

Zum Umgang mit Regressionsanalysen in Kartellschadensersatzfällen¹

Roman Inderst², Stefan Thomas³

Abstract

In seinem Urteil vom 13. April 2021 (KZR 19/20) hat sich der BGH im Rahmen von Schadensersatzklagen mit dem Instrument der Regressionsanalyse auseinandergesetzt und dabei drei Gütekriterien formuliert: eine hinreichend verlässliche Datengrundlage, die methodische Korrektheit sowie die Signifikanz der Ergebnisse. Wir gehen zunächst auf die allgemeinen Grundsätze dieses Instruments ein, was es sodann erlaubt, die drei Kriterien des BGH zu betrachten. Hierbei soll gezeigt werden, dass die Bewertung einer Regressionsanalyse nicht in einem festen Schema folgen kann, was sich auf die konkrete Anwendung der vom BGH formulierten Kriterien in der Tatsacheninstanz auswirkt. Insbesondere darf u.E. das vom BGH aufgestellte Prüfprogramm nicht als eine abschließende „Checkliste“ für die Berücksichtigungsfähigkeit der Aussagen einer Regressionsanalyse verstanden werden. So wäre ein Versuch, sich für eine Schadensschätzung allein auf statistische Gütemaße zu beschränken, einschließlich der Signifikanz der Ergebnisse, nicht sinnvoll. Stattdessen bieten sich gerade der in der Regel vergleichbare Aufbau der Regressionsanalysen der verschiedenen Gutachten in einem Verfahren dazu an, dass das Gericht die Ergebnisse dieser Regressionsanalysen differentialdiagnostisch bewertet, wobei es zwingend auch die Spezifika der Industrie, der vorhandenen Daten sowie des betrachteten Verstoßes im Auge behält. Dies stellt sicher, dass auch wesentliche Informationen, die sich nicht in das Raster einer Regressionsanalyse pressen lassen, im Rahmen einer umfassenden Schadensschätzung nach § 287 ZPO angemessen berücksichtigt werden können. Denn die Durchführung einer Regressionsanalyse ersetzt keine Gesamtbetrachtung aller verfügbaren Evidenz nach dem Maßstab der freien richterlichen Beweiswürdigung.

¹ Wir danken Dr. Frabian Griem und Dr. Raphael Kuhlmann für ihr kritisches Lesen der methodischen Ausführungen.

² Goethe University Frankfurt, inderst@finance.uni-frankfurt.de

³ Eberhard Karls University, Tübingen, thomas@jura.uni-tuebingen.de.

Inhalt

| | | |
|------------|---|-------------|
| I | Motivation und Fragestellungen | 3 - |
| II | Regressionsanalysen | 6 - |
| II.i | Der Schritt des Aufsetzens der zentralen Regressionsgleichung | 7 - |
| II.ii | Einbezug des Verstoßes im Rahmen der Indikatorvariablen-Methode..... | 8 - |
| II.iii | Der Schritt der Schätzung der Koeffizienten der Regressionsgleichung (einschließlich des Verstoßeffektes) – sowie die damit einhergehende unvermeidbare Schätzungenauigkeit | 8 - |
| II.iv | Der Schritt der ökonomischen Interpretation der Schätzergebnisse | 9 - |
| II.v | Interpretation des Koeffizienten der Verstoß-Indikatorvariablen (Verstoßeffekt)..... | 10 - |
| III | Verlässliche Datengrundlage | 12 - |
| III.i | Vorbemerkung | 12 - |
| III.ii | Aufbau des Datengerüsts für die erklärende Variable (Preis) | 15 - |
| III.iii | Datenfehler..... | 16 - |
| III.iv | Systematische, verstoßbedingte Beeinträchtigungen der Kontrollvariablen | 19 - |
| IV | Korrekte Methode | 20 - |
| IV.i | Zur Bedeutung der Methode..... | 20 - |
| IV.ii | Ökonomische Prüfung der Sinnhaftigkeit der (geschätzten) Koeffizienten der Kontrollvariablen | 21 - |
| IV.iii | Einbezug des konkreten Verstoßes | 23 - |
| IV.iv | Zur Rolle statistischer Gütemaße | 24 - |
| V | Signifikanz | 26 - |
| V.i | Das statistische Konzept der Signifikanz | 26 - |
| V.ii | Idealvorstellung eines Testverfahrens und Notwendigkeit der gemeinsamen Berücksichtigung von Fehlern 1. und 2. Art | 27 - |
| V.iii | Die Beschränktheit der Aussagekraft signifikanter und insignifikanter Ergebnisse | 29 - |
| V.iv | Das (quasi zwangsläufige) Ignorieren anderer Informationen und Evidenz in der Regressionsanalyse | 30 - |
| V.v | Der Umgang mit widersprüchlichen „signifikanten“ Ergebnissen | 32 - |
| VI | Schluss | 33 - |

I Motivation und Fragestellungen

In seinem Urteil vom 13. April 2021 (KZR 19/20) hat sich der BGH im Rahmen von Schadensersatzklagen explizit mit dem Instrument der Regressionsanalyse auseinandergesetzt. Damit die mittels Regressionsanalyse ermittelte kartellbedingte Preiserhöhung als relevantes Indiz für oder gegen einen entstandenen Schaden angeführt werden kann, hat der BGH zudem drei Kriterien formuliert.⁴ In diesem Artikel gehen wir auf die Frage nach der beweisrechtlichen Einordnung der Regressionsanalyse in der Schadensschätzung und der Bedeutung der vom BGH insoweit formulierten Gütekriterien ein. Wir greifen hierfür dezidiert die drei Kriterien des BGH – die hinreichend verlässliche Datengrundlage, die methodische Korrektheit sowie ein signifikantes Ergebnis – auf. Ausgehend sowohl von den ökonometrisch-statistischen⁵ Grundlagen als der juristischen Determinanten sowie unserer praktischen Erfahrung als Gutachter in Kartellschadensersatzfällen versuchen wir, diese Kriterien auszufüllen. Unsere nachfolgenden Ausführungen sind dabei gänzlich losgelöst von den Spezifika eines konkreten Falles und damit auch insbesondere des vom BGH in seinem Urteil behandelten Falles. An keiner Stelle müssen wir Bezug nehmen zu dem dort verhandelten konkreten Sachverhalt.

Hervorzuheben ist bereits am Anfang unseres Beitrags, dass der Umgang mit Regressionsanalysen eine Reihe von Angriffs- und Verteidigungsszenarien in einem Kartellschadensersatzfall betreffen kann. Die Methode kann den Nachweis des Umfangs eines Schadens betreffen ebenso wie die Widerlegung einer Schadensvermutung des § 33a Abs. 2 GWB oder die Erschütterung eines Anscheinsbeweises. Sie kann Umbrella-Schäden betreffen ebenso wie Nachteile indirekter Abnehmer, sie kann die Weiterwälzung betreffen ebenso wie die passing-on defence. Im Folgenden wird hierfür *pars pro toto* der Begriff der „Schadensschätzung“ verwendet. Der Grad der „Großzügigkeit“ oder „Engherzigkeit“, mit der man der Regressionsanalyse als relevantes Indiz *sub specie iuris* begegnet, korreliert daher nicht eindeutig positiv mit

⁴ Konkret führt der BGH aus: „Sie stellt damit zugleich - wenn sie auf einer hinreichend verlässlichen Datengrundlage methodisch korrekt und mit signifikanten Ergebnissen durchgeführt worden ist - ein relevantes Indiz für oder gegen den im Rahmen eines Grundurteils zu ermittelnden Umstand dar, dass der klagenden Partei durch den Kartellverstoß wahrscheinlich jedenfalls ein Schaden in irgendeiner Höhe entstanden ist“, BGH, Urteil v. 13.04.2021, KZR 19/20, Rn. 66 – *LKW-Kartell II*.

