2023

Leitfaden - Regressionsanalyse

INSTITUT FÜR BETRIEBLICHE FINANZWIRTSCHAFTEN (JKU)
ABTEILUNG FÜR ASSET MANAGEMENT

Inhalt

Grundlegende Informationen zum Leitfaden

Der vorliegende Leitfaden versteht sich als allgemeine Orientierungshilfe zur Vermittlung der grundlegenden Kompetenzen, welche zur Durchführung eigenständiger, empirisch-quantitativer Analysen im Bereich der Finanzforschung benötigt werden. Diesem fundamentalen Ziel entsprechend fokussieren sich die jeweils diskutierten Inhalte daher allen voran auf das Gebiet der (angewandten) Regressionsanalyse, welches häufig als Grundlage für die Durchführung (fortgeschrittener) finanzmarktökonometrischer Analysen herangezogen wird. Im Spezifischen werden in diesem Sinne die folgenden, übergeordneten Inhalte diskutiert:

- Basics der Regressionsanalyse
- Arten von Regressionsmodellen
- Allgemeines zu Schätzwerten & -verfahren
- Wichtige Kennzahlen zur Interpretation allenfalls erhaltener Ergebnisse
- Darstellung sowie Auswahl von Modelselektionsverfahren
- Notwendigkeit der Durchführung von Robustheitschecks
- Weiterführende Prognoseverfahren
- Basis der Zeitreihenanalyse
- Relevanz des Konzeptes der „Stationarität“
- Auswahl alternativer Modellierungsvarianten

Ergänzend zu den oberhalb angeführten, theoretischen Konzepterläuterungen ist dem vorliegenden Leitfaden darüber hinaus ein unmittelbares, kontextumfassendes Anwendungsbeispiel hinzugefügt worden, welches es den Studierenden ermöglichen soll, die jeweils besprochenen Inhalte (insbesondere hinsichtlich deren jeweiliger Einsatzmöglichkeiten sowie Interpretationsansätze) auch innerhalb eines konkreten Anwendungsbeispiels direkt verstehen zu lernen.

Basics der Regressionsanalyse

Innerhalb des ersten, größeren Subparts des Leitfadens sollen die Studierenden in einem ersten Schritt mit den relevantesten Konzepten statistischer Lehre vertraut gemacht werden, die es zur Durchführung wie auch zum Verständnis der Qualität allfällig durchgeführter empirischer Analysen bedarf. In diesem Sinne umfassen die folgenden Seiten insbesondere allgemeine Definitionen zu unterschiedlichen statistischen Konzepten, wobei ein besonderer Fokus auf das Gebiet der Regressionsanalyse gelegt wird.

Fundamentale Basics

Population vs. Stichprobe

Eine Population (N) bezieht sich auf die gesamte Gruppe von Individuen, Elementen oder Einheiten, die das Interesse einer Studie ausmachen. Eine Stichprobe ($n \subseteq N$) ist eine Teilmenge dieser Population, die ausgewählt wird, um repräsentative Informationen zu gewinnen, ohne dass die gesamte Population untersucht werden muss, wobei die Ergebnisse der Stichprobe wiederum auf die Gesamtpopulation verallgemeinert werden können.

Hypothesentests

Hypothesentests sind statistische Verfahren, die verwendet werden, um Annahmen über Populationsparameter auf Grundlage von Stichprobendaten zu überprüfen. In einem Hypothesentest wird eine Nullhypothese (H_0) aufgestellt, die besagt, dass es keinen signifikanten Effekt oder Unterschied gibt, und eine Alternativhypothese (H_1), die einen Effekt oder Unterschied behauptet. Einseitige Tests prüfen, ob der Effekt in eine bestimmte Richtung geht, während beidseitige Tests auf signifikante Unterschiede in beide Richtungen prüfen. Das Testergebnis wird anhand eines Signifikanzniveaus (α) interpretiert, wobei eine Ablehnung der Nullhypothese darauf hinweist, dass die Stichprobendaten signifikante Evidenz für die anzunehmende Gültigkeit der Alternative liefern.

Fehler 1. & 2. Art

Fehler erster Art, auch als „Alpha-Fehler“ bezeichnet, treten auf, wenn eine statistische Nullhypothese fälschlicherweise abgelehnt wird, obwohl sie tatsächlich wahr ist. Fehler zweiter Art, auch „Beta-Fehler“ genannt, treten auf, wenn eine Nullhypothese fälschlicherweise akzeptiert wird, obwohl sie tatsächlich falsch ist. Die Kontrolle über diese Fehler hängt oft vom gewählten Signifikanzniveau (α) und der Stichprobengröße (n) ab und beeinflusst die Zuverlässigkeit von statistischen Tests. Darüber hinaus gilt es zu beachten, dass eine Wechselwirkung zwischen Fehlern erster und zweiter Art besteht, als dass eine Verringerung des Fehlers erster Art (durch Senkung des Signifikanzniveaus) oft zu einer Zunahme des Fehlers zweiter Art führt, und umgekehrt, wodurch wiederum eine Abwägung zwischen beiden Fehlertypen erforderlich ist.

		Realer Zusammenhang in Population	
		H_0 *wäre richtig*	H_0 *wäre falsch*
Entscheidung auf Basis der Stichprobenergebnisse	H_0 *wird akzeptiert*	Richtige Entscheidung	„Beta-Fehler“
	H_0 *wird nicht akzeptiert*	„Alpha-Fehler“	Richtige Entscheidung

Kritische Werte, Signifikanz- & Konfidenzniveau

Kritische Werte (z) sind bestimmte Schwellenwerte in statistischen Tests, die verwendet werden, um zu entscheiden, ob eine Hypothese abgelehnt werden sollte. Das Signifikanzniveau, oft als Alpha (α) bezeichnet, legt die Wahrscheinlichkeit fest, mit der man bereit ist, einen Fehler erster Art (fälschlicherweise Ablehnung einer wahren Hypothese) in Kauf zu nehmen. Die Wahl des Signifikanzniveaus beeinflusst die Position der kritischen Werte und damit die Balance zwischen der Vermeidung von Fehlern erster Art und Fehlern zweiter Art in einem statistischen Test. Das Konfidenzniveau hingegen ist das Maß für die Zuversichtlichkeit, mit der wir annehmen können, dass ein statistischer Schätzwert oder ein Hypothesentest korrekt ist. Es steht in umgekehrtem Verhältnis zum Signifikanzniveau (→ $Konfidenzniveau = 1 - \alpha$): Ein höheres Konfidenzniveau führt zu einem niedrigeren Signifikanzniveau und umgekehrt, wodurch die Balance zwischen Präzision und Fehlerkontrolle in statistischen Analysen festgelegt wird.

Regressionsanalysen

Regressionsanalysen sind statistische Verfahren, die verwendet werden, um die Beziehung zwischen einer abhängigen Variable (Y) und einer oder mehreren unabhängigen Variablen (X_j) zu untersuchen und zu modellieren. Sie helfen dabei, Muster, Trends und potenzielle Zusammenhänge in den Daten zu identifizieren sowie Vorhersagen für zukünftige Werte der abhängigen Variable zu treffen. Im Kontext von Hypothesentests ermöglichen Regressionsanalysen die Überprüfung von Hypothesen über die Stärke und Richtung der Beziehungen zwischen den Variablen, indem sie statistische Tests auf die Koeffizienten des Regressionsmodells anwenden.

