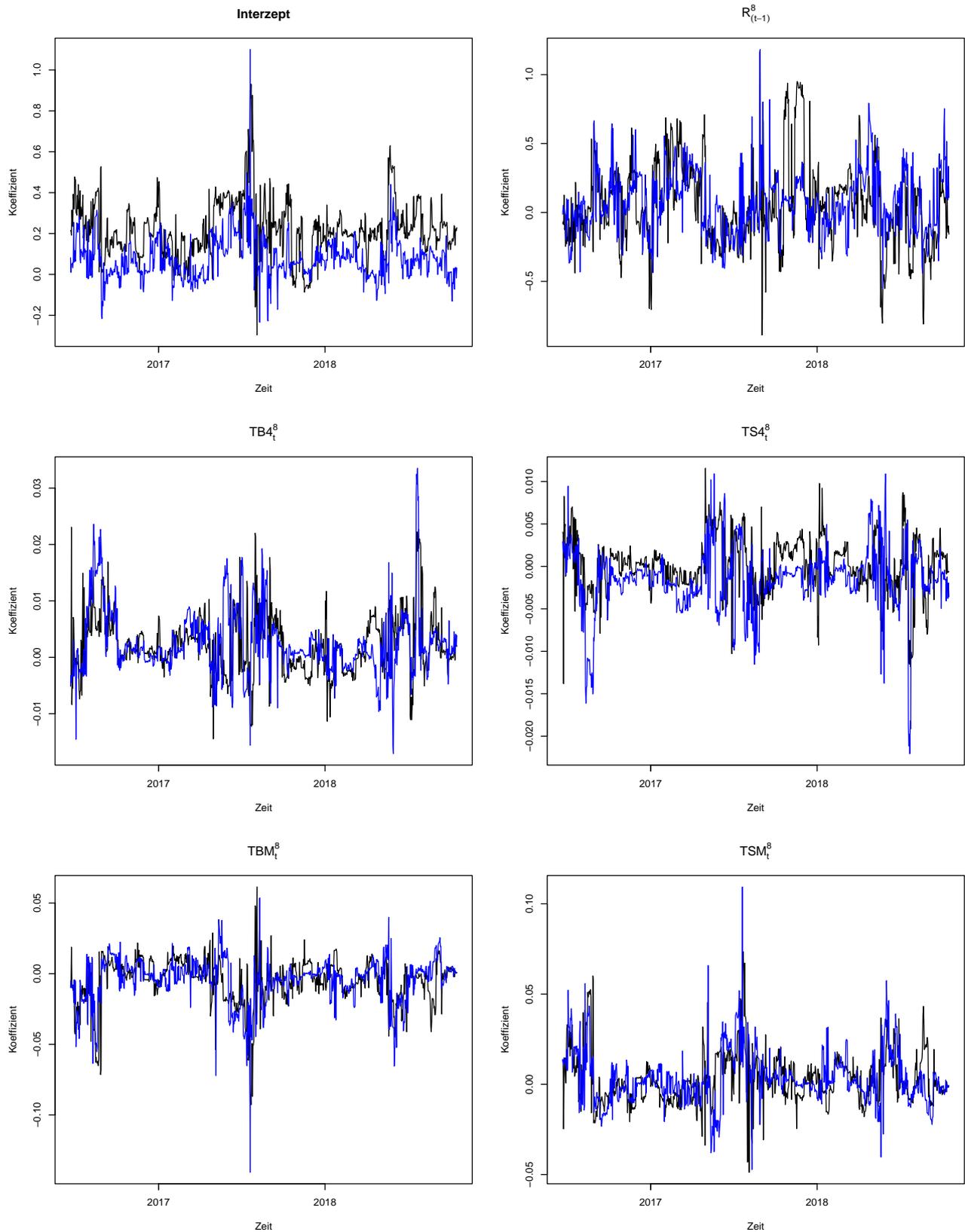


Abbildung 45: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 07:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 08:30 Uhr

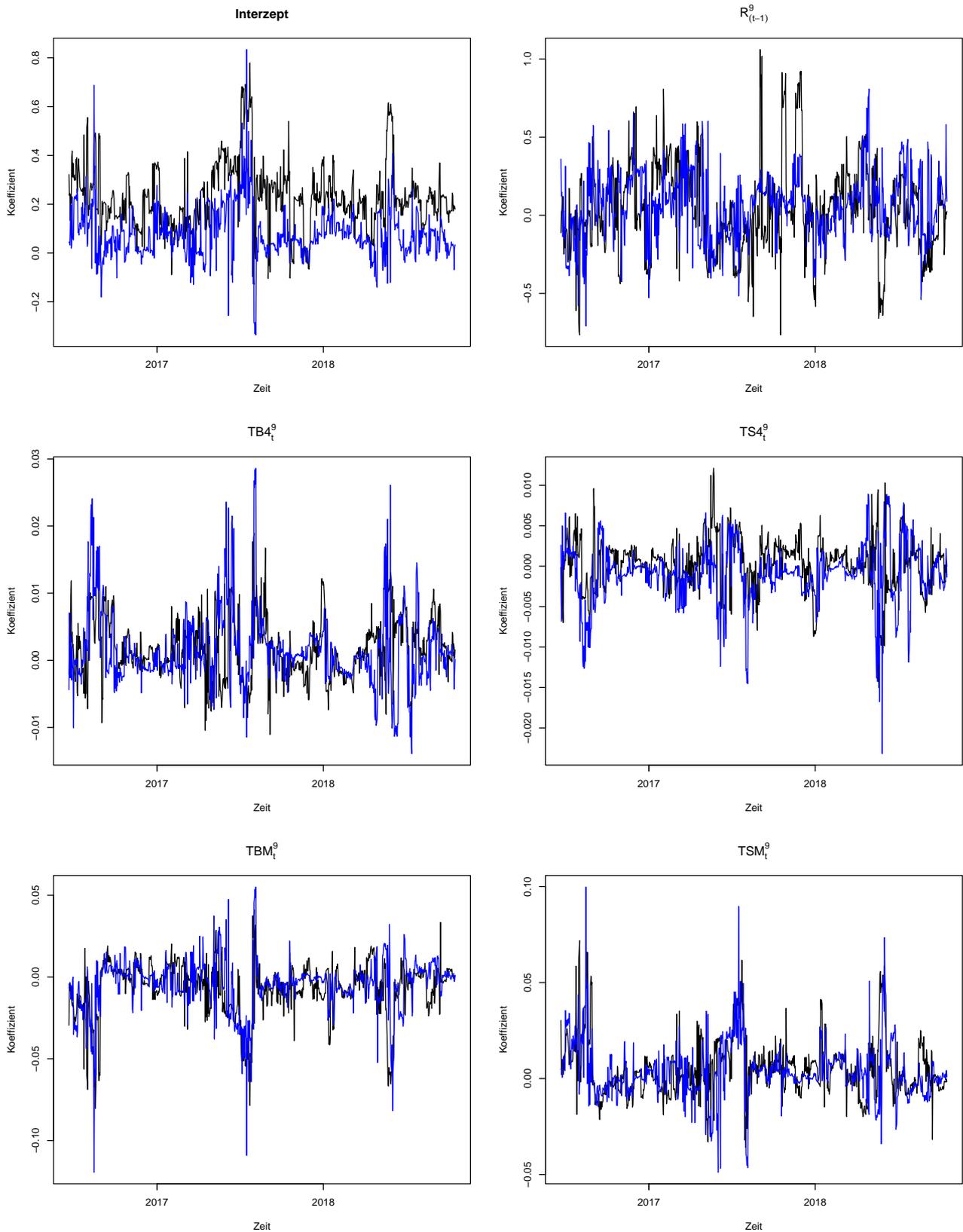


Abbildung 46: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 08:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 09:30 Uhr