⁵ Bereits an dieser Stelle scheint eine begriffliche Klärung hilfreich. Die Ökonometrie bezeichnet ein Teilgebiet der Wirtschaftswissenschaften, in dem ökonomische Theorien mit Daten zusammengeführt werden, um damit Hypothesen zu testen oder aber (lediglich) ökonomische Phänomene quantitativ zu analysieren. Die Ökonometrie bedient sich dabei (auch) Methoden der Statistik. Dies betrifft sowohl Methoden der deskriptiven Statistik, die zur Beschreibung der vorhandenen Daten verwendet werden. Die Inferenzstatistik versucht stattdessen Schlussfolgerungen zu ziehen, die über diese Daten hinausgehen. Das Paradebeispiel für Letztere ist der Schluss von einer Stichprobe auf die unbekannte Gesamtpopulation, welcher notwendigerweise mit Fehlern behaftet ist, deren wahrscheinliches Ausmaß mit abgeschätzt werden soll. Die Regressionsanalyse ist ein Instrument einer solchen schließenden, nicht lediglich beschreibenden Statistik.

den Interessen der Kläger- oder Beklagenseite, sondern er betrifft die Schätzgenauigkeit insgesamt. Daher sollte eine möglichst differenzierte Ausfüllung der vom BGH entwickelten Kriterien die Rechtspraxis bestimmen, weil durch eine Erhöhung der Schätzgenauigkeit insgesamt die rechtsbefriedigende Funktion eines Zivilurteils gesteigert wird. Es ist zu beachten, dass auch der Kartellschadensersatz zuvörderst eine Kompensationsfunktion hat, so dass die Wirksamkeit eines Kartellschadensersatzregimes sich in erster Linie danach beurteilen sollte, wie genau eine Schätzung in der Lage ist, den tatsächlich zu kompensierenden Nachteil zu bestimmen. Damit ist aber unmittelbar die Frage angesprochen, wie genau die Rückschlüsse sind, die aus einem signifikanten oder nicht-signifikanten Ergebnis für Vorliegen oder Fehlen eines Schadens gezogen werden können. Da das Gericht eine umfassende Beweiswürdigung unternimmt, muss es ein Verständnis dafür erlangen, weshalb eine Regressionsanalyse zu einem bestimmten Ergebnis gelangt ist. Eine Berücksichtigung der Ergebnisse anhand einer einfachen „Checkliste“ würde diesem Postulat nicht gerecht. In diesem Sinne sollte der BGH daher auch nicht verstanden werden. Die von ihm formulierten Gütekriterien sollten vielmehr der Tatsacheninstanz Anlass geben, sich mit den drei genannten Prüfungsebenen „Datengrundlage“, „Methode“ und „Frage der Signifikanz“ kritisch auseinanderzusetzen. Das erscheint insbesondere dann geboten und zielführend, wenn mehrere Regressionsanalysen mit unterschiedlichen Ergebnissen durch Kläger und Beklagte und ggf. gerichtlich bestellte Sachverständige vorgelegt werden.

Dies vorausgeschickt, wollen wir bereits vorab auf folgende Gefahren eines einseitigen Verständnisses der Regressionsmethoden hinweisen, um zu zeigen, worum es im Folgenden gehen wird:

Zur „verlässlichen Datengrundlage“: Es wäre nicht überzeugend, im Rahmen einer umfassenden Schadensschätzung oder anderer schadensersatzrechtlicher Kontexte (etwa der Widerlegung der Schadensvermutung des § 33a Abs. 2 GWB) eine Regressionsanalyse allein deshalb vollständig zu verwerfen, da im verwendeten Datensatz einzelne Fehler aufgedeckt wurden. Die Forderung einer lückenlosen „Fehlerfreiheit“ könnte das Instrument der Regressionsanalyse bereits angesichts der unvermeidbaren Schätzgenauigkeit der Methode faktisch entwerten. Das bedeutet umgekehrt nicht, dass der Grad der Fehlerbehaftung irrelevant wäre. So wäre es falsch, Ungenauigkeiten zu ignorieren mit dem Hinweis, diese würden sich im Durchschnitt schon „ausgleichen“. Vielmehr gilt es insbesondere zu unterscheiden zwischen unsystematischen und systematischen Fehlern, und es geht ferner darum, ob und wie diese das Ergebnis einer Regressionsanalyse verzerren können.

Zur „korrekten Methode“: Es wäre problematisch, aus dem Kriterium der „korrekten Methode“ ableiten zu wollen, dass es nur eine bestimmte korrekte Methode gäbe, um eine Regressionsanalyse in einem konkreten Fall durchzuführen, die gleichsam als wissenschaftliche Wahrheit zunächst zu ermitteln wäre. U.E. gibt das Judikat des BGH auch keinen Anlass, dieses Kriterium so zu verstehen. Daher kann es auch nicht darum gehen, eine ganz bestimmte Methode zu finden, auf die ein gerichtlich bestellter Sachverständiger nach den Regeln der ökonomischen Wissenschaft gleichsam festgelegt wäre. Ferner ist mit Blick auf Tatsacheninstanz im Gerichtsprozess zu beachten, dass das in den hier interessierenden Fällen

der Schadensschätzung praktizierte Vorgehen eines Gutachters i.d.R. von der Textbuch-Vorstellung der Durchführung eines (möglichst vollständig kontrollierten) Tests abweicht. Die in der Regel verwendeten (statistischen) Konzepte gehen allerdings von einer solchen (Ideal-)Vorstellung bei der Bewertung einer Methode aus. Bei der Frage nach der Korrektheit der Methode kann es daher auch nicht zuvörderst um die Auswahl einer bestimmten Methode anhand statistischer Gütekriterien gehen. Es geht stattdessen schlicht um eine hinreichend angemessene (in der Regressionsgleichung formalisierte) Beschreibung der betrachteten Industrie und des Sachverhalts (Verstoßes), was letztlich eine ökonomische und keine statistische Fragestellung darstellt. Die Korrektheit der angewendeten Methode lässt sich damit nicht (allein) an statistischen Kenngrößen ableiten, sondern sie erfordert vor allem hinreichende Kenntnisse der Industrie sowie der Natur des Verstoßes.⁶

Zur „Signifikanz“: Die statistische Signifikanz allein besitzt keine Aussagekraft für die Schlussfolgerungen, die aus dem Ergebnis einer Regressionsanalyse gezogen werden können. Insbesondere darf Signifikanz nicht im umgangssprachlichen Sinne als Wahrscheinlichkeit eines signifikanten Schadens verstanden werden, was die Entscheidung des BGH aber auch nicht nahelegt. Vielmehr ist die Signifikanz immer mit dem Test der sog. Nullhypothese verbunden. Lautet die Nullhypothese „kein Schaden“, dann bedeutet ein signifikantes Ergebnis, dass diese Hypothese mit einer gewissen, als noch hinnehmbar empfundenen Irrtumswahrscheinlichkeit verworfen werden kann. Ergibt die Regressionsanalyse hingegen ein nicht-signifikantes Ergebnis, würde man bei der Ablehnung der Nullhypothese „kein Schaden“ mit einer größeren, über dieser Schwelle liegenden Wahrscheinlichkeit einen Irrtum begehen.. Losgelöst von der Berücksichtigung dieses Aussagewertes eines Ergebnisses ist aber stets Folgendes zu berücksichtigen. Jegliche sinnhafte Interpretation der Signifikanz setzt eine korrekte Methode voraus – und damit zwingend ein (Daten-)Modell, das die Industrie und den Verstoß hinreichend gut abbildet. Daher könnte eine rein mechanische Interpretation der Signifikanz bzw. des Fehlens einer solchen zu schweren Fehlern führen, wenn man davon die weitere Schätzung abhängig machen wollte. Daher ist es wichtig, im Sinne des BGH zu beachten, dass es sich um ein Indiz handelt, dass also die Überzeugungsbildung sich nicht in der Hinnahme des Ergebnisses ohne weitere Auseinandersetzung mit der gewählten Methode begnügen kann.