Arten von Regressionsmodellen

Lineare vs. Nicht-lineare Regressionsmodelle

Lineare Regressionsmodelle beschreiben eine lineare Beziehung zwischen einer abhängigen Variable (Y) und einer oder mehreren unabhängigen Variablen (X_j), wobei die Änderung der abhängigen Variable proportional zur Änderung der unabhängigen Variablen ist. Nichtlineare Regressionsmodelle hingegen modellieren nicht-lineare Beziehungen zwischen den Variablen, was bedeutet, dass die Änderung der abhängigen Variable nicht proportional zur Änderung der unabhängigen Variablen ist und die Beziehung komplexere Formen annehmen kann, wie z.B. exponentiell, logarithmisch oder polynomial.

Einfache vs. Multiple Regressionsmodelle

Einfache Regressionsmodelle untersuchen die Beziehung zwischen einer abhängigen Variable und einer einzigen unabhängigen Variablen, um eine lineare Vorhersage zu erstellen. Multiple Regressionsmodelle hingegen berücksichtigen gleichzeitig den Einfluss mehrerer unabhängiger Variablen auf die abhängige Variable, wodurch komplexere Zusammenhänge und Interaktionen erfasst werden können. Die multiple Regression ermöglicht somit eine umfassendere Analyse von Variablen und deren kombiniertem Effekt auf die abhängige Variable.

Einfaches (lineares) Regressionsmodell	$\hat{Y}_i = c + \hat{\beta} X_i$
Multiples (lineares) Regressionsmodell	$\hat{Y}_i = c + \hat{\beta}_1 X_{i,1} + \hat{\beta}_2 X_{i,2} + \cdots + \hat{\beta}_p X_{i,h} = c + \sum_{j=1}^{h} \hat{\beta}_j X_{i,j}$

Modellierung angenommener Variablenverhältnisse

Das (linear-) lineare Modell postuliert eine lineare Beziehung zwischen der abhängigen und unabhängigen Variablen. Das log-lineare Modell verwendet eine lineare Beziehung zwischen der unabhängigen Variable und dem Logarithmus der abhängigen Variable, wodurch (vereinfacht) angenommen wird, dass Veränderungen in der unabhängigen Variable zu prozentuellen Änderungen in der abhängigen Variable führen. Das linear-log Modell wiederum nutzt eine lineare Beziehung zwischen der abhängigen Variable und der unabhängigen Variable im Logarithmus, was bedeutet, dass prozentuelle Veränderungen der unabhängigen Variable eine proportional-reale Veränderung im Gesamtwert der abhängigen Variable verursachen. Das log-log Modell verwendet schlussendlich eine logarithmische Transformation sowohl für die abhängige als auch für die unabhängige Variable und erfasst somit prozentuale Veränderungen beider Variablen in Bezug zueinander.

	X	
Y	X	log(X)
Y	*linear-linear* $\hat{Y}_i = c + \hat{\beta} X_i$	*linear-log* $\hat{Y}_i = c + \hat{\beta} * \log(X_i)$
log(Y)	*log-linear* $\log(\hat{Y}_i) = c + \hat{\beta} X_i$	*log-log* $\log(\hat{Y}_i) = c + \hat{\beta} * \log(X_i)$

Schätzwerte und -verfahren

Unterschied Schätzwerte und Reale Werte

Schätzwerte in Regressionsanalysen ($\hat{Y}$) sind vorhergesagte Werte für die abhängige Variable auf Grundlage des erstellten Regressionsmodells und der unabhängigen Variablen und repräsentieren die theoretischen Werte, die das Modell erwartet. Reale Werte für die abhängige Variable (Y) wiederum repräsentieren demgegenüber die tatsächlich beobachteten Werte in der Stichprobe oder Population. Der Unterschied zwischen Schätz- und Realwerten wird oft in Form von Residualwerten ($\epsilon = Y - \hat{Y}$) dargestellt, welche die Abweichung zwischen den geschätzten und beobachteten Werten messen und Informationen über die Genauigkeit des Regressionsmodells liefern. Residuen sind die vertikalen Abstände zwischen den tatsächlichen Datenpunkten und der Regressionsgeraden bzw. -kurve und sollten zufällig verteilt und möglichst nah an null sein, um ein gutes Modell zu bestätigen.

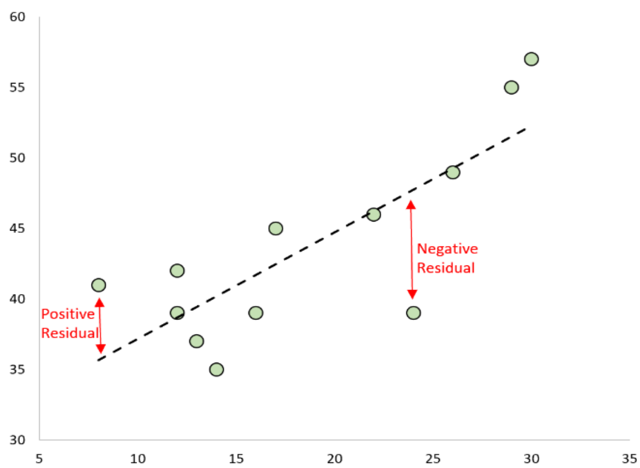

Abbildung 1: Ermittlung der einzelnen Residualwerte als Abstand der realen Messpunkte von der geschätzten Regressionslinie

Kleinste Quadrate Schätzer („*Oridnary Least Squares Estimation*")

Das bekannteste Verfahren zur Bestimmung der Verhältniskoeffizienten eines Regressionsmodelles ist die Methode der kleinsten Quadrate („*Ordinary Least Squares Estimation*"), welche darauf abzielt, die Parameter ($\hat{\beta}_j$) eines Regressionsmodells so zu schätzen, dass die Summe der quadrierten Residuen (das sind die Differenzen zwischen den beobachteten und vorhergesagten Werten der abhängigen Variable) minimiert wird. Dies bedeutet, dass das Modell die beste Anpassung an die vorliegenden Daten findet, indem es die Abweichungen zwischen den Datenpunkten und der Regressionslinie minimiert. Die Quadrierung der Residuen in der Methode der kleinsten Quadrate wird hierbei insbesondere verwendet, um sicherzustellen, dass sowohl positive als auch negative Abweichungen zwischen den beobachteten und vorhergesagten Werten gleichermaßen zur Gesamtabweichung beitragen, da sich solche bei einer alternativ einfachen Summation potentiell aufheben könnten und somit wichtige Informationen über die Fehler im Modell verloren gehen würden.

Maximum-Likelihood-Schätzer

Die Maximum-Likelihood-Schätzmethode in Regressionsanalysen zielt darauf ab, die Parameter des Modells so zu schätzen, dass die beobachteten Daten, unter der Annahme einer bestimmten Verteilung der Fehler, am wahrscheinlichsten sind. Sie sucht ergo nach denjenigen Werten der Regressionskoeffizienten, die die Wahrscheinlichkeit maximieren, die tatsächlichen Beobachtungen basierend auf dem Modell zu beobachten. Zu diesem Zweck werden komplexe Verfahren der Wahrscheinlichkeitsrechnung verwendet, die oftmals im Zuge fortgeschrittenerer Anwendungsfälle anlassbezogen modifiziert werden können.

Wichtige Interpretationskennzahlen

Regressionskoeffizienten

Regressionskoeffizienten (β_j) sind numerische Werte in Regressionsmodellen, die die Stärke und Richtung der Beziehung zwischen unabhängigen und abhängigen Variablen quantifizieren. Jeder Regressionskoeffizient repräsentiert den durchschnittlichen Effekt einer Einheitenerhöhung in der entsprechenden unabhängigen Variable auf die abhängige Variable, unter der Annahme, dass alle anderen unabhängigen Variablen konstant gehalten werden. Sie dienen dazu, die Beziehung zwischen den Variablen zu verstehen, Vorhersagen zu treffen und Hypothesen über den Einfluss der unabhängigen Variablen zu testen.