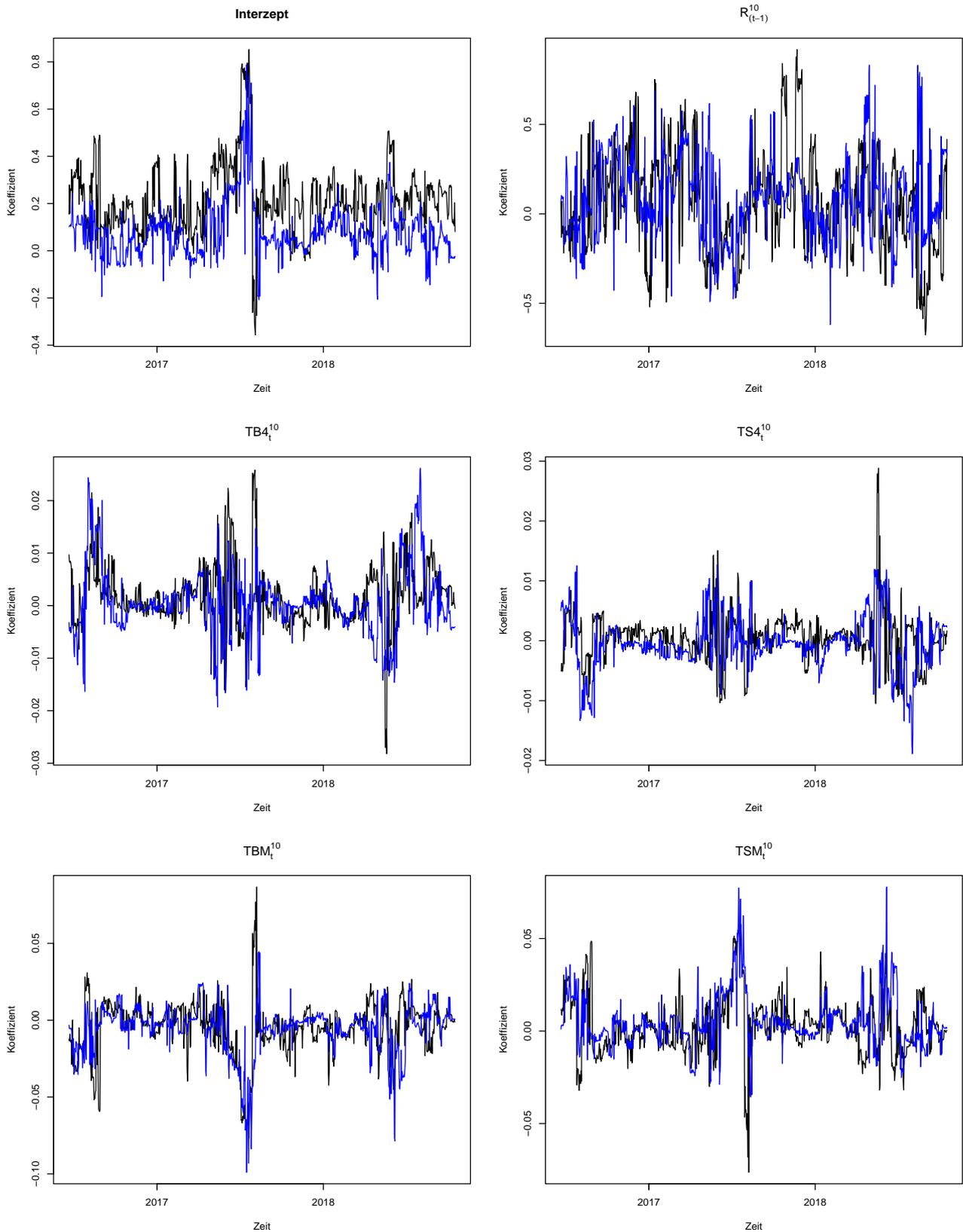


Abbildung 47: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 09:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 10:30 Uhr

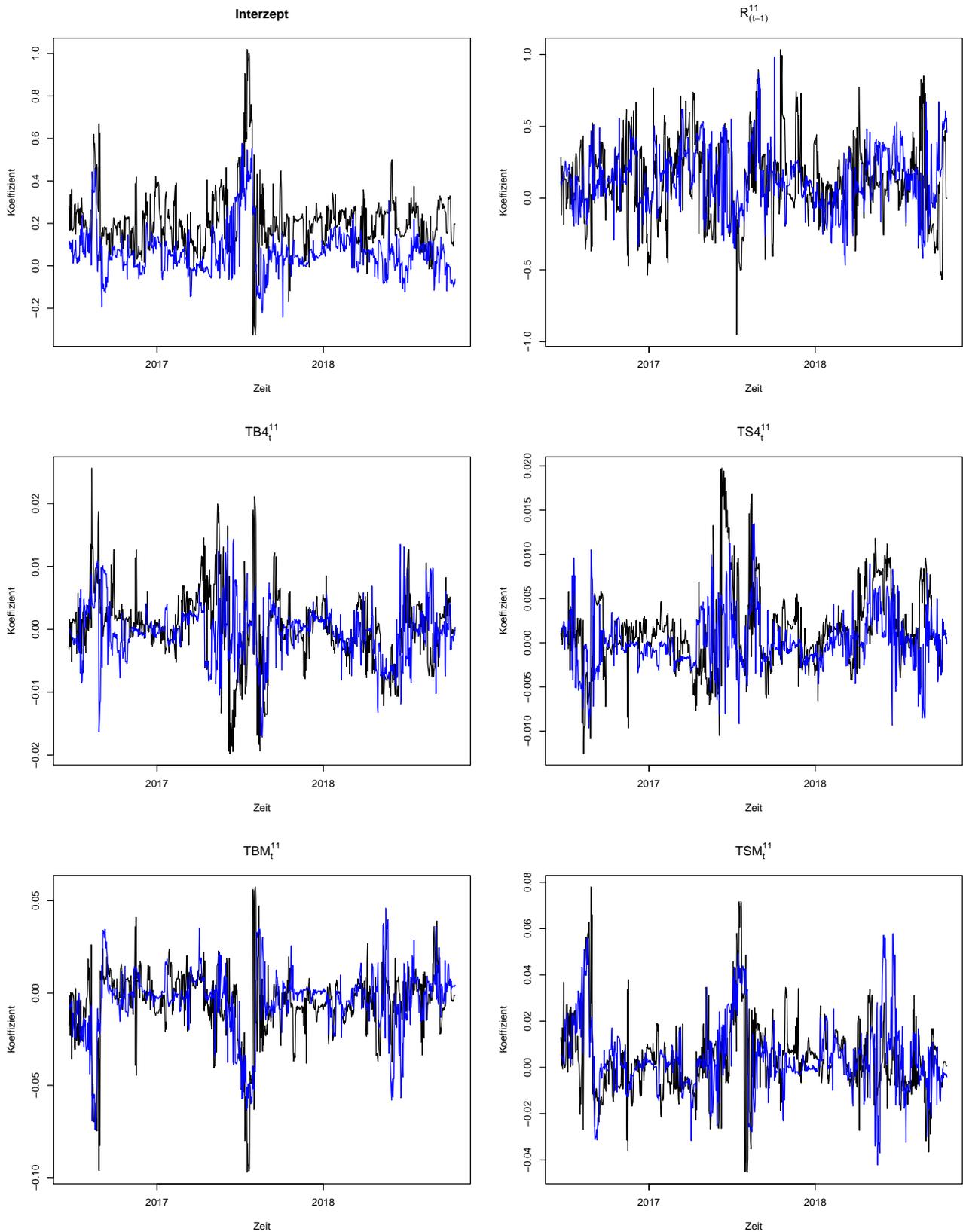


Abbildung 48: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 10:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 11:30 Uhr

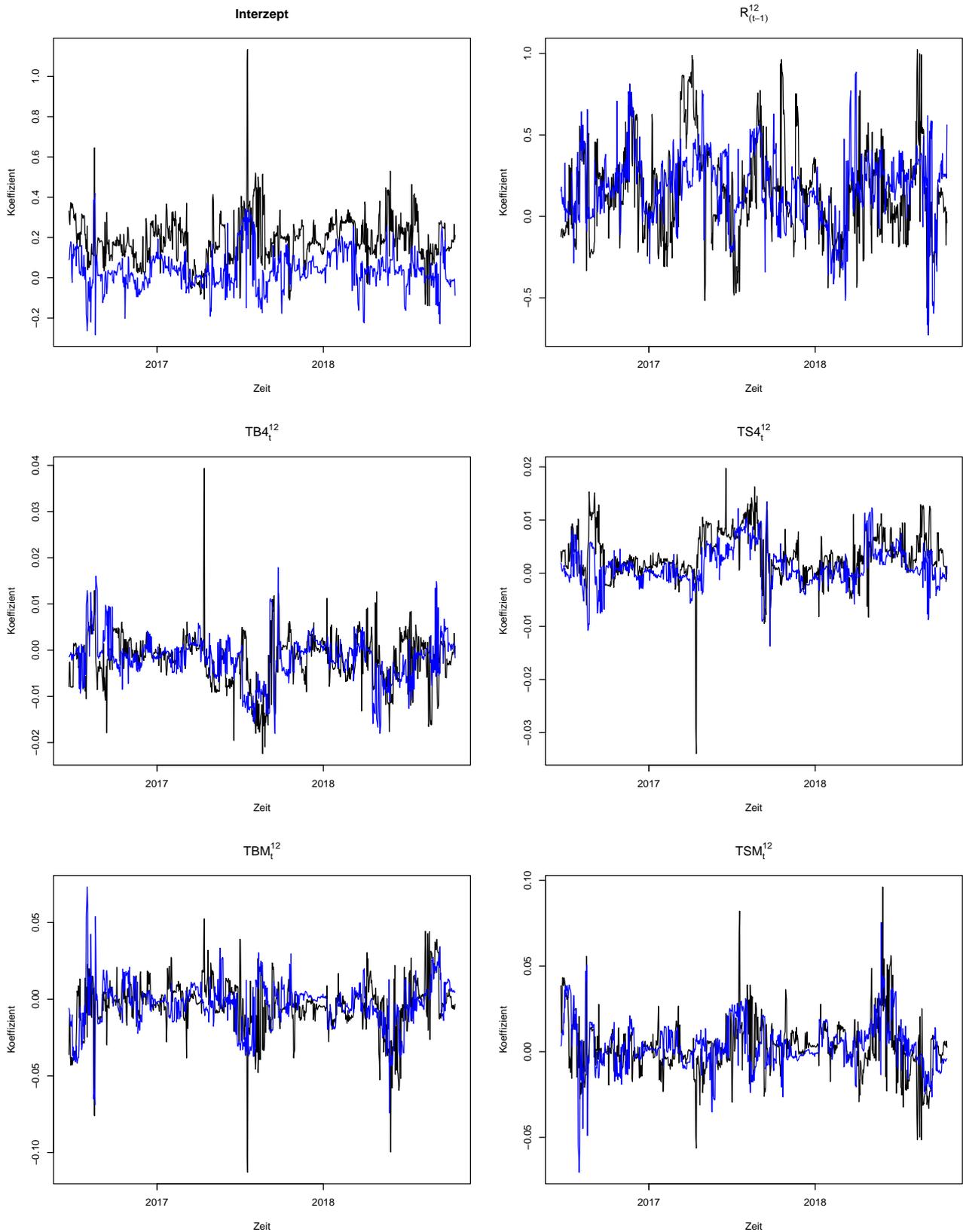


Abbildung 49: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 11:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 12:30 Uhr