Die damit bereits im Überblick und zunächst nur skizzenhaft aufgezeigten Gefahren einer Fehlinterpretation der Regressionsanalyse sollen allerdings auf keinen Fall die Nützlichkeit dieser Methode in Abrede stellen. Im Gegenteil ist dieser Methode, bei richtiger ökonomischer Anwendung und vor allem

⁶ Dies bedeutet nicht, dass nicht auch Ansätze, die hinsichtlich der Natur des Verstoßes agnostisch sind, zum Einsatz kommen können. Wie noch dargestellt wird, laufen diese unter anderem aber Gefahr, alle nicht erklärten Unterschiede zwischen beispielsweise dem Verstoßzeitraum und einem Vergleichszeitraum fälschlicherweise eben jenem Verstoß zuzuschreiben. Die falsche Spezifikation mag auch, wie wir noch darstellen, dazu führen, dass der Verstoß fälschlicherweise „wegerklärt“ wird. Eine ökonomische Gesamtschau ist stets unerlässlich. Genauso wenig soll nahegelegt werden, dass die gesamte Analyse zwingend von Industrieexperten durchzuführen ist. Allerdings legt es unsere praktische Erfahrung nahe, hier keine wissenschaftstheoretisch gebotene Trennung von Entdeckungs- und Begründungszusammenhang einzufordern.

bei richtiger Interpretation, der Vorzug zu geben gegenüber einem einfachen Mittelwertvergleich mit ggf. ökonomisch unbefriedigenden Zu- oder Abschlägen durch das Gericht. Insbesondere ist dieser Methode aber auch der Vorzug zu geben vor einer Schätzung, die die in den Daten enthaltenen Informationen schlicht ignoriert. Allerdings wollen wir mit diesem Aufsatz auch davor warnen, das Instrument der Regressionsanalyse und die damit erzielten Ergebnisse in ein Schema zu pressen, das es dann vermeintlich erlauben würde, die Relevanz und Belastbarkeit der Ergebnisse anhand weniger Kriterien zu prüfen. Ein solches Schema kann es nicht geben, muss es aber für eine Schadensschätzung auch nicht geben. So besitzt die Signifikanz von Ergebnissen für sich genommen keine Aussagekraft, und statistische Gütemaße alleine sind nicht hinreichend, um die Korrektheit der Methode zu prüfen. Ebenso wäre ein Gericht nicht gut beraten, die Regressionsergebnisse etwa eines bestellten Sachverständigen ohne weiteres Hinterfragen der dahinterliegenden Annahmen, so insbesondere des unterstellten (Daten-)Modells, zu akzeptieren. Ebenso sollte stets hinterfragt werden, wie sich das Ergebnis der Regressionsanalyse zu anderer verfügbarer Evidenz verhält. Darauf gehen wir nachfolgend im Detail ein.

Wir gehen wie folgt vor: Abschnitt II stellt die Regressionsanalyse im Kontext einer Vergleichsmarktanalyse kurz dar.⁷ In Abschnitt III führen wir zur (hinreichend belastbaren) Datengrundlage einer Regressionsanalyse aus, in Abschnitt IV zur (korrekten) Methode und in Abschnitt V zur Signifikanz. Wir versuchen hierbei stets, unsere konzeptionellen Ausführungen mit fiktiven Beispielen zu illustrieren. Diesen Beispielen räumen wir relativ viel Raum ein. Sie sollten nicht in einen Zusammenhang mit bestimmten tatsächlichen Fällen gestellt werden. Abschnitt VI enthält Schlussfolgerungen.

II Regressionsanalysen

Es mangelt nicht an verständlichen Einführungen in die prinzipiell einfachen Grundlagen der Regressionsanalyse. Nachfolgend sollen daher nur die wesentlichen Begrifflichkeiten sowie das grundsätzliche Vorgehen bei einer Regressionsanalyse dargestellt werden.⁸ Hierfür sowie auch für die nachfolgenden Ausführungen ist kein statistisches Vorwissen nötig.⁹

⁷ Viele der in diesem Artikel angesprochenen Probleme der Durchführung einer Regressionsanalyse in einer Vergleichsmarktbetrachtung treten bei einer sogenannten Marktsimulation nicht auf. Insbesondere zwingt diese den Gutachter dazu, seine Vorstellung zur Industrie und zum Markt explizit zu machen, statt fehlendes Wissen hinter einem „Fehlerterm“ in einer Regressionsanalyse zu verstecken. Leider besteht ein wesentlicher Nachteil der Methode der Marktsimulation darin, dass die Komplexität vieler Märkte nicht oder aber nicht hinreichend gut mit den in der Industrieökonomie bislang entwickelten Modellen in Einklang zu bringen ist.

⁸ Hierzu *Inderst/Thomas*, Schadensersatz bei Kartellverstößen, 2. Aufl., 2018, Kap. III 3.5.3.

⁹ Wir verzichten deshalb auch durchgehend auf den Verweis auf statistische oder ökonometrische Standardwerke. Inderst und Thomas (Fn. 8) referieren darauf genauso wie auf vorhandene, vornehmlich an Praktiker orientierte Literatur.

II.i Der Schritt des Aufsetzens der zentralen Regressionsgleichung

Ausgangspunkt einer Regressionsanalyse ist die Regressionsgleichung. Diese ist für alle nachfolgenden Überlegungen zentral. Sie beschreibt mathematisch den Zusammenhang zwischen Kontrollvariablen (auch erklärende Variablen oder Regressoren genannt) und der zu erklärenden Variablen (auch abhängige Variable oder Regressand genannt). Der einfachste Fall ist dabei offensichtlich der einer Regressionsgleichung mit einer einzigen Kontrollvariablen. Ein Beispiel: So mag sich ein Gutachter damit begnügen, den Salatpreis (in Abwesenheit eines Verstoßes) allein in Abhängigkeit vom Niederschlag zu sehen. Oft wird dieser Zusammenhang mittels einer einfachen linearen Beziehung erfasst, so dass in diesem Beispiel das Modell unterstellt, dass der Salatpreis bestimmt ist durch eine einfache multiplikative Beziehung mit der Regenmenge, welche dann durch einen einfachen Koeffizienten charakterisiert ist.¹⁰ Der so ermittelte Koeffizient ließe sich dann beispielsweise wie folgt interpretieren: Wenn die Regenmenge um x Liter steigt, dann ändert sich der Salatpreis um das Produkt aus dem ermittelten Koeffizienten b und der Regenmenge x (also um b mal x). Dass dies natürlich nur eine Näherung sein kann und dass der ausschließlich aus diesem Zusammenhang prognostizierte Salatpreis fast immer vom tatsächlichen Preis abweicht, da dieser von vielen weiteren Faktoren abhängt, wird nachfolgend noch dargestellt. Nachfolgend wenden wir uns auch der Frage nach der Bestimmung dieses Koeffizienten anhand der Daten, d.h. seiner Schätzung.

Zuvor weisen wir zunächst darauf hin, dass in der Regel komplexere Regressionsgleichungen verwendet werden. Beinhalten diese mehr als eine Kontrollvariable, so spricht man von einer multivariaten Regression. Auch werden oft sogenannte Transformationen der Variablen verwendet, allen voran die Logarithmierung. Ohne auf die formalen Grundlagen einzugehen, sei lediglich gesagt, dass damit ein gleichbleibender prozentualer (statt einem gleichbleibendem nominalen) Zusammenhang zwischen den so transformierten Variablen unterstellt wird. Für die nachfolgenden Betrachtungen ist allerdings einzig und allein wesentlich, dass die Regressionsgleichung das entweder einfache oder komplexe Modell darstellt, mit dem der Gutachter den Markt mathematisch-formal beschreibt (mit und ohne Verstoß, siehe nachfolgend). Die Koeffizienten dieses Modells werden mittels der Daten geschätzt.¹¹

¹⁰ Genauer handelt es sich hierbei etwas allgemeiner um eine sogenannte affine Beziehung, da auch eine Konstante mit einbezogen wird. Anders ausgedrückt muss damit in diesem einfachen (univariaten) Beispiel die Gerade nicht durch den Ursprung geben („kein Niederschlag gleich Salatpreis von Null“), sondern durch einen beliebigen Achsenabschnittspunkt. Damit müssen auch in diesem Fall zwei Variablen bestimmt bzw. geschätzt werden, daher die Steigung der Geraden und ihr Achsenabschnitt.