Konstante

Die Konstante (c) in einem Regressionsmodell, auch als „y-Achsenabschnitt“ oder „*Intercept*“ bezeichnet, repräsentiert den geschätzten Wert der abhängigen Variable, wenn alle unabhängigen Variablen gleich Null sind. Sie gibt den Anfangspunkt der Regressionslinie an der y-Achse an und zeigt, wie sich die abhängige Variable verhält, wenn keine unabhängigen Variablen berücksichtigt werden. Die Konstante ist ein wichtiger Teil des Modells, da sie den Basiswert darstellt, von dem aus Änderungen in den unabhängigen Variablen ihren Einfluss auf die abhängige Variable entfalten.

t-Teststatistik (für Regressionskoeffizienten)

Die t-Teststatistik für Regressionskoeffizienten wird verwendet, um zu überprüfen, ob die geschätzten Koeffizienten signifikant von null verschieden sind, was darauf hinweisen würde, dass die jeweiligen, unabhängigen Variablen einen statistisch bedeutsamen Einfluss auf die abhängige Variable haben. Die Teststatistik berechnet das Verhältnis der geschätzten Koeffizienten zu ihrer Standardfehlerabschätzung und vergleicht es mit einer „Student t-Verteilung“, um festzustellen, ob der geschätzte Koeffizient statistisch signifikant ist. Ein t-Wert, der weit genug von null entfernt ist, führt zu einer Ablehnung der Nullhypothese und zeigt an, dass die zugehörige unabhängige Variable einen signifikanten Einfluss auf die abhängige Variable hat.

p-Wert (für Regressionskoeffizienten)

Quantifizierte p-Werte für Regressionskoeffizienten sind statistische Werte, die anzeigen, wie wahrscheinlich es ist, den geschätzten Wert eines Regressionskoeffizienten zu erhalten, wenn die Nullhypothese wahr wäre (d.h. wenn kein Effekt der unabhängigen Variable anzunehmen wäre). Ein niedriger p-Wert deutet darauf hin, dass der geschätzte Koeffizient signifikant von null abweicht, was wiederum darauf hindeutet, dass die unabhängige Variable einen statistisch bedeutsamen Einfluss auf die abhängige Variable hat. Ein p -Wert unter einem vorher festgelegten Signifikanzniveau (z.B. 0,05) führt zur Ablehnung der Nullhypothese und zeigt an, dass der Koeffizient statistisch signifikant ist.

(Adjustiertes) Bestimmtheitsmaß (R^2)

Das Bestimmtheitsmaß, auch als „R-Quadrat“ (R^2) bezeichnet, gibt an, wie gut die unabhängigen Variablen in einem Regressionsmodell die Variation in der abhängigen Variable erklären. Ein höheres R^2 bedeutet, dass eine größere proportionale Varianz der abhängigen Variable durch die unabhängigen Variablen erklärt wird. Das adjustierte Bestimmtheitsmaß (adjustiertes R^2) ist eine angepasste Version von R^2, welches die Anzahl der in das Modell aufgenommenen unabhängigen Variablen berücksichtigt und daher hilfreich ist, um die Modellkomplexität zu bewerten und sicherzustellen, dass zusätzliche Variablen tatsächlich zur Verbesserung der Modellgenauigkeit beitragen.

F-Teststatistik

Die F-Teststatistik wird im Kontext von Regressionsmodellen dazu verwendet, um die Gesamthypothese zu prüfen, ob das gesamte Regressionsmodell statistisch signifikant ist, das heißt, ob die unabhängigen Variablen als Ganzes einen Einfluss auf die abhängige Variable haben oder ob diese gesamtheitlich irrelevant sind. Sie vergleicht in diesem Sinne die Varianz zwischen den vorhergesagten Werten des Modells und der durchschnittlichen abhängigen Variable mit der Varianz allein aufgrund des Zufalls. Ein signifikanter F-Test weist darauf hin, dass zumindest eine der unabhängigen Variablen im Modell einen statistisch bedeutsamen Beitrag zur Erklärung der Variation der abhängigen Variable leistet.

Modelselektionsverfahren

Faktorkorrelationstests

Eine initiale Möglichkeit zur Bestimmung, welche Variablenzusammenhänge innerhalb einer Regressionsanalyse potentiell überprüft werden sollen, bietet die Prüfung allfälliger Beziehungen mittels Faktorkorrelationstests. Grundsätzlich dienen entsprechende Korrelationstests zwischen abhängigen und unabhängigen Variablen hierbei dazu, anfänglich zu prüfen, ob es eine statistische Beziehung zwischen diesen Variablen gibt, wobei ein signifikanter Korrelationskoeffizient darauf hinweisen kann, dass eine lineare Beziehung zwischen ihnen besteht, was wiederum eine mögliche Grundlage für ein Regressionsmodell darstellt. Interpretationsbezogen dient dieser Test allerdings nur als erster Hinweis betreffend die mögliche Eignung einer Regressionsanalyse, um den Einfluss der unabhängigen Variable auf die abhängige Variable zu untersuchen, und sollte im Anschluss durch weiterführende Selektionsevaluationsverfahren (wie etwa Informationskriterientests) ergänzt werden.

Informationskriterien

Informationskriterien (ICs) sind statistische Bewertungswerkzeuge, die in der Modellselektion verwendet werden, um sinnvolle, unabhängige Variablen in Regressionsmodellen zu identifizieren. Sie ermöglichen eine Abwägung zwischen Modellanpassung und Modellkomplexität, indem sie Strafen für zusätzliche unabhängige Variablen einführen. Interpretationsbezogen gilt für die meisten

Informationskriterien, dass ein niedrigerer IC-Wert anzeigt, dass das Modell besser abschneidet, indem es die Daten genauer erklärt, ohne zu komplex zu sein, was wiederum darauf hinweist, dass die ausgewählten unabhängigen Variablen sinnvoll sind. Bekannte Beispiele für häufig verwendete Informationskriterien sind z.B. das Akaike-Informationskriterium (*AIC*), das Bayes- (bzw. Schwarz-) Informationskriterium (*BIC*) sowie das Hannan-Quinn-Informationskriterium (*HQIC*), wobei das *AIC* eine leichtere Tendenz zur Modellkomplexität aufweist, das *BIC* eine stärkere Abneigung gegenüber Komplexität beinhaltet, und das *HQIC* eine Mittelposition zwischen den beiden einnimmt.

Robustheitschecks

Notwendigkeit der Durchführung von Robustheitschecks

Die Durchführung von Robustheitschecks nach initialen Regressionsanalysen ist notwendig, um die Stabilität und Verlässlichkeit der Ergebnisse sicherzustellen. Dies hilft dabei, mögliche Einflüsse von Ausreißern sowie ungewöhnlichen Datenmustern oder Verletzungen der Modellannahmen zu erkennen und sicherzustellen, dass die Schlussfolgerungen der Analyse widerstandsfähig gegenüber solchen Anomalien sind.

Normalverteilung der Fehlerterme

In Regressionsanalysen wird oft die Annahme vertreten, dass die Fehlerterme (= Residuen), also die Abweichungen zwischen den beobachteten Werten und den vorhergesagten Werten der abhängigen Variable, normalverteilt sind. Diese Annahme ermöglicht es, statistische Schlussfolgerungen zu ziehen und Hypothesentests durchzuführen. Eine Verletzung dieser Annahme kann die Gültigkeit der Schätzungen (inklusive der verwendeten Schätzverfahren) sowie statistischen Tests beeinflussen, insbesondere bei kleineren Stichproben, weshalb Tests auf die Normalverteilung der Fehlerterme zu den grundlegenden Testverfahren zählen, die im Kontext der Überprüfung der Robustheit der initial erhaltenen Ergebnisse durchzuführen sind.