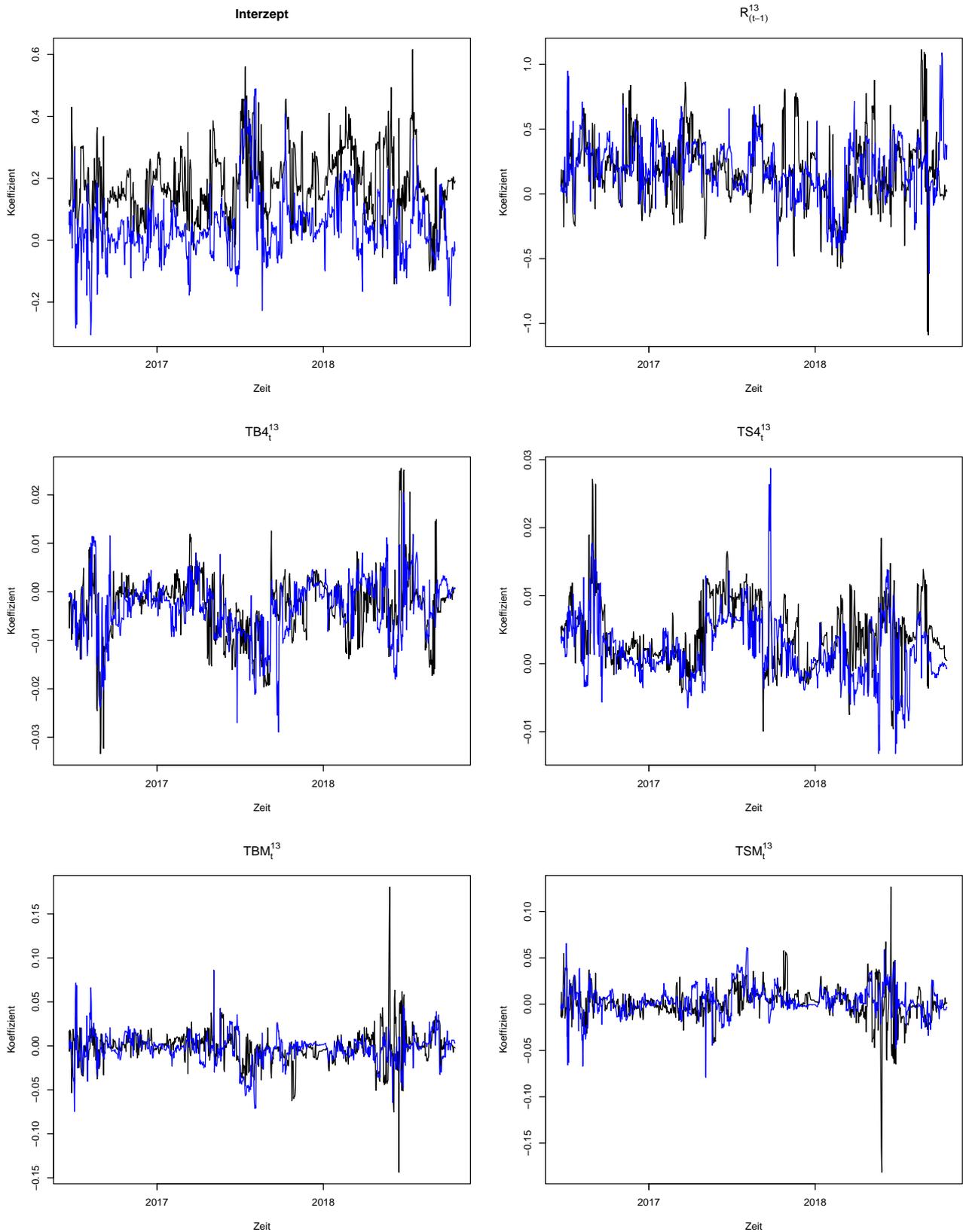


Abbildung 50: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 12:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 13:30 Uhr

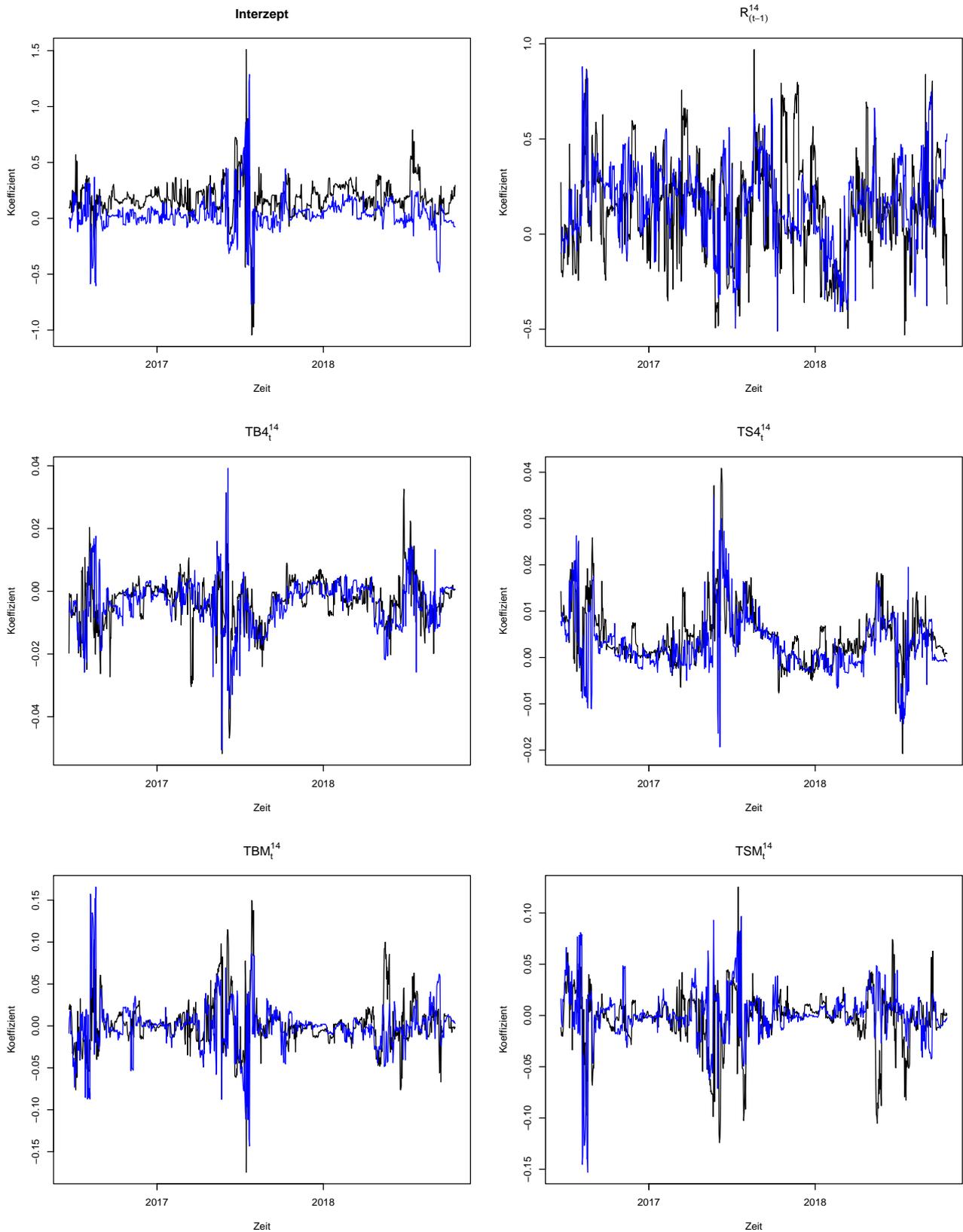


Abbildung 51: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 13:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 14:30 Uhr