¹¹ Für die Rezeption der Regressionsanalyse ist es sicherlich nicht hilfreich, dass der Begriff keine Anhaltspunkte über das zugrundeliegende Vorgehen liefert. In der Regel wird der Begriff der „Regression“ auf den im 19. Jahrhundert lebenden britischen Naturforscher Francis Galton zurückgeführt und bezog sich zunächst schlicht auf die statistische Beschreibung des Phänomens, dass nach sogenannten Ausreißern in den Daten die nachfolgenden Beobachtungen wieder näher am Durchschnitt liegen (so in der Arbeit von Galton in Bezug auf die Größe von Vorfahren und Nachkommen). Intuitiv kann nun ein Bezug zur Regressionsanalyse dadurch hergestellt werden, dass die Regressionsgleichung quasi den „Regelfall“, nicht aber Ausreißer, beschreiben soll. In einer einfachen Regressionsanalyse geht die sogenannte Regressionsgerade, d.h. die geschätzte lineare Beziehung zwischen der

II.ii Einbezug des Verstoßes im Rahmen der Indikatorvariablen-Methode

Die zentrale Regressionsgleichung setzt damit die Kontrollvariablen in Bezug zu der interessierenden abhängigen Variablen. Bevor wir auf die Schätzung dieser Regressionsgleichung bzw. der in ihr enthaltenen Koeffizienten eingehen, soll dies zunächst in Bezug gesetzt werden zur uns interessierenden Frage der Schätzung eines möglichen Verstoßeffektes bzw. der Frage des „Ob“ eines solchen Effektes. Ein an der Schätzung eines Verstoßeffektes interessierter Gutachter wird den Verstoß als eine erklärende Variable, d.h. als eine Kontrollvariable, in sein Modell mit aufnehmen.

Nachfolgend beschränken wir uns ausschließlich auf die in der Praxis wohl wichtigste sogenannte Indikatorenvariablen-Methode (auch Dummy-Methode genannt). Wir stellen diese für den Fall der Umsetzung einer zeitlichen Vergleichsmarktanalyse dar.¹² In diesem Fall werden (noch näher zu beschreibende) Daten für den Verstoßzeitraum und einen Vergleichszeitraum, in der Regel den Zeitraum nach dem Verstoß (ggf. unter Ausklammern eines Nachlaufzeitraums) erhoben und in einer gemeinsamen Regressionsanalyse zusammen betrachtet. Dabei wird in die bereits beschriebene Regressionsgleichung eine sogenannte Indikatorvariable als weitere Kontrollvariable aufgenommen. Diese gibt für jeden betrachteten Zeitpunkt an (d.h. sie „indiziert“), ob dieser dem Verstoßzeitraum oder aber dem Vergleichszeitraum zuzurechnen ist. Typischerweise wird dies so umgesetzt, dass die Indikatorvariable im Verstoßzeitraum den Wert 1 annimmt und im Vergleichszeitraum den Wert 0 (d.h. „1“ bezeichnet den Zustand, dass der Verstoß aktiv war, und „0“ den Zustand, dass dieser nicht mehr aktiv war und auch keine Nachwirkungen mehr bestanden). Auch die Indikatorvariable erhält einen Koeffizienten (so wie die Niederschlagsmenge im Salat-Beispiel). Wir beschreiben nun insgesamt, wie die Koeffizienten, einschließlich dem Koeffizienten dieser Verstoß-Indikatorvariablen, geschätzt werden – und wie insbesondere letzterer zu interpretieren ist.

II.iii Der Schritt der Schätzung der Koeffizienten der Regressionsgleichung (einschließlich des Verstoßeffektes) – sowie die damit einhergehende unvermeidbare Schätzungenauigkeit

Das Regressionsmodell spezifiziert in der beschriebenen Form einen exakten und damit deterministischen, d.h. nicht durch weitere Determinanten oder gar Zufälligkeiten mitbestimmten, Zusammenhang zwischen den aufgenommenen Kontrollvariablen und der zu erklärenden Variablen. Natürlich lassen sich keine Koeffizienten finden, bei denen der dann konkretisierte mathematische Zusammenhang wirklich für alle

abhängigen und der unabhängigen Variablen, auch durch die jeweiligen Mittelpunkte (Durchschnittswerte) der Variablen.

¹² Damit bekommen die zu betrachteten Daten zwangsläufig eine zeitliche Dimension. Die verschiedenen Beobachtungen der gleichen Variablen, so des Transaktionspreises eines bestimmten Produktes, über die Zeit hinweg bezeichnet man auch als Zeitreihe. Werden beispielsweise mehrere Produkte zusammen betrachtet, so um einen durchschnittlichen Aufschlag zu schätzen (siehe nachfolgend), so bezeichnet man diesen Datensatz auch als Panel. Betrachtet man in diesem Fall die unterschiedlichen Preise, ggf. auch über verschiedene Märkte hinweg, zu einem bestimmten Zeitpunkt, so wird dies als Querschnittsanalyse bezeichnet.

Beobachtungen genau zutrifft: Der Salatpreis wird nicht zu jedem Zeitpunkt gleich dem Produkt aus dem Niederschlag und einem festen Koeffizienten sein, selbst wenn die Niederschlagsmenge eine wesentliche Determinante darstellt. Im einfachsten (Textbuch-)Fall einer Regressionsanalyse werden die Koeffizienten nun so gewählt, dass das Modell die vorhandenen Daten allerdings „am besten“ beschreibt, nämlich in dem Sinne, dass die damit prognostizierten Werte im Durchschnitt möglichst wenig von den tatsächlich beobachteten Werten der zu erklärenden Variablen abweichen. Die Koeffizienten werden im einfachsten (Textbuch-)Fall damit so gewählt, dass das mittels der Regressionsgleichung postulierte Modell möglichst geringe Fehler aufweist.¹³

Eine kurze Nebenbemerkung zur Einführung häufig gebrauchter Begriffe: Die Differenzen zwischen den tatsächlichen Werten der abhängigen Variablen und den durch das Modell prognostizierten Werten werden oft als „Residuen“ bezeichnet. Um solchen Residuen bzw. die diesen zugrunde liegenden anderen, in der Regressionsgleichung nicht explizit spezifizierten Determinanten der erklärenden Variablen Rechnung zu tragen, wird in der Regel die Regressionsgleichung um einen sogenannten „Störterm“ oder „Fehlerterm“ erweitert.

Mit der Schätzung haben nun die in der Regressionsgleichung zuvor noch unbestimmten Koeffizienten für die jeweiligen Kontrollvariablen einen Wert erhalten. Wir kommen nun zur Interpretation der geschätzten Koeffizienten. Zuvor scheint uns ein Verweis auf die folgende Terminologie sinnhaft. Die Koeffizienten werden aus den Daten berechnet. Man bezeichnet diesen Vorgang allerdings als Schätzung und die ermittelten Werte als Schätzer, da dies die Auffassung widerspiegelt, dass es gemäß der unterstellten Regressionsgleichung quasi wahre Werte für diese Koeffizienten gäbe, die man allerdings aufgrund beschränkter Datenverfügbarkeit oder aber da das Modell die Wirklichkeit nur unvollständig abbilden kann, nicht präzise ermitteln kann. Damit ist die Ermittlung der Koeffizienten auch immer mit einer Schätzungenauigkeit verbunden. Der statistische Apparat hinter einer Regressionsanalyse dient dazu, Maße für diese Schätzungenauigkeit anzugeben – und das später diskutierte Signifikanzkriterium ist ein solches Maß.