Multikollinearität

Multikollinearität in Regressionsanalysen tritt auf, wenn zwei oder mehrere unabhängige Variablen in einem Modell stark miteinander korreliert sind, was dazu führt, dass es schwierig ist, den individuellen Einfluss jeder Variablen auf die abhängige Variable zu unterscheiden. Dies kann die Stabilität der Regressionskoeffizienten beeinträchtigen und die Interpretation der Ergebnisse erschweren. Tests auf Multikollinearität sind wichtig für Robustheitstests, da sie sicherstellen, dass die unabhängigen Variablen im Modell tatsächlich unabhängige Beiträge zur Vorhersage der abhängigen Variable leisten, und helfen, mögliche Probleme der Mehrdeutigkeit in den Koeffizienten zu identifizieren und zu adressieren.

Robuste Schätzer

Robuste Schätzer in Regressionsanalysen sind eine Methode, um die Parameter eines Modells so zu schätzen, dass sie widerstandsfähig gegenüber Abweichungen von den Annahmen der klassischen linearen Regression sind. Sie sind insbesondere dann nützlich, wenn Probleme wie Autokorrelation der Fehlerterme und Heteroskedastizität in den Residuen vorliegen. Autokorrelation der Fehlerterme tritt auf, wenn die Residuen in einem Zeitreihenmodell oder einem Paneldatenmodell nicht unabhängig voneinander sind, sondern eine systematische Abhängigkeit aufweisen, was die Schätzungen der Regressionskoeffizienten beeinflussen kann. Heteroskedastizität bedeutet, dass die Varianz der Fehlerterme nicht konstant ist, sondern sich mit den Werten der unabhängigen Variablen ändert, was die Annahme homoskedastischer Fehler verletzt und die Effizienz der Schätzungen beeinträchtigt. Robuste

Schätzungen sind darauf ausgelegt, diese Probleme zu bewältigen und genaue Schätzungen der Regressionskoeffizienten zu liefern, auch wenn die Annahmen der klassischen linearen Regression verletzt sind. Dies erhöht die Zuverlässigkeit der Regressionsanalysen und ermöglicht belastbarere Schlussfolgerungen.

Omitted Variable Bias

Der *„Omitted Variable Bias"* tritt auf, wenn wichtige unabhängige Variablen in einem Regressionsmodell ausgelassen werden, obwohl sie tatsächlich Einfluss auf die abhängige Variable haben. Dies führt zu verzerrten Schätzungen der Regressionskoeffizienten und damit zu falschen Schlussfolgerungen über die Beziehungen zwischen den Variablen. Tests auf den *„Omitted Variable Bias"* (z.B. durch Inklusion allfälliger Proxy-Variablen in multiplen Regressionsmodellen) sind wichtig für Robustheitstests, um sicherzustellen, dass das Modell alle relevanten Variablen enthält und die Zuverlässigkeit der Ergebnisse gewährleistet wird.

Prädiktionsverfahren

In-Sample- vs. Out-of-Sample-Analysen

In-Sample-Analysen beziehen sich auf die Bewertung eines Regressionsmodells anhand der gleichen Daten, die zur Modellerstellung verwendet wurden. Diese Methode testet, wie gut das Modell die bekannten Daten erklärt, kann jedoch die Fähigkeit zur Vorhersage neuer, unbekannter Daten überschätzen. Out-of-Sample-Analysen hingegen bewerten die Vorhersageleistung eines Modells anhand von Daten, die es zuvor nicht gesehen hat. Dies stellt sicher, dass das Modell auf neue, unabhängige Daten robuste Vorhersagen treffen kann und hilft, Überanpassung (*„Overfitting"*) an die Trainingsdaten zu vermeiden. Der wesentliche Unterschied liegt in der Prüfung der Modellleistung auf bekannten Daten (In-Sample) gegenüber der Fähigkeit zur Vorhersage zunächst unbekannter Daten (Out-of-Sample), wodurch die Verallgemeinerungsfähigkeit des Modells besser beurteilt werden kann.

Evaluationskriterien

Evaluationskriterien im Kontext von Prognoseverfahren dienen dazu, die Genauigkeit und Leistung eines Modells bei der Vorhersage von Daten zu bewerten. Sie ermöglichen die quantitative Beurteilung, wie gut die Modellvorhersagen mit den tatsächlichen Daten übereinstimmen. Zu den bekanntesten Gütekriterien zählen in dieser Hinsicht der *„Mean Error"* (ME), der *„Mean Average Error"* (MAE) wie auch der *„Root Mean Squared Error"* (RMSE), wobei der ME den durchschnittlichen Vorhersagefehler, der MAE den durchschnittlichen Betrag der Fehler und der RMSE die Wurzel des durchschnittlichen quadratischen Fehlers berechnet (wodurch der RMSE sowohl die Größe als auch die Richtung der Fehler berücksichtigt und dadurch auch als relevantester der genannten Metriken gilt).

Mean Error	$ME = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)}{n}$
Mean Average Error	$MAE = \frac{\sum_{i=1}^{n}\lvert Y_i - \hat{Y}_i\rvert}{n}$
Root Mean Squared Error	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}}$

Spezifikum Zeitreihenanalyse

Innerhalb des zweiten, größeren Subparts des vorliegenden Leitfadens soll, aufgrund der fundamentalen Relevanz dieses Analysegebiets für den Bereich empirisch-quantitativer Untersuchungen im Finanzbereich, ein besonderer Fokus auf das Instrumentarium sowie Anwendungsfeld der Zeitreihenanalyse gelegt werden. Um das inhärente Ziel des Leitfadens im Sinne einer kompakten Übersicht wissenswerter Verfahren zum Zwecke der Durchführung eigenständiger, finanzmarktökonometrischer Untersuchungen jedoch nicht zu verkomplizieren, wird sich im Folgenden insbesondere mit relevanten Aspekten der Zeitreihenanalyse beschäftigt, die mittels der bereits vorgestellten Konzepte der Regressionsanalyse evaluiert werden können.

Fundamentale Basics

Grundlegende Definition

Die Zeitreihenanalyse ist eine statistische Methode, die sich auf die Untersuchung und Modellierung von Daten bezieht, die in zeitlicher Reihenfolge erfasst wurden. Im Allgemeinen befasst sie sich mit der Identifizierung von Mustern, Trends und Abhängigkeiten in den Daten über die Zeit. Im spezifischen Kontext der Regressionsanalyse kann die Zeitreihenanalyse verwendet werden, um den Einfluss von zeitabhängigen Variablen auf eine abhängige Variable zu quantifizieren, wobei die Zeit als eine der unabhängigen Variablen betrachtet wird. Dies ermöglicht es, Beziehungen und Vorhersagen in zeitlichen Datenverläufen zu formulieren. So kann das Instrumentarium der Zeitreihenanalyse in einem Finanzmarktkontext beispielsweise dazu verwendet werden, um historische Preis- und Handelsdaten zu analysieren und damit Vorhersagen über zukünftige Preisbewegungen, Risiken und Portfolio-Optimierungen auf Basis vergangener, bereits beobachteter Bewegungen zu treffen.