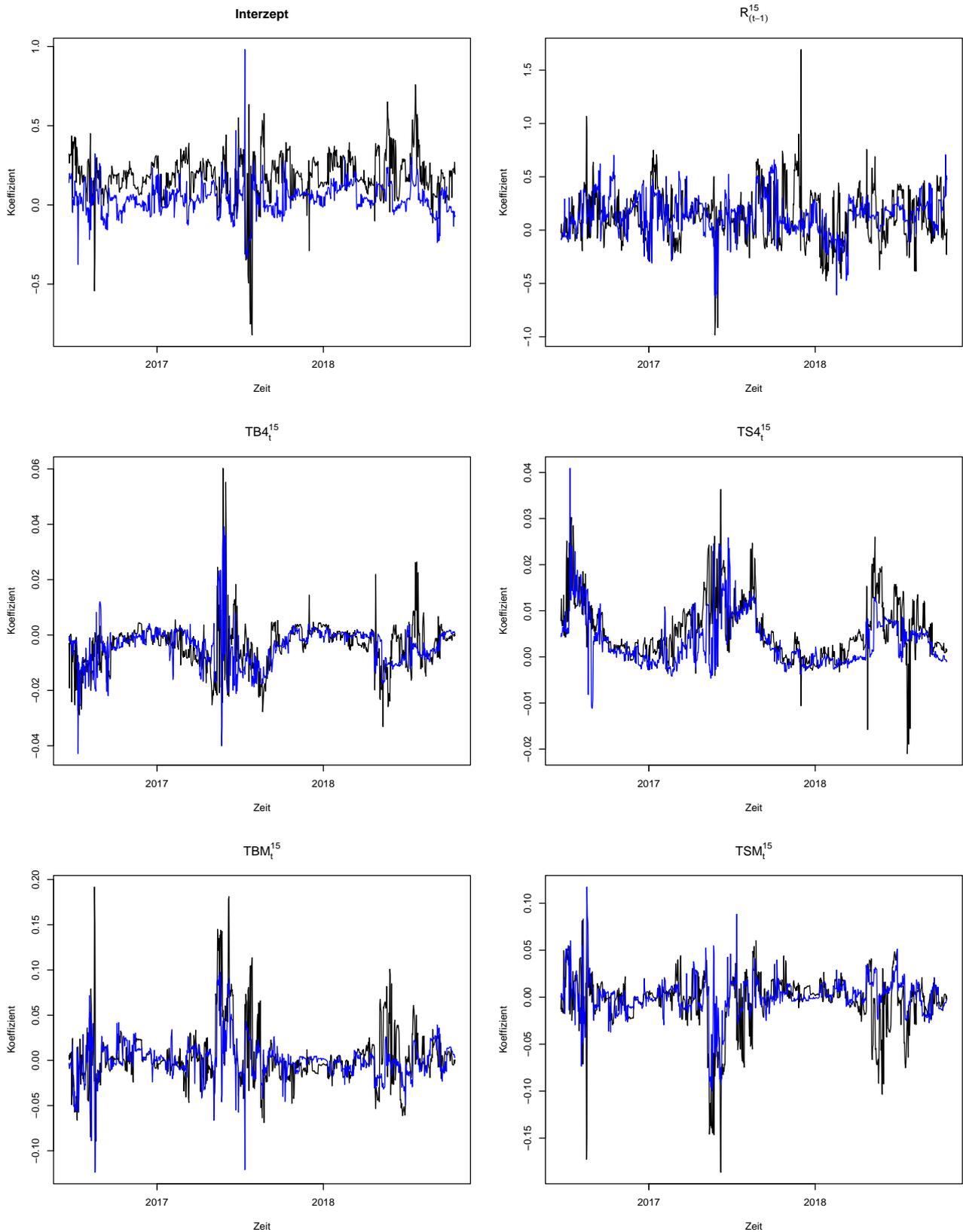


Abbildung 52: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 14:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 15:30 Uhr

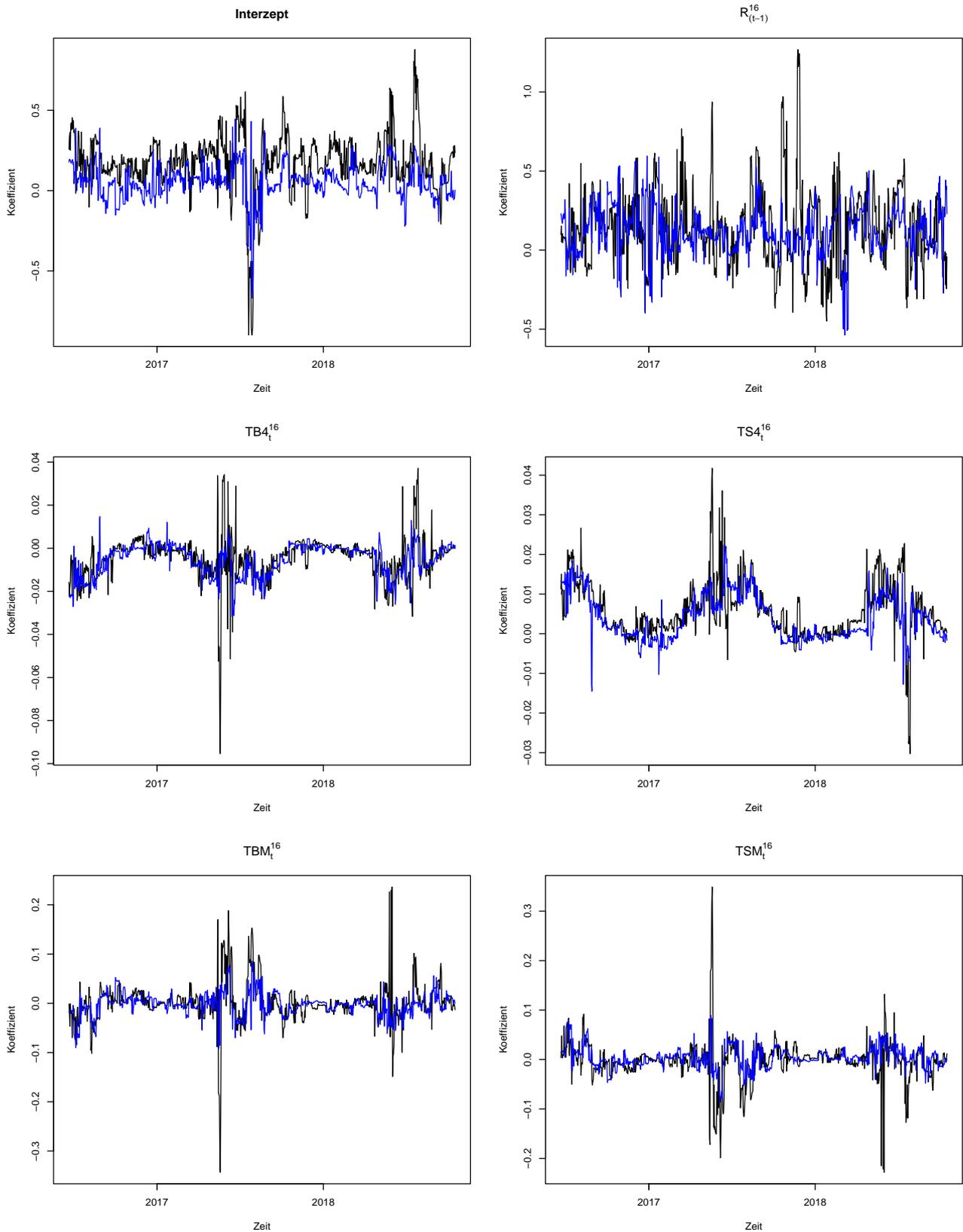


Abbildung 53: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 15:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 16:30 Uhr