II.iv Der Schritt der ökonomischen Interpretation der Schätzergebnisse

Die ermittelten Koeffizienten beschreiben die Stärke der Beziehung zwischen der zu erklärenden Variablen und der jeweiligen Kontrollvariablen, so dem Salatpreis und dem Niederschlag in unserem einfachen

¹³ Im Textbuch-Fall wird dabei nicht etwa der Durchschnitt (oder die Summe) aller solche Fehler minimiert, sondern die Summe der quadrierten Fehler. Dadurch erhalten größere Fehler ein höheres Gewicht. Der Grund für die Quadrierung liegt nicht, wie oft angegeben, darin, dass so negative und positive Abweichungen auf gleiche Weise betrachtet werden können. Stattdessen ist es das Ziel, einen Koeffizienten zu schätzen, der möglichst genau ist im Sinne einer möglichst geringen Varianz (welche sich eben aus den quadrierten Abweichungen vom Mittelwert zusammensetzt).

Beispiel. Enthält die Regressionsgleichung mehrere Kontrollvariablen, so erfasst der Koeffizient einer Kontrollvariablen die entsprechende Beziehung *ceteris paribus*.¹⁴

Der durch den Koeffizienten beschriebene Zusammenhang wird oft einfach als (partieller) Effekt beschrieben. Diese Beschreibung ist intuitiv, und sie entspricht auch der Natur der Regressionsgleichung, so wie sie üblicherweise in den Textbüchern verwendet wird. Denn dort drückt sie in der Regel auch eine kausale Beziehung aus, ausgehend von einem ökonomischen Modell, so dass auch in einer Prognose davon auszugehen ist, dass sich, *ceteris paribus*, die abhängige Variable bei einer Änderung der Kontrollvariablen (etwa im Falle eines höheren Niederschlags) entsprechend ändert. Die Verwendung des Koeffizienten der Verstoß-Indikatorvariablen für einen Test des „Ob“ eines Schadens sowie auch für eine Quantifizierung der Höhe verlangt ebenfalls eine solche (ideale) kausale Interpretation. Wie wir bei der nachfolgenden Diskussion der Korrektheit des Modells ausführlich darlegen, ist eine solche Idealsituation allerdings praktisch nie gegeben, was sich in der Gefahr verzerrter Schätzungen niederschlägt. Eine solche Verzerrung kann etwa daraus resultieren, dass wichtige Kontrollvariablen vergessen wurden. Sie kann aber auch daran liegen, dass Kontrollvariablen mit aufgenommen wurden, die ökonomisch keinen Sinn ergeben. Wir beschreiben nun genauer die Interpretation der Verstoß-Indikatorvariablen bzw. des geschätzten Verstoßkoeffizienten.

II.v Interpretation des Koeffizienten der Verstoß-Indikatorvariablen (Verstoßeffect)

Wir rekapitulieren kurz: Wir betrachten in diesem Artikel ausschließlich die sogenannte Indikatorvariablen-Methode, bei der, neben anderen Kontrollvariablen, ein zeitlicher Indikator in die Regressionsgleichung aufgenommen wird, der die dem Verstoß(-zeitraum) zuzurechnenden Beobachtungen kennzeichnet. Analog etwa zur Kontrollvariable „Niederschlag“ erhält auch die Indikatorvariable einen Koeffizienten. Mittels der Schätzung der Regressionsgleichung wird nun allen Koeffizienten ein Wert zugewiesen. Sofern wir die „richtige“ Datengrundlage benutzen und wenn das Modell „korrekt“ spezifiziert ist (wobei wir gerade auf diese Punkte nachfolgend eingehen), gibt der Koeffizient der Indikatorvariablen den durchschnittlichen Verstoßeffect an. Sofern wir nominale Preise verwendet haben, gibt der Koeffizient

¹⁴ Man mag nun einwenden, dass dies nicht immer logisch sei, da beispielsweise zwei Kontrollvariablen in einer bestimmten Beziehung zueinanderstehen. Trivialerweise wäre dies der Fall, wenn man gleichzeitig die Niederschlagsmenge in Litern und in Millilitern als Kontrollvariablen aufnehmen würde. Dass dies ökonomischer Unfug ist, zeigt die Schätzung unmittelbar an: Die Koeffizienten der beiden vermeintlichen Kontrollvariablen können gar nicht mathematisch zusammen ermittelt werden. Dies liegt anders, wenn man beispielsweise den Niederschlag in der gleichen Woche (bei der Verwendung wöchentlicher Salatpreise) und in der Vorwoche mit aufnimmt. Auch zwischen diesen Variablen besteht eine Beziehung, d.h. sie sind nicht statistisch unabhängig voneinander. Dennoch können hohe Niederschlagsmengen in einer bestimmten Woche auftreten sowohl im Falle hoher als auch im Falle niedriger Niederschlagsmengen in der Vorwoche. Eine *ceteris paribus* Betrachtung, bei der der Niederschlag der Vorwoche konstant bleibt aber derjenige der aktuellen Woche sich verändert, ist dann in der Tat sinnhaft.

folglich den durchschnittlichen nominalen Verstoßeffekt (bspw. in Euro) an. Wurden die Preise logarithmisch transformiert, so erhalten wir den durchschnittlichen prozentualen Verstoßeffekt.¹⁵

Wir stellen dieses für die Interpretation zentrale Ergebnis noch einmal anders dar, um dadurch auch den Einfluss der weiteren Kontrollvariablen herauszustellen. Dafür lassen wir diese allerdings zunächst weg. Denn es ist natürlich möglich, eine Regressionsgleichung aufzustellen, in der die einzige Kontrollvariable die Verstoß-Indikatorvariable ist. Hierbei würde man implizit annehmen, dass Verstoß- und Vergleichszeitraum – mit Ausnahme des Verstoßes selbst – vollständig miteinander vergleichbar wären, so dass weitere Kontrollvariablen nicht nötig sind. Führt man eine solche degenerierte Regressionsanalyse durch, so erhält man tatsächlich nichts anderes als einen Mittelwertvergleich: Der Koeffizient der Indikatorvariablen entspricht dem Unterschied zwischen den Mittelwerten der abhängigen (Preis-)Variablen in den beiden Zeiträumen. Die Aufnahme von Kontrollvariablen, so etwa des Niederschlags in unserer Salatpreisregression, erlaubt es nun allerdings, hinsichtlich der Unterschiede zwischen den beiden Zeiträumen zu „kontrollieren“ bzw. diese herauszurechnen (sofern die unterstellte Beziehung der tatsächlichen entspricht). Wenn beispielsweise ein höherer Niederschlag in der Regel das Wachstum von Salat fördert und mit der höheren Menge die Preise sinken, sowie wenn es in den beiden Zeiträumen unterschiedlich stark geregnet hätte, so wäre ein reiner Mittelwertvergleich verzerrt. Über die Aufnahme des Niederschlags als Kontrollvariable und die Schätzung des Niederschlags-Koeffizienten soll der Einfluss des Niederschlags herausgerechnet werden, so dass der Verstoßeffekt dann in der Tat *ceteris paribus* den Effekt des Verstoßes widerspiegelt, nicht aber auch den Effekt von anderen Veränderungen über die Zeit. Ob dies die aufgestellte Regressionsgleichung und die geschätzten Koeffizienten aber tatsächlich leisten, wird eine zentrale Frage bei der Bewertung der Korrektheit des Modells sein.