Auswahl der zu berücksichtigenden Vergangenheitswerte

Die optimale Auswahl der zu berücksichtigenden, vergangenen Beobachtungspunkte (= Lag-Länge) in zeitbasierten Regressionsmodellen kann durch Verwendung verschiedener Evaluationskriterien ermittelt werden. Eine gängige Methode ist die Verwendung klassischer Informationskriterien wie etwa das Akaike-Informationskriterium (*AIC*) oder das Bayesianischen Informationskriterium (*BIC*), bei denen jene Lag-Länge ausgewählt wird, die das Kriterium minimiert und damit das beste Gleichgewicht zwischen Modellkomplexität und Passung zu den Daten bietet. Alternativ können jedoch auch spezifischere Instrumente wie die Ljung-Box-Teststatistik oder der (partielle) Autokorrelationstest zur Identifikation signifikanter Verzögerungen im Modell herangezogen werden. Diese Verzögerungen können wiederum darauf hinweisen, wie viele Lags in das Modell aufgenommen werden sollten. Die Kombination aus Informationskriterien und Autokorrelationstests kann daher eine solide Grundlage bieten, um die optimale Lag-Länge in zeitlichen Regressionsmodellen zu bestimmen.

Stationarität

Grundidee

Das Konzept der Stationarität von Zeitreihen ist im Kontext von Zeitreihenanalysen von entscheidender Bedeutung, da es sicherstellt, dass statistische Eigenschaften einer Zeitreihe im Laufe der Zeit konstant bleiben. Stationäre Zeitreihen ermöglichen stabilere Modellierung und Vorhersagen, da sie frei von systematischen Veränderungen oder Trends sind, was die Analyse und Interpretation erheblich vereinfacht und zuverlässigere Schlussfolgerungen ermöglicht.

Strenge Form vs. Schwache Form

Grundsätzlich unterscheidet man zwischen zwei Formen der Stationarität. Während die strenge Form der Stationarität („*strong stationarity*“) erfordert, dass die gesamte Wahrscheinlichkeitsverteilung einer Zeitreihe über die Zeit konstant bleibt (was bedeutet, dass die statistischen Eigenschaften wie Mittelwert und Varianz nicht von der Zeit abhängen dürfen), erfordert die schwache Form der Stationarität („*weak stationarity*“) lediglich, dass der Erwartungswert konstant sowie die Varianz endlich bleibt und die Kovarianz zwischen zwei Zeitpunkten lediglich aufgrund der zeitlichen Länge zwischen den gewählten Zeitpunkten, nicht jedoch aufgrund der gewählten Zeitpunkte selbst variiert. Die schwache Form ist daher weniger restriktiv und in der Praxis häufiger anwendbar als die strenge Form der Stationarität.

Prüfmechanismen & Stationarisierungsprozesse

Die Ermittlung der Stationarität einer Zeitreihe kann durch visuelle Inspektion von Zeitreihenplots, Autokorrelationsfunktionen und Teiltrends erfolgen, um offensichtliche Veränderungen im Mittelwert und der Varianz zu erkennen. Statistische Tests wie der Augmented Dickey-Fuller-Test oder der Kwiatkowski-Phillips-Schmidt-Shin-Test (*KPSS*) können verwendet werden, um die Stationarität quantitativ zu überprüfen. Zusätzlich können Transformationen wie die Differenzierung oder Log-Transformation angewendet werden, um eine Zeitreihe stationär zu machen. In diesem Falle spricht man von integrierten Zeitreihen, wobei die Notation $I(d)$ anzeigt, dass für die originale Zeitreihe insgesamt d Transformationen vollzogen wurden, um eine final als stationär anzusehende Zeitreihe zu erhalten. Obgleich solche Transformationen allgemein hilfreich zur Stationarisierung einer etwaig zu untersuchenden Zeitreihe sind, gilt es zu berücksichtigen, dass der Nachteil darin besteht, dass mit jedem Transformationsschritt Informationen über die absoluten Werte der ursprünglichen Daten verloren gehen.

Modellvarianten

Autoregressive Regressionsmodelle

Autoregressive Regressionsmodelle (kurz *AR-Modelle*) sind eine wichtige Komponente der Zeitreihenanalyse. Die Grundidee besteht darin, den aktuellen Wert einer Zeitreihe als eine gewichtete Linearkombination seiner vergangenen Werte zu modellieren. Diese Modelle nehmen an, dass vergangene Werte der Zeitreihe Informationen enthalten, die verwendet werden können, um zukünftige Werte vorherzusagen, und sie ermöglichen die Erfassung von Mustern, Trends und Abhängigkeiten in zeitlichen Datenverläufen.

$$Y_t = \delta_0 + \sum_{i=1}^{p} \delta_i Y_{t-i} + \epsilon_t$$

Moving Average Modelle

Moving Average Regressionsmodelle (kurz *MA-Modelle*) sind ein Konzept in der Zeitreihenanalyse, bei denen der aktuelle Wert einer Zeitreihe als Linearkombination vergangener Fehlerwerte (Residuen) modelliert wird. Die Grundidee besteht darin, dass die aktuellen Beobachtungen von den zufälligen Schwankungen oder Fehlern vergangener Zeitpunkte beeinflusst werden, wobei die Werte durch eine gewichtete Summe dieser Fehler modelliert werden. Diese Modelle ermöglichen die Erfassung von Glättungseffekten und können verwendet werden, um Vorhersagen für zeitliche Datenverläufe zu erstellen.

$$Y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

Autoregressive (Integrierte) Moving Average Modelle

Autoregressive Moving Average (*ARMA*) Regressionsmodelle kombinieren die Ideen der autoregressiven (*AR*) und Moving Average (*MA*) Modelle in der Zeitreihenanalyse. Die Grundidee besteht darin, die aktuelle Beobachtung einer Zeitreihe als eine Linearkombination ihrer vergangenen Werte und der Fehler (Residuen) vergangener Beobachtungen zu modellieren. ARMA-Modelle können dazu beitragen, sowohl zeitliche Abhängigkeiten als auch Glättungseffekte zu erfassen. Einer Erweiterung erfährt dieser Ansatz darüber hinaus durch sogenannte *ARIMA* (Autoregressive Integrierte Moving Average) Modelle, welche auf der ARMA-Grundidee aufbauen, indem sie zusätzlich jene Integration $I(d)$ einbeziehen, welche die entsprechend relevanten Zeitreihen stationär werden lässt. ARIMA-Modelle sind besonders nützlich, um saisonale Trends und zyklische Muster in den Daten zu modellieren und sind ein leistungsstarkes Instrument in der Zeitreihenanalyse um Vorhersagen für zeitliche Datenverläufe zu generieren.

$$Y_t = \mathrm{c} + \sum_{i=1}^{p} \delta_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

Vektorautoregressive Modelle

Vektorautoregressive (*VAR*) Modelle sind ein Konzept in der Zeitreihenanalyse, bei denen mehrere Zeitreihenvariablen miteinander modelliert werden. Die Grundidee besteht darin, dass jede Variable als Linearkombination ihrer eigenen vergangenen Werte und der vergangenen Werte aller anderen Variablen modelliert wird. VAR-Modelle ermöglichen die Erfassung von komplexen Wechselwirkungen zwischen verschiedenen Variablen im Laufe der Zeit und sind besonders nützlich für die Analyse von multivariaten Zeitreihendaten.

$$Y_{h,t} = c_h + \sum_{m=1}^{H} \sum_{i=1}^{p} (\delta_{h,m,i} Y_{m,t-i}) + \epsilon_{h,t} \; \forall \; h = 1 \to H$$