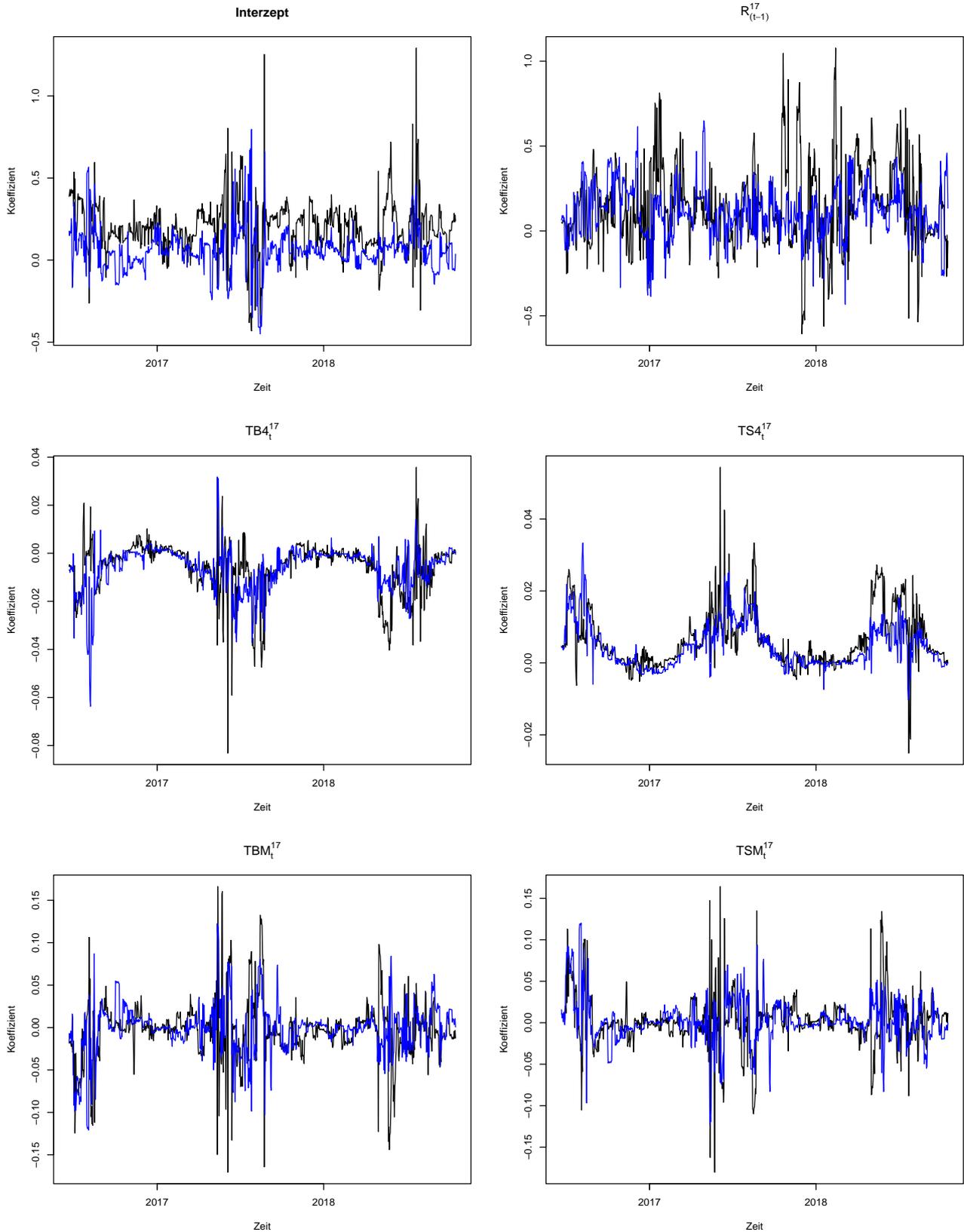


Abbildung 54: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 16:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 17:30 Uhr

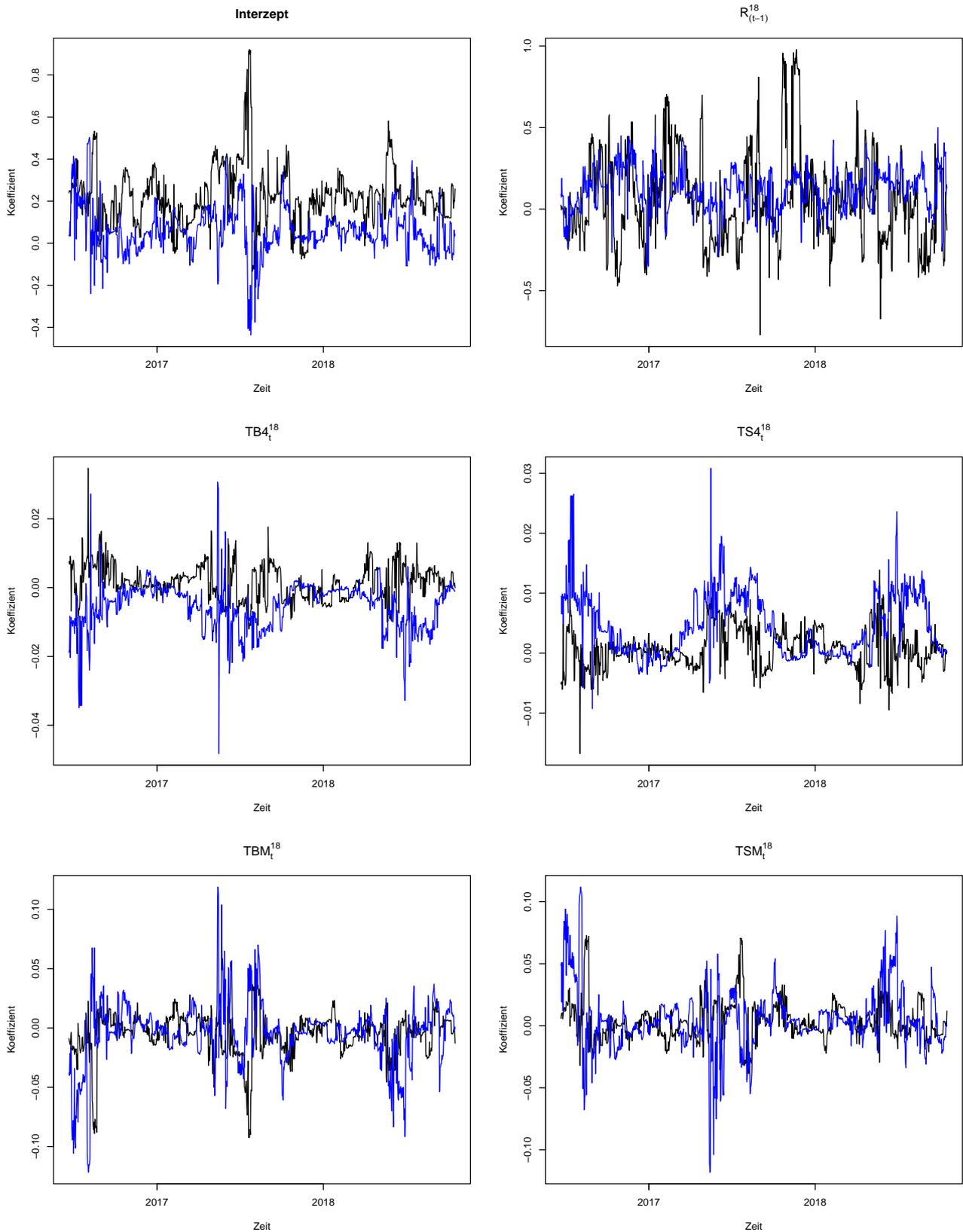


Abbildung 55: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 17:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 18:30 Uhr

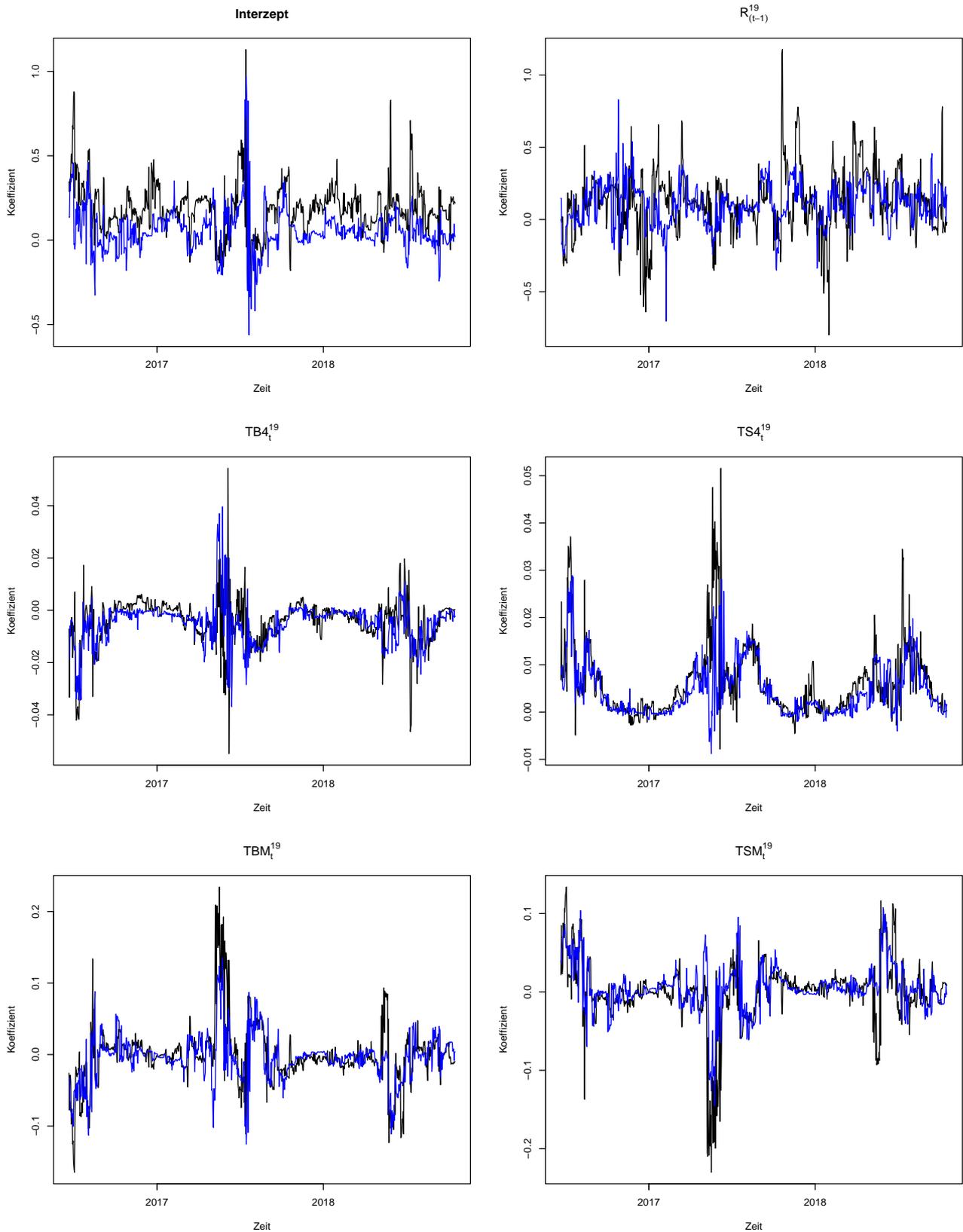


Abbildung 56: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 18:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 19:30 Uhr