Damit schließen wir die einführende Darstellung der Regressionsanalyse ab. Wir haben die zentrale Bedeutung der Regressionsgleichung herausgestellt und damit die Bedeutung des vom Gutachter unterstellten Modells. Dieses Modell formalisiert die Annahmen des Gutachters darüber, wie seiner Sachkenntnis nach die verwendeten (Preis-)Daten generiert wurden,¹⁶ d.h. was in der betrachteten Industrie und für die betrachteten Zeiträume die wesentlichen Determinanten der Preise sind. Wie wir noch darstellen, verlangt dies zwingend eine hinreichende Kenntnis der Industrie. Dies sollte bei der Bestellung und Instruktion von Sachverständigen berücksichtigt werden. Oft verfügen die Parteigutachter über die

¹⁵ Wie zuvor dargestellt, gehen wir nicht auf die (allerdings mathematisch einfachen) Grundlagen der Logarithmierung ein. Unsere Aussagen sind auch nur näherungsweise richtig, da der in diesem Fall erhaltene Koeffizient wiederum „rücktransformiert“ werden muss. Im Falle einer solchen prozentualen Schätzung ist auch zu fragen, ob sich der Prozentsatz auf den erhöhten Kartellpreis oder aber auf den kontrafaktischen Preis bezieht. Zuletzt darf nicht vergessen werden, dass dies, sofern nicht eine entsprechend angepasste Schätzmethode verwendet wird, einen ungewichteten Durchschnitt darstellt. War der Aufschlag tatsächlich nicht konstant über die Zeit und wurden zu verschiedenen Zeitpunkten verschiedenen Mengen bezogen, so kann die Multiplikation eines solchen ungewichteten Durchschnitts mit dem betroffenen Volumen zu einer Verzerrung führen.

¹⁶ Man spricht daher auch von dem unterstellten „datengenerierenden (statistischen) Prozess“.

nötigen Ressourcen und insbesondere den Zugang zu Industrie-Insidern, während dem gerichtlich bestellten ökonomischen Sachverständigen ggf. sogar die Kontaktaufnahme zu den Parteien untersagt wird. Sowohl mangelnde Kenntnisse der Spezifika einer Industrie aber auch ggf. mangelnder Datenzugang schlagen sich dann aber in einer unzureichenden oder gar falschen Spezifikation der Regressionsanalyse nieder. Die darauf aufbauende statistische Analyse kann dies nicht wettmachen. Sie vermittelt stattdessen ein falsches Bild einer lediglich vermeintlich richtigen Schätzung. Die zentrale Rolle der Wahl der Regressionsgleichung steht folglich auch im Mittelpunkt der nachfolgenden Diskussion der korrekten Methode. Wir diskutieren allerdings zunächst das erste vom BGH formulierte Kriterium einer verlässlichen Datengrundlage.

III Verlässliche Datengrundlage

III.i Vorbemerkung

Der BGH formuliert als erste Voraussetzung an eine zulässige Regressionsanalyse die Verwendung einer verlässlichen Datengrundlage. Nachfolgend leisten wir erste Schritte zur konkreten Ausfüllung dieses Kriteriums. Dabei zeigt sich allerdings bereits bei diesem zunächst offensichtlich erscheinendem Kriterium, dass es keine abschließende Checkliste geben kann, anhand derer die Verlässlichkeit der Datengrundlage geprüft werden könnte, sondern dass hierzu eine ökonomisch fundierte Gesamtschau unabdingbar ist, die den normativen Maßstäben einer Schadensschätzung Rechnung trägt.

Was diese normativen Maßstäbe anbelangt, ist zwischen dem Vollbeweis und der Schätzung zu unterscheiden. Nach dem Maßstab des Vollbeweises des § 286 ZPO ist zu entscheiden, ob die dargelegte Tatsache „wahr oder nicht wahr“ ist. Dies fordert keine absolute Gewissheit oder mathematisch-naturwissenschaftlich zwingende Evidenz¹⁷, aber es ist eine Gewissheit in der richterlichen Überzeugungsbildung nötig, die über eine einfache Wahrscheinlichkeit hinausgeht.¹⁸ Der Richter hat „nur seinem Gewissen unterworfen die Entscheidung zu treffen, ob er die an sich möglichen Zweifel überwinden und sich von einem bestimmten Sachverhalt als wahr überzeugen kann.“¹⁹ Sofern ein

17 Vgl. BGH 26.10.1993, VI ZR 155/92, NJW 1994, 801, 802; Foerste, in: Musielak/Voit, Kommentar zur ZPO, 18. Aufl. 2021, § 286 Rn. 19.

18 BGH 17.2.1970, III ZR 139/67, NJW 1970, 946, 948: „Denn nach § 286 ZPO muß der Richter aufgrund der Beweisaufnahme entscheiden, ob er eine Behauptung für wahr oder nicht für wahr hält, er darf sich also gerade nicht mit einer bloßen Wahrscheinlichkeit beruhigen.“; Foerste, in: Musielak/Voit, Kommentar zur ZPO, 18. Aufl. 2021, § 286 Rn. 18.

19 BGH 17.2.1970, III ZR 139/67, NJW 1970, 946, 948.

Vollbeweis nicht oder nicht mit verhältnismäßigem Aufwand²⁰ möglich ist²¹, kommt hinsichtlich des Kartellschadens eine Schätzung nach § 287 ZPO in Betracht. Dies betrifft sowohl das „Ob“ eines Schadens als auch dessen Höhe²² (wobei für Ersteres die Vermutung des § 33a Abs. 2 GWB streitet²³). Eine Schätzung muss auf Anknüpfungstatsachen beruhen. Sie darf nicht „völlig in der Luft hängen“. Im Rahmen des § 278 ZPO ist es somit Sache des Gerichts, „unter Berücksichtigung aller wesentlichen Umstände die Grenze zu ermitteln, bis zu der für die Schätzung eines Schadens eine ausreichende Grundlage vorhanden ist“.²⁴ Nach der Rspr. des BGH muss eine Schätzung nachvollziehbar sein²⁵, und sie muss einen

²⁰ Foerste, in: Musielak/Voit, ZPO, 18. Aufl. 2021, § 287 Rn. 6: „Eine Schätzung setzt daher voraus, dass ein Vollbeweis vom Geschädigten nicht angetreten wird (ggf. Nachfrage, § 139), scheitert oder mit Schwierigkeiten verbunden ist, die außer Verhältnis zur Höhe des streitigen Anspruchs stehen. Demgegenüber lässt die Rspr. teilweise ausreichen, dass der Tatrichter der Beweisführung eine ungünstige Prognose stellt“; dazu BGH 13.6.1996, IX ZR 233/95, NJW 1996, 2501, 2502: „Im Unterschied zu den Anforderungen des § 286 Abs. 1 ZPO kann er von einer weiteren Beweisaufnahme absehen, wenn ihm bereits hinreichende Grundlagen für ein Wahrscheinlichkeitsurteil zur Verfügung stehen ... Das hat für den Geschädigten eine Beweiserleichterung zur Folge, bedeutet aber auf der anderen Seite auch, daß der Richter die Tatsachen nicht weiter aufzuklären braucht, wenn der Nachweis bisher nicht einmal ansatzweise geführt und bereits hinreichend erkennbar ist, daß die noch zur Verfügung stehenden Beweise nicht ausreichen werden, die Behauptung des Kl. mit Wahrscheinlichkeit zu belegen.“

²¹ BVerfG 8.12.2009, 1 BvR 3041/06, NJW 2010, 1870, 1871: „Die Norm soll in erster Linie den Geschädigten entlasten, das Gericht hingegen nur bedingt ... Für eine Schätzung ist deshalb dann kein Raum, wenn das Fachgericht den Schaden ohne Schwierigkeiten exakt berechnen kann.“; Leisse, GRUR 1988, 88, 90. Zum Kartellschaden insoweit Logemann, Der kartellrechtliche Schadensersatz, 2009, S. 457 f.: „Eine konkrete Berechnung der verursachten Marktschäden setzt eine nicht bloß vage Vorstellung von den hypothetischen Ergebnissen auf einem Wettbewerbsmarkt voraus. Die damit verbundenen Unwägbarkeiten sind angesichts der vielen potentiellen Einflussfaktoren entsprechend groß und gehen im Zweifel zulasten des Geschädigten, trägt dieser doch die Darlegungs- und Beweislast für die Schadenshöhe.“

²² BGH 12.7.2016, KZR 25/14 – Lottoblock II.

²³ Zur zeitlichen Anwendbarkeit der neuen Norm siehe § 186 Abs. 3 GWB.

²⁴ BGH 16.12.1963, III ZR 47/63, NJW 1964, 589, 590.