Granger-Kausalität

Granger-Kausalitätsmodelle sind ein Konzept in der Zeitreihenanalyse, das verwendet wird, um festzustellen, ob eine Zeitreihe eine (zusätzliche) Vorhersagekraft für eine andere hat. Die Grundidee besteht darin, dass eine Zeitreihe X eine Granger-Kausalität für eine Zeitreihe Y aufweist, wenn die Vergangenheitswerte von X dazu beitragen, die aktuellen Werte von Y besser vorherzusagen als durch eine reine Berücksichtigung der Vergangenheitswerten von Y bewerkstelligt werden könnte. Granger-Kausalitätsmodelle helfen darüber hinaus dabei, Richtungen und Stärke von Ursache-Wirkungs-Beziehungen zwischen unterschiedlichen Zeitreihen zu identifizieren wie auch quantifizieren.

$$Y_t = \mathrm{c} + \sum_{i=1}^{n} \delta_i Y_{t-i} + \sum_{j=1}^{m} \beta_j X_{t-j} + \epsilon_t$$

Anwendungsbeispiel

Zum Zwecke der hinreichenden Illustration der konkreten, praktischen Anwendungsmöglichkeiten der oberhalb beschriebenen Konzepte sollen einige dieser nunmehr auch noch einmal innerhalb eines abschließenden Anwendungsbeispiels aus dem Bereich der empirischen Finanzwissenschaften präsentiert sowie kurz beschrieben werden. Für das konkrete Beispiel selbst soll sich hierbei auf zwei grundlegende Datensets bezogen werden:

- Zum einen auf die täglichen (korrigierten) Schlusswerte des ***S&P 500 Index*** (kurz „*S&P 500*") im Zeitraum zwischen 01.01.2020 und 31.12.2022
- Und zum anderen auf die tagesbezogen publizierten Werte des ***CBOE Volatility Index*** (kurz „VIX") für denselben Zeitraum (d.h. für die Periode 01.01.2020 bis 31.12.2022)[1]

Konkret sollen für diese beiden Datensets zwei Dinge untersucht werden:

1. Lassen sich aktuelle Werte des S&P 500 aufgrund der vergangenen Entwicklung des S&P 500 erklären?
2. Kann die Berücksichtigung der vergangenen Entwicklung des VIX ebenfalls zur Erklärung der aktuellen Werte des S&P 500 beitragen?

Auf Basis dieser grundsätzlichen Fragestellungen können in weiterer Hinsicht die zu überprüfenden Null- (H_0) wie auch Alternativhypothesen ($H_{1.a}$ & $H_{1.b}$), die es anschließend über ein entsprechendes Zeitreihenmodell zu analysieren gilt, formuliert werden.[2] Exemplarisch lauten diese z.B. wie folgt:

$\boldsymbol{H_0}$**:** Die aktuellen Realisationen des S&P 500 lassen sich auf reine Zufälligkeit zurückführen.

$\boldsymbol{H_{1.a}}$**:** Es besteht ein zeitlicher Zusammenhang zwischen den vergangenen sowie aktuellen Realisationen des S&P 500.

$\boldsymbol{H_{1.b}}$: Es besteht ein zeitlicher Zusammenhang zwischen den vergangenen Werten des VIX und den aktuellen Realisationen des S&P 500.

Um die etwaige Gültigkeit (bzw. Ablehnbarkeit) der aufgestellten Hypothesen fortfolgend auch empirisch testen zu können, bedarf es eines entsprechenden Modelles. Zu diesem Zwecke soll sich, dem grundlegenden Ziel des vorliegenden Leitfadens entsprechend, einer Methode aus dem Bereich der regressionsbasierten Zeitreihenanalyse bedient werden. Konkret wird ein empirisches Modell aufgestellt, mit welchem überprüft werden soll, ob zeitlich vorangehende Beobachtungen des VIX einen Granger-kausalen Zusammenhang zu den jeweils unmittelbar vorliegenden Realisationen des S&P 500 aufweisen. Ein entsprechend zweckdienliches Modell könnte hierbei zum Beispiel wie folgt aussehen:

---

[1] **Achtung**: Für den Fall, dass im Zuge eigens durchgeführter Analysen die jeweiligen Beobachtungen innerhalb der entsprechenden Erhebungsperioden voneinander abweichen (z.B. weil sich ein Datenset nur auf Börsentage bezieht, während das andere auch separate Beobachtungen für Wochentage enthält oder weil die Datensets Erhebungspunkte an unterschiedlichen Wochentagen enthalten), müssen die vorliegenden Datensets zunächst noch einmal entsprechend bereinigt bzw. die grundlegenden Annahmen der Studie noch einmal entsprechend adaptiert werden [→ *glücklicherweise ist dieses Problem allerdings im vorliegenden Anwendungsbeispiel nicht anzutreffen*].

[2] Das besondere an den Alternativhypothesen $H_{1.a}$ & $H_{1.b}$ ist hierbei, dass diese nebeneinander existieren können (d.h. H_0 kann abgelehnt werden, wenn entweder nur $H_{1.a}$, nur $H_{1.b}$ oder $H_{1.a}$ und $H_{1.b}$ korrekt sind [*wobei im letzteren Falle weitere Analysen betreffend die gleichzeitige Gültigkeit mehrerer Hypothesen zu vollziehen wären, worauf jedoch aufgrund der Komplexität dieser Materie an dieser Stelle verzichtet wird*])

$$SP\_LRET_t = \mathrm{c} + \sum_{i=1}^{n} \delta_i SP\_LRET_{t-i} + \sum_{j=1}^{m} \beta_j VIX\_PC_{t-j} + \epsilon_t$$

SP_LRET_t stellt hierbei den jeweiligen Wert der logarithmierten Tagesrendite des S&P 500 für die Periode $t-1$ bis t dar (d.h. die stetige Rendite basierend auf dem adjustierten Schlusskurs des S&P 500 an Tag $t-1$ und dem adjustierten Schlusskurs des S&P 500 am Folgetag t), wohingegen SP_LRET_{t-i} den Wert einer vergangenen Beobachtung der stetigen Rendite des S&P 500 beschriebt, welche insgesamt i Tage vor t liegt.[3,4] Der übergeordnete Term $\sum_{i=1}^{n} \delta_i SP\_LRET_{t-i}$ wiederum besagt, dass im Modell für jede aktuelle, zu erklärende Beobachtung $SP_{LRET\,t}$ alle vergangenen Beobachtungen der logarithmierten Rendite des S&P 500 als potentiell einen Zusammenhang beschreibende Faktoren inkludiert werden, die mindestens 1 Tag und maximal n Tage vor der aktuellem Tag t liegen.[5,6] Die darüber hinaus inkludierten Variablen δ_i können hierbei als jeweilige Regressionsparameter für die einzelnen, vergangenen Renditebeobachtungen des S&P 500 angesehen werden, die es in weiterer Hinsicht zur Überprüfung der Ablehnbarkeit der Nullhypothese H_0 zugunsten der Alternativhypothese $H_{1.a}$ zu schätzen gilt. Die spezifischen VIX_PD_{t-j}, hingegen, können als repräsentativ für die vergangenen, tagesbezogenen (in Prozent gemessene) Änderungen des VIX angesehen werden, wobei erneut jede vergangene Beobachtung ab dem Vortag des aktuellen Tages t bis zu einem neuen, maximalen Look-Back von m Tagen vor dem aktuellen Tag t im final aufgestellten Modell inkludiert wird.[7] Gleichermaßen gilt für die einzelnen β_j, dass diese repräsentativ für die zugehörigen Sensitivitätsgewichte (d.h. zu schätzenden Regressionsparameter) der prozentuellen Änderungen des VIX anzusehen sind, auf deren Basis in weiterer Hinsicht die Ablehnbarkeit der Nullhypothese H_0 zugunsten der Alternativhypothese $H_{1.b}$ überprüft wird. Die final inkludierten Variablen c wie auch ϵ_t stellen wiederum zum einen die Konstante des Regressionsmodells (im Sinne des Erwartungswertes für SP_LRET_t, falls weder für die vergangenen Werte des S&P 500 noch für die vergangenen Werte des VIX ein zeitlicher Zusammenhang zu den aktuellen werten des S&P 500 feststellbar ist) und zum anderen den jeweiligen, beobachtungsspezifischen Residualwert laut Modell (als Differenz zwischen dem real beobachteten Wert SP_LRET_t und dem laut Modell geschätzten Wert $\widehat{SP\_LRET_t}$) dar.