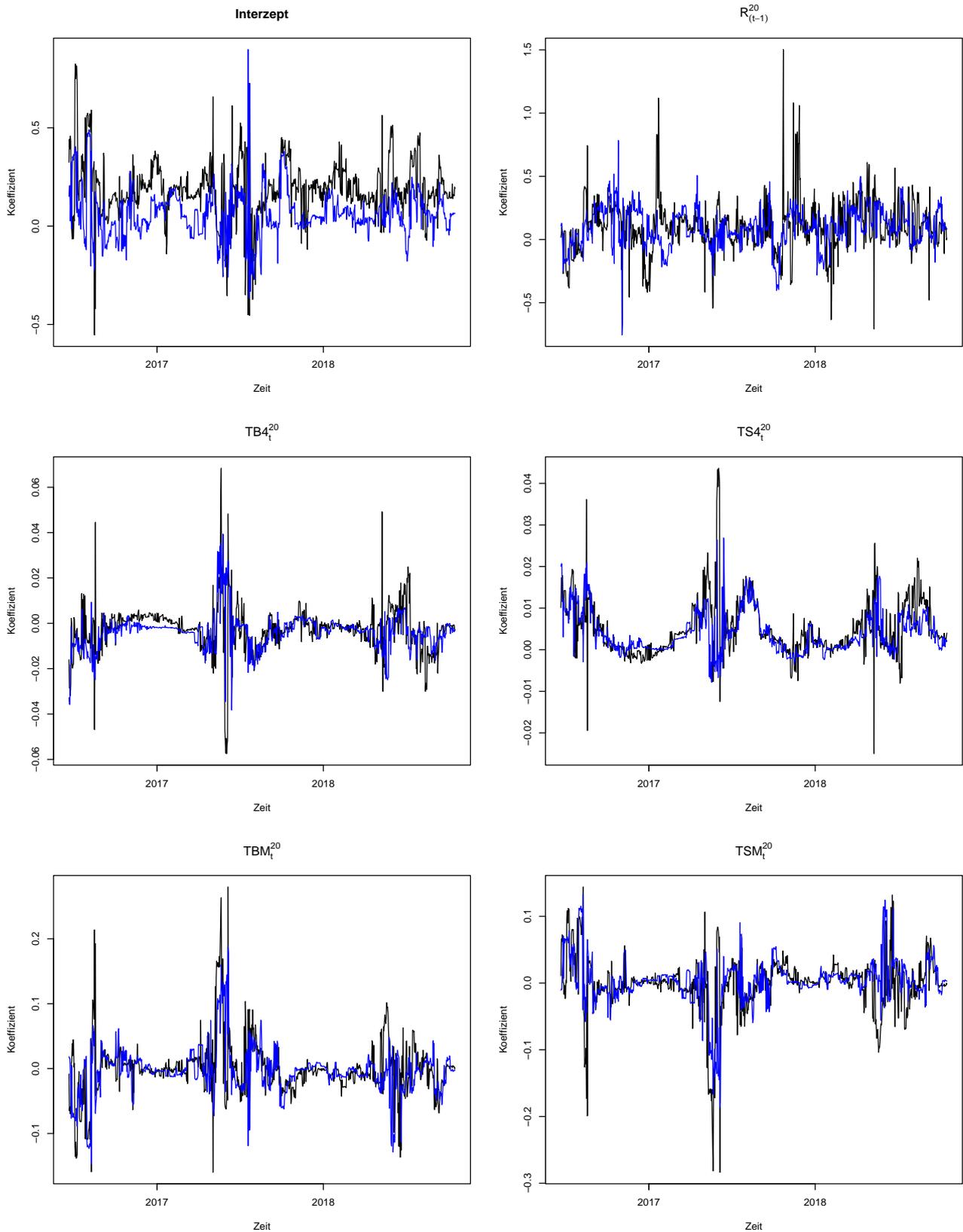


Abbildung 57: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 19:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 20:30 Uhr

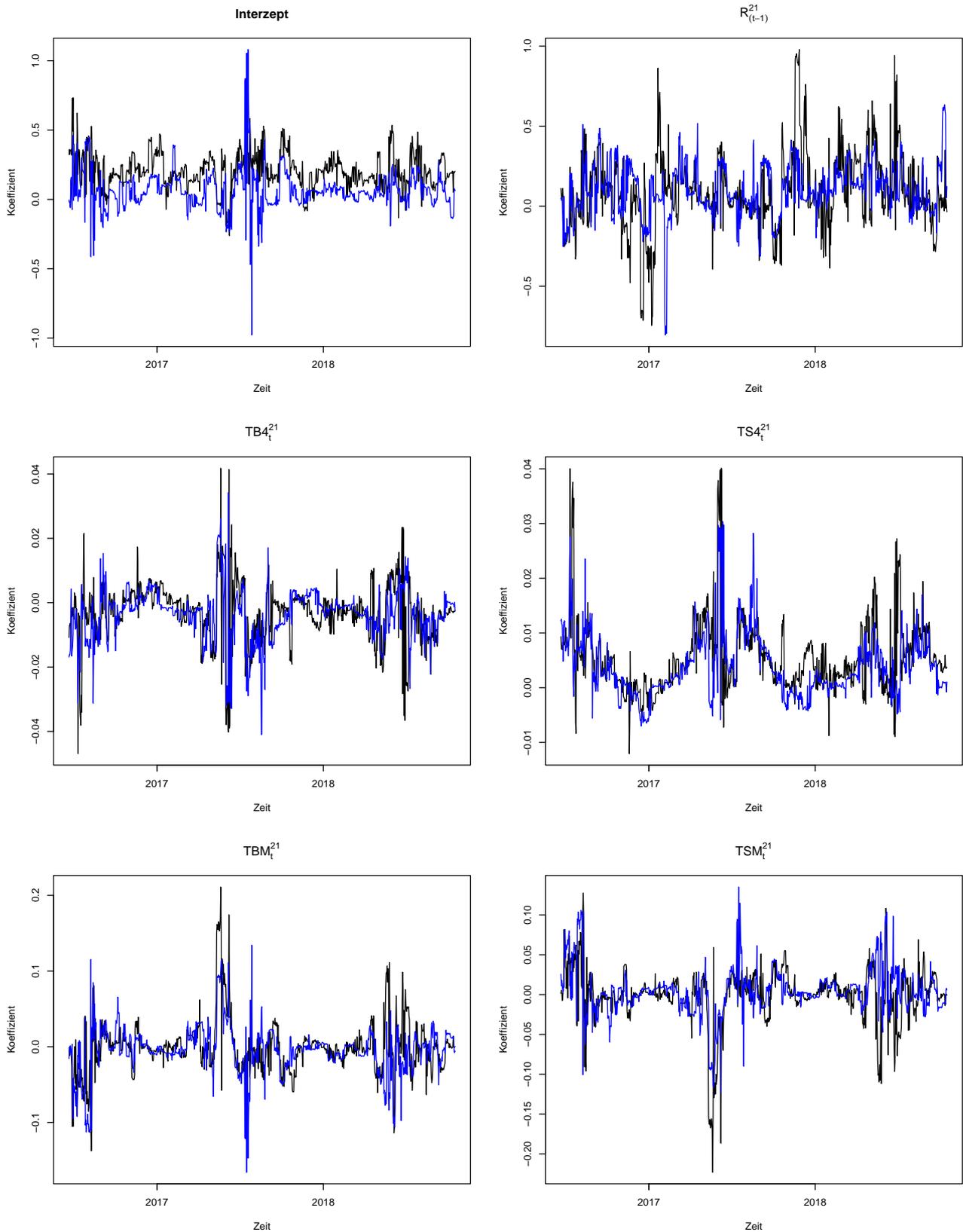


Abbildung 58: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 20:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 21:30 Uhr