²⁵ BGH 22.5.1984, III ZR 18/83, BGHZ 91, 243, 256 f.; BGH 25.2.1997, VI ZR 101/96, NJW 1997, 1640, 1641 m. w. N.; OLG München 23.2.2012 Az.: U 3365/11 (Kart), juris-Rn. 62 – Schätzung eines billigen Entgelts: „Nachdem somit keine brauchbaren Anknüpfungstatsachen gegeben sind, ist eine Schätzung des Umfangs einer angemessenen Erhöhung der Trassenentgelte durch die Beklagte unzulässig, weil sie völlig in der Luft hängen würde (vgl. Greger in: Zöller, a.a.O., §§ 287 ZPO Rn. 4). Die Auffassung der Beklagten, dem Gericht lägen sämtliche für eine Schätzung erforderlichen Informationen vor [...], findet in den Verfahrensakten offensichtlich keine Grundlage.“; LG Dortmund 1.4.2004, 13 O 55/02 Kart, WuW/E DE-R 1352, 1354 – Vitaminpreise Dortmund: „Für die Feststellung der Schadenshöhe gilt § 287 ZPO, da ein Schaden eingetreten ist und ausreichende Anhaltspunkte zur näheren Bestimmung dieses Schadens gegeben sind.“

ausreichenden Realitätsbezug aufweisen.²⁶ Insoweit gilt nach Ansicht des BGH²⁷: „Entscheidend ist ... nicht der Grad der Überzeugung des Sachverständigen, sondern die persönliche Überzeugung des Tatrichters [...]“. Hierfür kommt es auf eine Gesamtschau des Beweis- und Streitstoffs an.²⁸ Mit Blick auf § 286 ZPO hat. Das „schließt [...] generalisierende Erwägungen zum Beweiswert eines Beweismittels nicht aus [...] solange nur eine einzelfallbezogene Gesamtabwägung stattfindet.“

Dies zugrunde gelegt, kann es sich bei der Frage nach der Datenverlässlichkeit für eine Regressionsanalyse nicht darum handeln, dass die Daten, auf die die Regression sich stützt, lückenlos fehlerfrei sein müssen. Zum einen würde die Überprüfung der Fehlerlosigkeit vielfach vor dem Problem stehen, dass sich die für die Regressionsanalyse nötige Datenbasis nicht automatisch erschließt. Dies gilt, wie wir darstellen, sogar für die Verwendung von Transaktionspreisen. Es ist oft gerade bei einem zeitlichen Vergleich nicht offensichtlich, welche Preise verwendet werden sollen, um diesen unverzerrt durchzuführen. Darüber hinaus muss gesehen werden, dass der Einsatz einer Regressionsanalyse, wie jede Schätzung, auch bei fehlerfreien Daten mit Schätzungenauigkeiten verbunden ist. Die Feststellung irgendeiner Fehlerhaftigkeit einzelner Daten lässt daher noch nicht den Schluss zu, dass die Analyseergebnisse insgesamt unbrauchbar sind. Eine prinzipielle Forderung nach lückenlos fehlerfreien Daten könnte daher das Institut der Regressionsanalyse insgesamt entwerten.²⁹

Wie wir an Beispielen illustrieren, ist die Frage, welche Fehler oder Ungenauigkeiten noch vertretbar sind, wiederum keine Frage, die sich mit einer Checkliste oder aber einer (statistischen) Kenngröße beantworten ließe (bspw. dass mindestens x % der Preise fehlerfrei sein müssten). Auch sie verlangt ein genaueres Eingehen auf die (ökonomischen) Gründe und möglichen Implikationen dieser Fehler und Ungenauigkeiten, die im Rahmen des Schadensersatzprozesses nach § 287 ZPO das Gewicht der

²⁶ Inderst/Thomas, Schadensersatz bei Kartellverstößen, 2. Aufl. 2018, S. 258; Rauh, in: WRP 2012, 173, 179: „Dabei ist es Aufgabe des Klägers, für eine ausreichende Schätzungsgrundlage zu sorgen, indem er die Anknüpfungstatsachen darlegt und ggf. beweist. Für die Kartellschadensersatzklage bedeutet dies, dass der Kläger Anhaltspunkte aufzeigen muss, die auf für ihn günstigere Marktbedingungen ohne Wettbewerbsbeschränkungen hinweisen.“; Meeßen, Der Anspruch auf Schadensersatz bei Verstößen gegen EU-Kartellrecht, 2011, S. 414: „Die Schätzung kommt nur in Betracht, wenn hierfür vom Kläger greifbare Anhaltspunkte vorgetragen werden.“

²⁷ BGH 26.10.1993, VI ZR 155/92, NJW 1994, 801, 802.

²⁸ Schellhammer in Schellhammer, Zivilprozess, 16. Aufl. 2020, Rn. 562 (Juris): „Seine Überzeugung gewinnt das Gericht aus einem Gesamturteil über Beweis- und Streitstoff. Es wägt alle Argumente, die für und wider das Beweisthema sprechen, gegeneinander ab.“ (Fn. ausgelassen).

²⁹ Hinweis: Die Frage der Fehlerlosigkeit der Datengrundlage im hiesigen Beitrag bezieht sich auf die Verwendung zwecks Konstruktion eines wettbewerblichen counterfactual, i.e. der Ermittlung hypothetischer Wettbewerbspreise mithilfe der Regressionsanalyse. Sie behandelt nicht die Anforderungen an den Nachweis tatsächlich gezahlter Preise, etwa durch Vorlage entsprechender Zahlungsbelege bei einzelnen Bezugsvorgängen. Diese Nachweisebene hat nichts mit der Regressionsanalyse zu tun. Auch zur Frage, inwieweit einzelne Bezugsvorgänge eigene Ansprüche bilden, die eines jeweils separaten Schadensnachweises bedürfen, behandelt der vorliegende Beitrag nicht, siehe dazu Thole, NZKart 2020, 227 ff.

Anknüpfungstatsache betrifft und die vom Gericht daher vorgenommen werden sollte. Unsere nachfolgenden Ausführungen hierzu beschränken sich zunächst auf die zu erklärende Variable. Wir nehmen hierbei wieder konkret an, dass es sich dabei um Transaktionspreise handelt.³⁰ Abschließend gehen wir kurz auf die Kontrollvariablen ein.

III.ii Aufbau des Datengerüsts für die erklärende Variable (Preis)

Das einfache (Textbuch)-Modell sieht eine einzige Zeitreihe vor, die für jeden betrachteten Zeitpunkt den (ggf. durchschnittlichen) Transaktionspreis eines in seiner Beschaffenheit gleichbleibenden Produktes angibt. Damit eine statistisch belastbare Regressionsanalyse durchgeführt werden kann, muss die Zeitreihe (mit dem einen Produkt) hinreichend lang sein, d.h. es müssen hinreichend viele Preissetzungen stattgefunden haben. In vielen Fällen werden die tatsächlichen Transaktionsdaten allerdings erheblich von dieser Vorgabe abweichen. Zum einen mag es sich bei dem vom Verstoß betroffenen Volumen nicht um ein Produkt allein handeln, sondern um verschiedene, differenzierte Produkte. Zum anderen kann es sein, dass die vom Kartellanten abgesetzten Produkte auch nicht stets gleich geblieben sind: Bestehende Produkte wurden ggf. verändert oder aber ihr Verkauf wurde ausgesetzt, während neue Produkte eingeführt wurden. Zuletzt mag gerade bei Transaktionen zwischen zwei Unternehmen der Preis nur in längeren Abständen angepasst worden sein, so beispielsweise bei jährlichen Verhandlungen, was die Zahl von (unabhängigen) Beobachtungen erheblich einschränkt.