Unter Verwendung des oberhalb definierten Modelles kann nunmehr mit der konkreten, regressionsbasierten Berechnung des Modells sowie der darin enthaltenen Koeffizienten begonnen werden.[8] Die Ergebnisse für eine Modellschätzung mit $n = m = 3$ Lags würde hierbei zum Beispiel zu folgendem Resultat führen:

---

3 $SP\_LRET_t = \ln\left(\frac{Adj.\ Schlusskurs\ S\&P\ 500\ am\ Tag\ t}{Adj.\ Schlusskurs\ S\&P\ 500\ am\ Vortag\ von\ t\ (d.h.\ am\ Tag\ t-1)}\right)$

4 So würde z.B. bei $i = 2$ der Wert von $SP_{LRET\,t-2}$ der logarithmierten Rendite für die Periode $t-3$ bis $t-2$ entsprechen.

5 Ein Wert von $n = 3$ würde z.B. bedeuten, dass die vergangenen Renditen von $t-1$, $t-2$ sowie $t-3$ gesamthaft als mögliche, erklärende Variablen im Regressionsmodell inkludiert werden würden.

6 Es sei an dieser Stelle noch einmal erwähnt, dass es allgemein keinen singulären, allgemeingültigen Ansatz gibt, welcher vorschreibt, wie die optimale Lag- bzw. Look-Back-Periode zu bestimmen ist (bzw. ob auch sämtliche vergangenen Beobachtungen oder nur spezifisch selektierte i zu berücksichtigen sind), entsprechende Entscheidungen jedoch nichtsdestotrotz unter Bezugnahme auf allfällige theoretische bzw. literaturgestützte Erkenntnisse durch Nutzung empirischer Hilfswerkzeuge hinreichend getroffen sowie gestützt werden können.

7 $VIX\_PC_{t-1} = \frac{Publizierter\ Wert\ des\ VIX\ am\ Vortag\ von\ t\ (d.h.\ am\ Tag\ t-1)}{Publizierter\ Wert\ des\ VIX\ zwei\ Tage\ vor\ t\ (d.h.\ am\ Tag\ t-2)} - 1$

8 Klassische Software-Tools zur Durchführung solcher Berechnungen sind unter anderem MS Excel, Stata, R, Python, EViews, etc.

Regressions-Statistik	
Multipler Korrelationskoeffizient	0,271
Bestimmtheitsmaß	0,073
Adjustiertes Bestimmtheitsmaß	0,066
Standardfehler	1,561
Beobachtungen	752

ANOVA

	Freiheits-grade (df)	*Quadrat summen (SS)*	*Mittlere Quadratsumme (MS)*	*Prüfgröße (F)*	*F krit*
Regression	6	143,85	23,97	9,84	0,00
Residue	745	1814,24	2,44		
Gesamt	751	1958,09			

	Koeffi-zienten	*Standard-fehler*	*t-Statistik*	*P-Wert*	*Untere 95%*	*Obere 95%*
Schnittpunkt	0,047	0,058	0,814	0,416	-0,066	0,160
S&P - L1	-0,290	0,052	-5,614	0,000	-0,392	-0,189
S&P - L2	0,101	0,052	1,932	0,054	-0,002	0,203
S&P - L3	-0,030	0,052	-0,578	0,563	-0,131	0,071
VIX - L1	-0,024	0,009	-2,641	0,008	-0,043	-0,006
VIX - L2	-0,004	0,009	-0,468	0,640	-0,023	0,014
VIX - L3	-0,017	0,009	-1,826	0,068	-0,035	0,001

Tabelle 1: Berechnung des aufgestellten Zeitreihenmodells für $\boldsymbol{n = m = 3}$

Auf Basis dieser Ergebnisse können nunmehr einige Aussagen gemacht werden. Zum einen lässt sich festhalten, dass vier Faktoren auf Basis ihrer zugehörigen Kriterienwerte (→ t-Statistik, p-Wert) als signifikant auf Basis eines minimalen Konfidenzniveaus von 90,00% (entspricht einem maximalen Signifikanzniveau bzw. zugelassenem α-Fehler von 10,00%) anzusehen sind.[9] Konkret handelt es sich hierbei zum einen um den „$t-1$“-Lag wie auch um den „$t-2$“-Lag der S&P 500 Renditen (was bedeuten würde, dass die stetige Rendite des Vortages wie auch des Vorvortages im Aggregat einen entsprechenden Zusammenhang mit dem jeweils aktuellen, logarithmierten Renditewert des S&P 500 aufweist) und zum anderen um den „$t-1$“-Lag wie auch um den „$t-3$“-Lag der prozentuellen Entwicklung des VIX (was bedeutet, dass im Schnitt ein Zusammenhang zwischen den aktuell beobachtbaren, stetigen Renditenwerten des S&P 500 und der publizierten, prozentuellen Änderung des VIX am Vortag von t wie auch am Tag drei Tage vor t zu existieren scheint). Gleichermaßen lässt sich erkennen, dass der signifikante „S&P 500“-Faktor von einem eher gemischten Einfluss ausgeht, während die beiden signifikanten VIX-Faktoren einen allgemein negativen Zusammenhang zu beschreiben vermögen. [10]

[9] Genauer gesagt kann der „$t-1$“-Lag-Faktor des VIX als signifikant auf einem 99,00%-Konfidenzniveau [SN: 1,00%] angesehen werden, wohingegen der „$t-2$“-Lag-Faktor des S&P 500 wie auch der „$t-3$“-Lag-Faktor des VIX nur ein Signifikanzniveau von 10,00% [KN: 90,00%] aufzuweisen vermögen.

[10] Interpretationsbezogen lässt sich festhalten, dass nach einer Steigerung des S&P 500 um einen Prozentpunkt einen Tag darauf ein Wert erwartet werden kann, welcher im Schnitt um 0,29 Prozentpunkt niedriger liegt als der zu erwartende Durchschnitt (welcher aufgrund der insignifikanten Konstante bei 0,00% angenommen werden kann) und zwei Tage später ein solcher erwartet werden kann, welcher um 0,1 Prozentpunkte höher als ein entsprechender Erwartungswert liegt. Eine Steigerung des VIX um einen Prozentpunkt würde wiederum zur Erwartung eines um 0,024 Prozentpunkte verminderten Renditewert beim S&P 500 für den Folgetag und eines um

Betreffend die Überprüfbarkeit der aufgestellten Nullhypothese H_0 sowie der allfälligen Alternativhypothesen ($H_{1.a}$ sowie $H_{1.b}$), lässt sich wiederum festhalten, dass sowohl die Signifikanz einzelner Faktoren wie auch die Tatsache, dass der modellspezifische F-Wert (9,84) größer als der zugehörige kritische F-Wert (0,00) ist, als ausreichende Kriterien zur Ablehnung der Nullhypothese angesehen werden können.[11] Stattdessen können durch die Signifikanz sowohl zweier „S&P 500“-lag-Faktors wie auch mindestens eines „VIX“-Lag-Faktors (vorerst) beide Alternativhypothesen als allgemeinhin sinnvollere Alternativannahmen für die Durchführung aufbauender Studien sowie Theorien anstelle der Annahme vollständiger Zufälligkeit der Renditenentwicklung des S&P 500 angesehen werden.