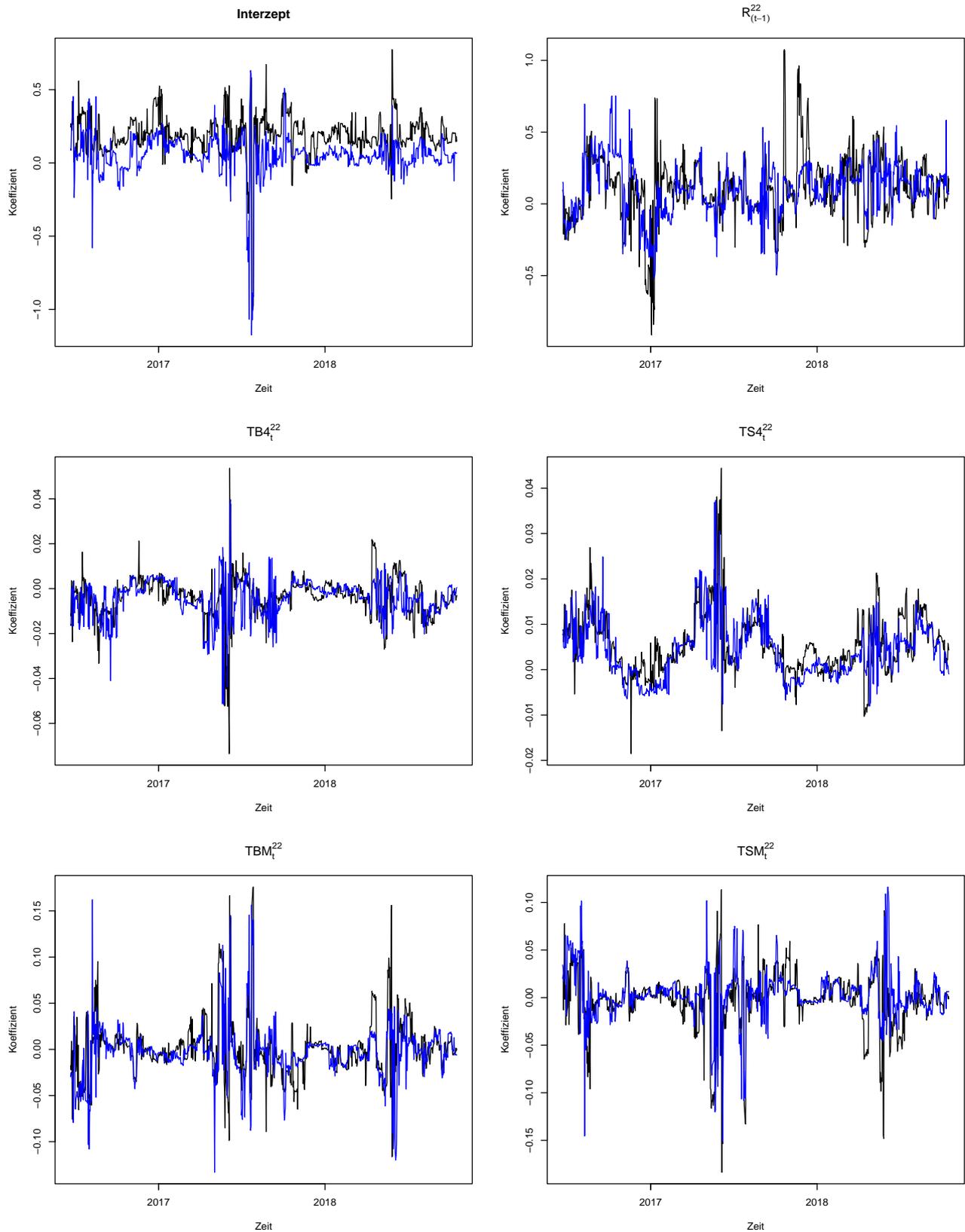


Abbildung 59: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 21:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 22:30 Uhr

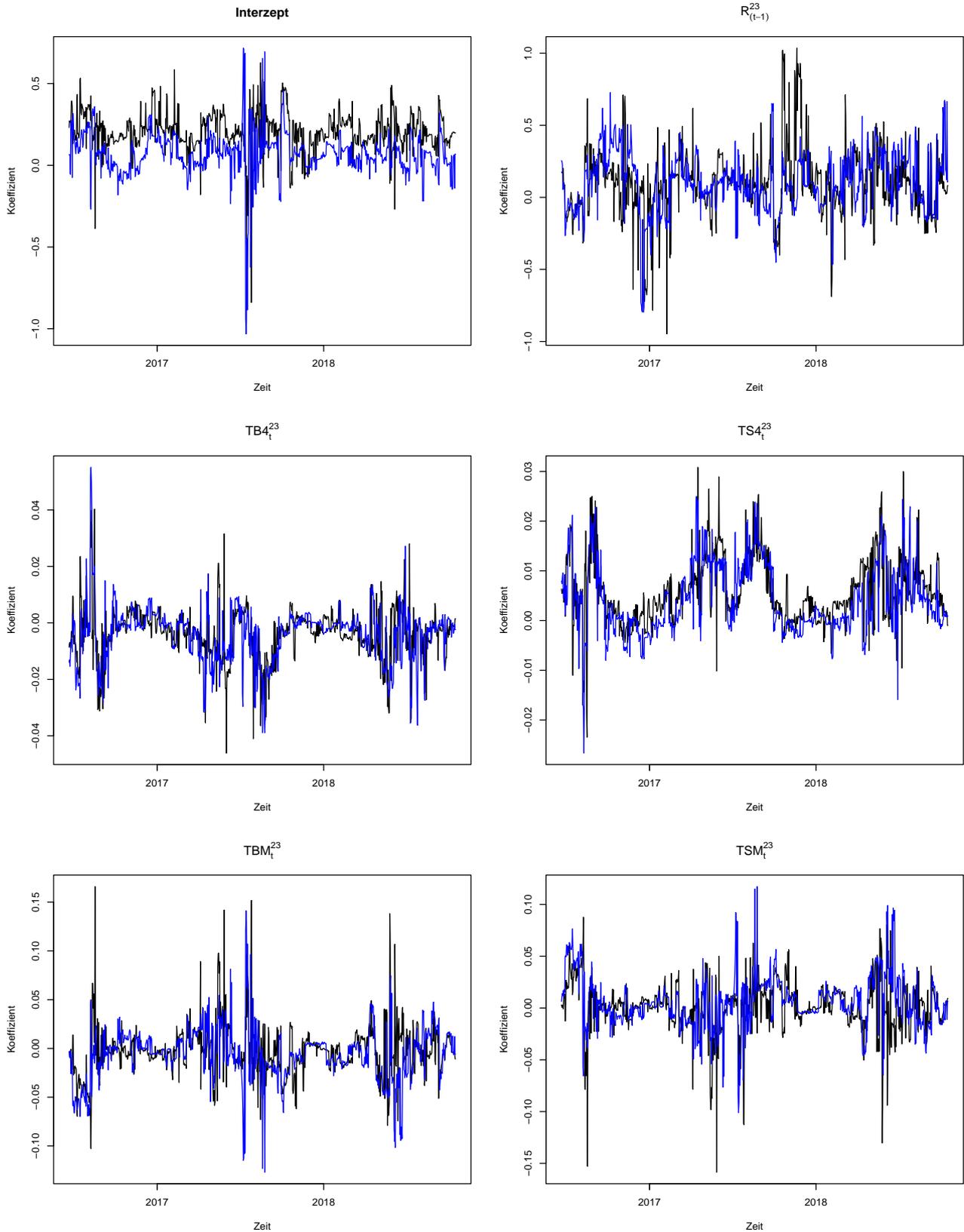


Abbildung 60: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 22:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# 23:30 Uhr

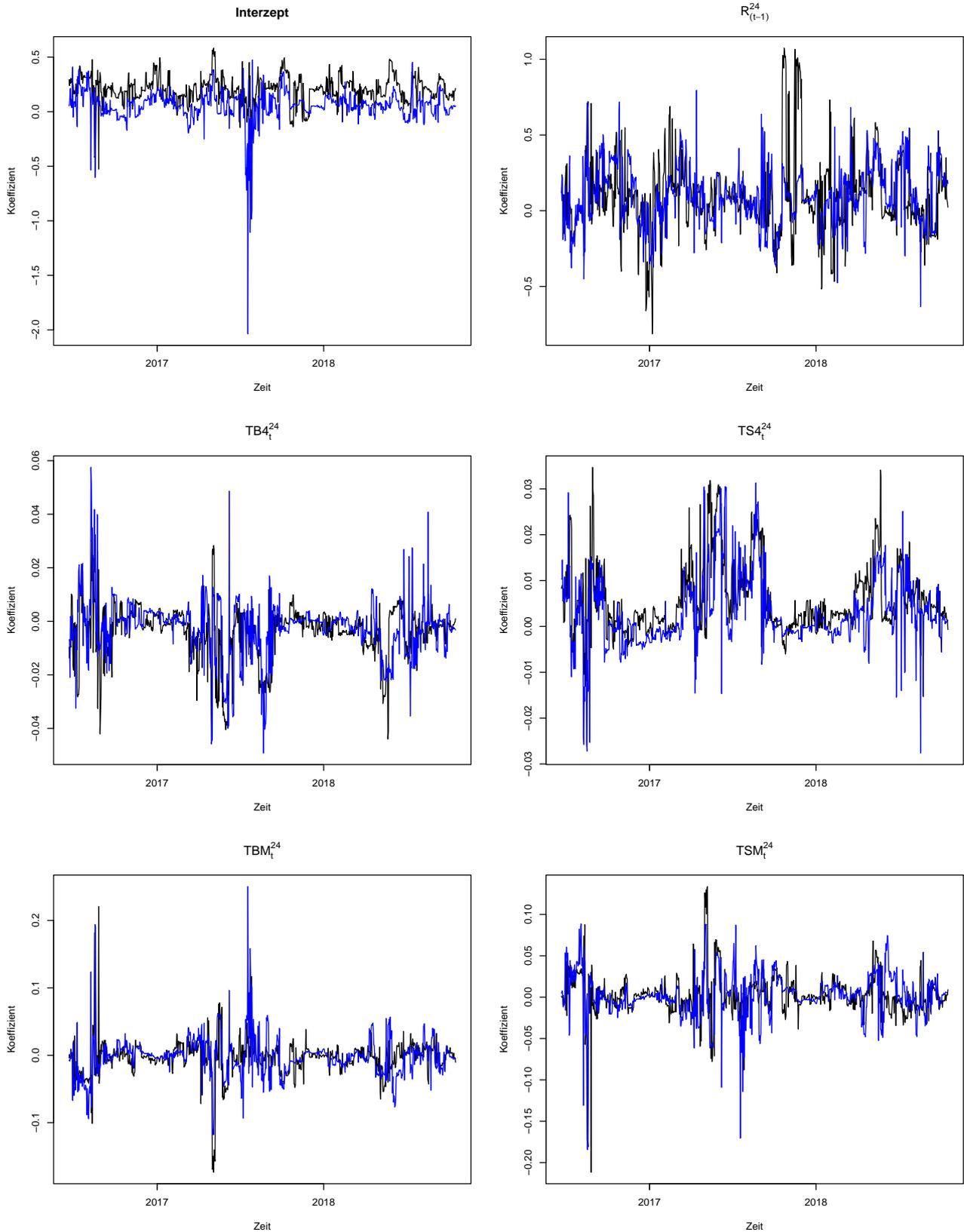


Abbildung 61: robuste Koeffizientenschätzungen der Multivariaten Linearen Modelle für den täglichen Messzeitpunkt 23:30 Uhr über den gesamten Zeitraum, wobei die schwarze Linie die geschätzten Koeffizienten für Riss 1 und die blaue Linie die geschätzten Koeffizienten für Riss 2 darstellt

# robustes Gütekriterium basierend auf der MVE-Schätzung

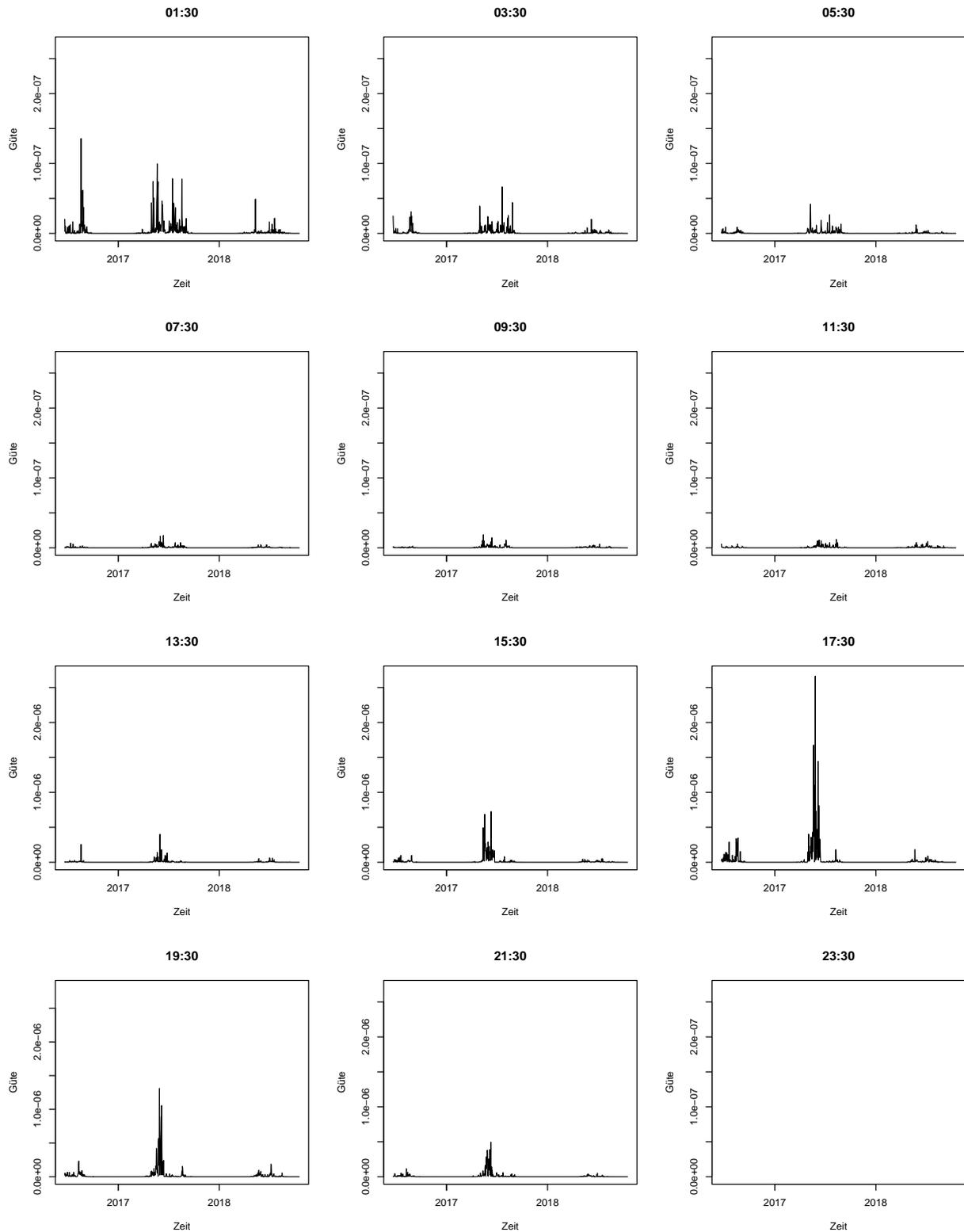


Abbildung 62: Gütekriterium basierend auf der MVE-Schätzung zur Ermittlung der Modellanpassung eines robusten Multivariaten Linearen Modells zur Vorhersage von Rissen im Baumaterial einer Brücke

# robustes Gütekriterium basierend auf der MCD-Schätzung

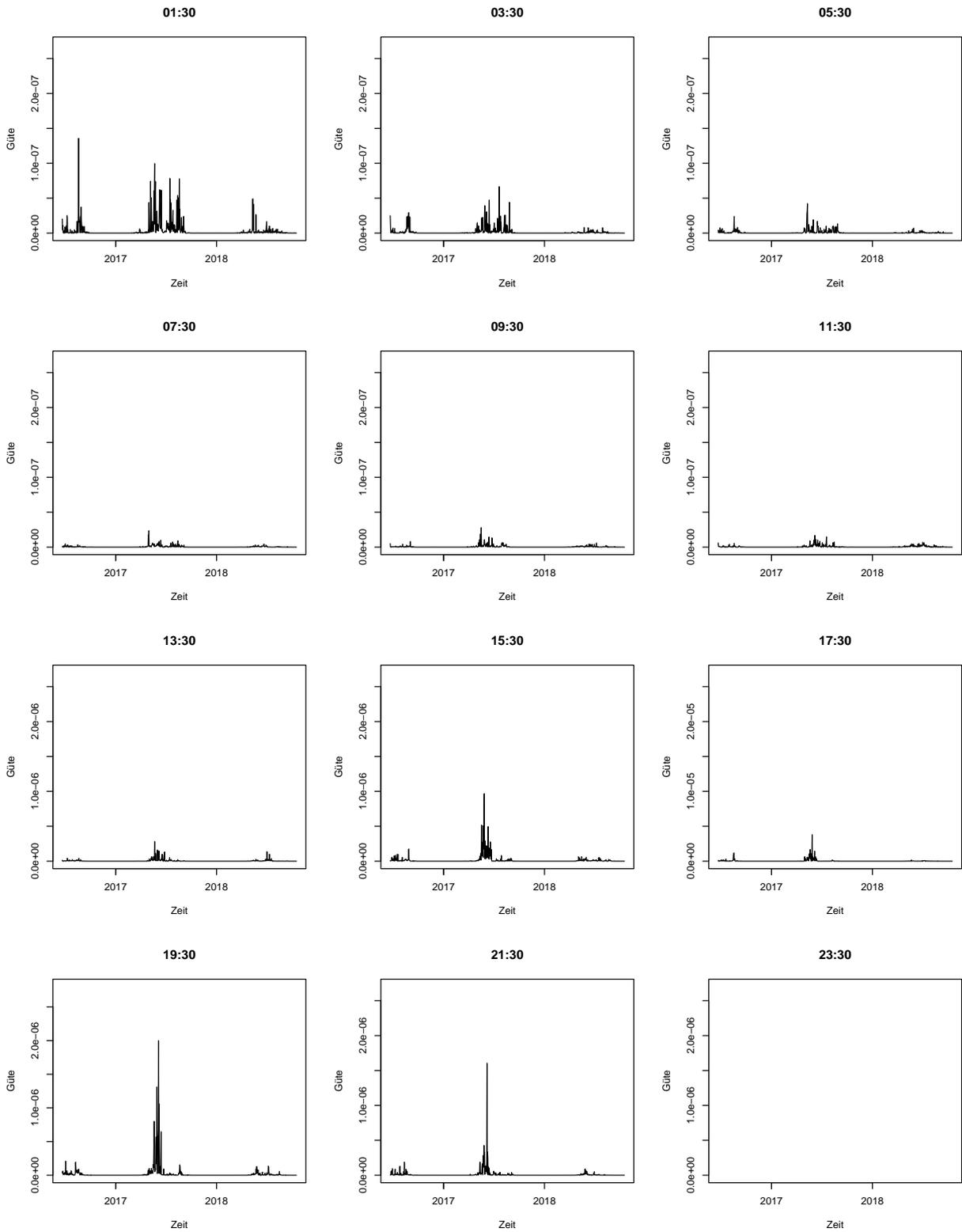


Abbildung 63: Gütekriterium basierend auf der MCD-Schätzung zur Ermittlung der Modellanpassung eines robusten Multivariaten Linearen Modells zur Vorhersage von Rissen im Baumaterial einer Brücke