Um final noch zu überprüfen, ob qualitativ bessere Ergebnisse erreicht werden können, indem ein alternatives Modell geschätzt wird, soll zum allgemeinen Vergleich noch ein Modell mit 2 Lags (d.h. mit $n = m = 3$) berechnet werden. Die konkreten Ergebnisse der Regression einer solchen Modellvariante zeigen hierbei folgendes Bild:

Regressions-Statistik	
Multipler Korrelationskoeffizient	0,261
Bestimmtheitsmaß	0,068
Adjustiertes Bestimmtheitsmaß	0,063
Standardfehler	1,563
Beobachtungen	752

ANOVA

	Freiheitsgrade (df)	*Quadratsummen (SS)*	*Mittlere Quadratsumme (MS)*	*Prüfgröße (F)*	*F krit*
Regression	4	133,31	33,33	13,64	0,00
Residue	747	1824,78	2,44		
Gesamt	751	1958,09			

	Koeffizienten	*Standardfehler*	*t-Statistik*	*P-Wert*	*Untere 95%*	*Obere 95%*
Schnittpunkt	0,038	0,057	0,653	0,514	-0,075	0,150
S&P - L1	-0,282	0,051	-5,513	0,000	-0,382	-0,181
S&P - L2	0,108	0,051	2,100	0,036	0,007	0,209
VIX - L1	-0,023	0,009	-2,553	0,011	-0,042	-0,005
VIX - L2	-0,002	0,009	-0,263	0,793	-0,021	0,016

Tabelle 2: Berechnung des aufgestellten Zeitreihenmodells für $n = m = 2$

Wie sich erkennen lässt, bleiben die Konklusionen betreffend die (ungefähre) Wirkungsstärke sowie -richtung als auch die Signifikanz des „$t - 1$“-Lag und „$t - 2$“-Lag der S&P 500 Renditen wie auch des „$t - 1$“-Lag der prozentuellen Entwicklung des VIX auch in diesem Modell bestehen. Nichtsdestotrotz gilt es allerdings anzumerken, dass obgleich sämtliche signifikanten Koeffizienten nunmehr auf Basis eines Konfidenzniveaus von 95,00% (entspricht einem Signifikanzniveau von 5,00%) zur Ablehnung der

einen um 0,017 Prozentpunkte verminderten Renditewert beim S&P 500 am 3. Tag nach der VIX-Beobachtung führen.

[11] Zur Erinnerung: Wäre der modellspezifische F-Wert nicht größer als der zugehörige, kritische Schwellenwert, würde dies bedeuten, dass die Annahme, dass alle Regressionsparameter einen Wert von 0,00 haben (und dadurch keine der getesteten Vergangenheitsvariablen einen Zusammenhang zu den aktuellen Rendite-realisierungen des S&P 500 aufweist) nicht abgelehnt werden könnte.

Nullhypothese H_0 herangezogen werden können, ein Vergleich der (adjustierten) Bestimmtheitsniveaus erkennen lässt, dass das erste Modell ($n = m = 3$) gegenüber dem zweiten, alternativen Modell (mit $n = m = 2$) zu bevorzugen ist, da die (wenn auch leichte) Signifikanz des zusätzlichen „$t - 3$"-Lag-Faktors des VIX im ersten Modell zu einer dermaßen höheren Güte der modellbasierten Schätzwerte für die stetigen Renditen des S&P 500 zu führen scheint, dass der dadurch entstehende Koeffizientenmalus im Sinne eines höheren Strafterms, welcher zur Berechnung des adjustierten Bestimmtheitsmaßes verwendet wird, gegenüber dem zweiten Modell kompensiert werden kann.

Abschließend sollen zur Abrundung des vorgestellten Anwendungsbeispiels noch einige Bemerkungen zu potenziell fortfolgenden Robustheitstests gemacht werden (die im Zuge vollwertiger Studien in weiterer Hinsicht durchzuführen wären):

- Allgemein wurde im Zusammenhang mit der Berechnung des oberhalb dargestellten Regressionsmodelles mit klassischen Schätzmethoden gearbeitet, die neben anderen Charakteristika auch von einer (*i.i.d.*) Normalverteilung der Beobachtungswerte wie auch der Modellresiduen ausgehen. Da dieser Aspekt jedoch allen voran im Bereich der Zeitreihenanalyse nicht ohne weiteres angenommen werden kann, sollten die erhaltenen Ergebnisse noch einmal mittels hinreichender Gütekriterientests auf Aspekte wie etwa der Multikollinearität in den beschreibenden Variablen sowie der Autokorrelation oder Heteroskedastizität in den Modellresiduen überprüft werden. Weiterführend können in diesem Zusammenhang vereinfachend auch allgemein alternative Schätzmethoden verwendet werden, welche Aspekte wie die soeben beschriebenen von vorneherein berücksichtigen (wie etwa *HAC*-Schätzmethoden oder Ähnliches)
- Unter Berücksichtigung des herangezogenen Datenzeitraumes lässt sich festhalten, dass dieser im vorliegenden Falle durch einige besondere, nicht regelmäßig wiederkehrende respektive nicht von vorneherein erwartbare Ereignisse externer Natur geprägt war (→ COVID-19, Börsenschock & -rally, Rekordinflation , etc.). Wissen um solche Aspekte kann daher genutzt werden, um z.B. durch die Inklusion spezifischer Alternativvariablen zu eruieren, ob die Ergebnisse des Grundmodelles konstant und frei von allfälligen *Omitted Variable Biases* sind (ähnliches kann auch erreicht werden, indem das originale Modell noch einmal für separate Unterperioden berechnet wird, um zu erfahren, ob allfällige Zusammenhänge bloß Zeichen ihrer Zeit oder von allgemeingültiger Natur sind).
- Um zu eruieren, ob die erhaltenen Ergebnisse auch einen unmittelbar praktischen Anwendungszweck haben oder nur dem besseren Verständnis wissenschaftlicher Zusammenhänge diesen, bietet es sich an, mittels der erhaltenen Schätzmodellvarianten die jeweils relevanten Werte für die einzelnen Beobachtungspunkte zu schätzen sowie auch mit einem alternativen Set an Daten abzugleichen (→ In-Sample- vs. Out-of-Sample Testverfahren). Sollten die Ergebnisse hierbei zeigen, dass die erstellten Modelle auch vorher nicht bekannte Werte gut schätzen können, bietet es sich beispielsweise an, diese im Zuge weiterführender Profitabilitätsstudien zu untersuchen.

Literaturempfehlungen

Dougherty, C. (2016). *Introduction to Econometrics* (5. Ausg.). Oxford.

Enders, W. (2015). *Applied Econometric Time Series* (4. Ausg.). Hoboken.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics.* (5. Ausg.) New York.

Komlos, J., & Süssmuth, B. (2022). *Empirische Ökonomie. Eine Einführung in Methoden und Anwendungen* (2. Ausg.). Berlin.

Schröder, M. (2012). *Finanzmarkt-Ökonometrie. Basistechniken, Fortgeschrittene Verfahren, Prognosemodelle.* Stuttgart.

Stoetzer, M.-W. (2020). *Regressionsanalyse in der empirischen Wirtschafts- und Sozialforschung Band 2. Komplexe Verfahren.* Berlin.

Winker, P. (2010). *Empirische Wirtschaftsforschung und Ökonometrie.* Heidelberg.