Es liegt nun an dem Gutachter, sich erst eine Datengrundlage, dh. das Datengerüst, zu schaffen. Wurde beispielsweise ein bestimmtes Produkt nur im Verstoßzeitraum verkauft, nicht aber im herangezogenen Vergleichszeitraum, so kann es immer noch möglich sein, ein entsprechendes vergleichbares Nachfolgeprodukt zu identifizieren. Unterscheidet sich dieses allerdings, so muss die Regressionsgleichung hinsichtlich dieses Unterschieds „kontrollieren“, nämlich durch die Aufnahme von produktspezifischen Informationen.³¹ Der Gutachter mag ferner versuchen, die Analyse nur auf die Produkte zu beschränken, die hinreichend lange sowohl im Verstoß- wie auch im Vergleichszeitraum abgesetzt wurden, so dass für den Vergleich genügend Beobachtungen in beiden Zeiträumen verfügbar sind. Eine solche Selektion läuft aber die Gefahr, dass die so gewählten Produkte nicht hinreichend repräsentativ sind, beispielsweise da im Vergleichszeitraum Preiserhöhungen über die Einführung neuer Produkte umgesetzt wurden.

³⁰ Mit zusätzlichen Informationen ließe sich ein Verstoßeffekt beispielsweise auch aus einer Regressionsanalyse ableiten, bei der die abhängige Variable die Marge des vom Verstoß betroffenen Produktes ist.

³¹ Die Aufnahme solcher Informationen ist quasi unerlässlich, wenn die betroffenen Produkte für jeden einzelnen Abnehmer individuell konfiguriert werden. Gelingt es aber, entsprechende Attribute hinreichend zu erfassen, so kann eine Vergleichbarkeit sowohl im Querschnitt, d.h. für Transaktionen zum gleichen Zeitpunkt, wie auch im Längsschnitt, d.h. über die Zeit hinweg, hergestellt werden. Gerade bei (Marken-)Produkten, die letztlich an Endkonsumenten verkauft werden, ist die Erfassung derjenigen Attribute, die bsp. die Wertschätzung der Kunden bestimmt, allerdings oft schwierig, schlicht da sie sich nicht einfach an „harten“ Fakten, wie den Inhaltsstoffen oder aber der Leistung, ablesen lassen.

Betrachtet man lediglich die letztendlich durchgeführte Regressionsanalyse mitsamt ihrer Datengrundlage, so würde ein Blick allein auf die lückenlose Fehlerfreiheit dieser Daten damit zu kurz greifen. Stattdessen muss die Frage nach der Verlässlichkeit der Daten die gesamte Konstruktion des verwendeten Datensatzes mitberücksichtigen, einschließlich der Selektion bestimmter Produkte oder aber des Bildens von Zeitreihen aus verschiedenen Produkten. Es liegt auf der Hand, dass diese Prüfung eine hinreichende Kenntnis des spezifischen Sachverhalts, u.a. über die Beschaffenheit der Produkte, voraussetzt. Die Begründung der Konstruktion des Datensatzes muss letztlich eine ökonomische sein.³² Sie muss der Realität der Industrie gerecht werden, was gerichtlich im Rahmen der Schätzung berücksichtigt werden muss.

III.iii Datenfehler

Auch hier stellen wir zuerst auf die zu erklärende Variable, hier den Transaktionspreis, ab. Es ist offensichtlich, dass eine fehlerfreie Datenbasis zu bevorzugen ist. Allerdings liegen solche fehlerfreien Daten nicht immer vor. Welcher Transaktionspreis tatsächlich gezahlt wurde, mag sich beispielsweise nur aus der Finanzbuchhaltung erschließen, da selbst auf einer Rechnung nicht verzeichnet ist, ob der Käufer einen gewährten Skonto tatsächlich gezogen hat. Weder aus der Rechnung noch aus der Finanzbuchhaltung lässt sich aber ggf. einfach ableiten, in welchem Umfang und bezogen auf welche Transaktionen nachträgliche Rabatte gewährt wurden. Daten der Finanzbuchhaltung mögen weiterhin zwar genaue(re) Transaktionsbeträge liefern, allerdings dürften dort wenig weitere Daten zu der betroffenen Transaktion vorliegen. Solche Daten finden sich stattdessen in der Regel im Warenwirtschaftssystem – und die genaue Beschreibung der Transaktion ist gerade bei nicht standardisierten (und insbesondere unverpackten) Gütern wesentlich, um in einer Regressionsanalyse eine Vergleichbarkeit sicherzustellen. Eine prinzipielle Forderung nach einer Datengrundlage für die Regressionsanalyse, die zwingend fehlerfrei die tatsächlich gezahlten Transaktionspreise abbildet, wäre bereits angesichts solcher exemplarischer Komplikationen in der Praxis nicht sinnvoll auch angesichts des Umstands, dass jede Aufschlagsschätzung mit Schätzungenauigkeiten verbunden ist.³³

Sofern beispielsweise die Gutachter in einem Verfahren begründet einzelne Fehler und Ungenauigkeiten in der Datengrundlage einer anderen Partei aufzeigen, sollte dies nicht per se ein Grund sein, um die Analyse unter dem Gesichtspunkt ökonomischer Fehlerhaftigkeit zu verwerfen. Wiegen die Einwände hinreichend schwer, so liegt es zunächst auch nahe, die Analyse ohne die beanstandeten Daten zu wiederholen. In der Regel sollte dies mit geringem Aufwand durchführbar sein, da, um es kurz zu fassen, der gleiche Programmcode auf eine nun veränderte Datenbasis angewendet wird.

³² Während auch Robustheitsanalysen hilfreich sind, verlangen diese allerdings in der Regel eine Erweiterung des Datensatzes und damit ggf. die Aufbereitung weiterer (Produkt-)Preisreihen.

³³ Erneut ist zu beachten, dass sich dieser Beitrag nur mit der Datengrundlage für die Regressionsanalyse befasst. Davon zu trennen ist die Frage, wie der Kläger hinsichtlich der einzelnen Bezugsvorgänge beweist, welchen Preis er tatsächlich gezahlt hat. Dazu bereits oben Fn. 29.

Nachfolgend diskutieren wir an Beispielen, wie sich der Nutzen einer Verbesserung der Datengrundlage bereits a priori abschätzen lässt – und damit auch, ob und in welchem Ausmaß eine unvollständige und fehlerhafte Datengrundlage überhaupt das Ergebnis wahrscheinlich verzerrt. Hierbei nehmen wir die folgenden Unterscheidungen vor, die über die Beispiele hinaus hilfreich erscheinen. In Anlehnung an die ökonomische Literatur sprechen wir dabei von „Messfehlern“ und unterscheiden sodann wie folgt: i) Prinzipiell unsystematische Messfehler, ii) in Hinblick auf die verwendete Methode unsystematische Messfehler und iii) systematische Messfehler.

Wir stellen den Beispielen die folgenden Bemerkungen voran, wodurch vermieden werden soll, dass dies als Plädoyer für unsaubere Datenarbeit missverstanden wird. Zunächst verlangt eine Schadensschätzung als Anknüpfungstatsache im Sinne von § 278 ZPO nach einer dokumentierten Erfüllung der gutachterlichen Sorgfalt bei der Sammlung und Aufbereitung der Daten. Eine Dokumentation ermöglicht es dem Gericht, zu ermitteln, ob und wo ggf. systematische Fehler eingeflossen sein können, die dann die Regressionsanalyse verzerren.³⁴ Ferner obliegt es dem Gutachter, darzustellen, warum in den Daten enthaltene Ungenauigkeiten und Fehler entweder nur sehr beschränkt auftreten oder aber nicht zu einer Verzerrung der Ergebnisse führen. Dadurch kann das Gericht ebenfalls die Verlässlichkeit der Regressionsanalyse in diesem Punkt beurteilen.

Prinzipiell unsystematische Messfehler: Hierbei handelt es sich um Fehler und Ungenauigkeiten, die quasi zufällig auftreten. Ein Beispiel mag die Verwendung von Preisen sein, die aus eingescannten Rechnungen mittels OCR-Verfahren („Optical Character Recognition“) ausgelesen wurden. Hier ist zunächst nicht zu erwarten, dass falsch ausgelesene Zahlen entweder systematisch zu hoch oder zu niedrig sind, oder aber, dass es dabei systematische Unterschiede zwischen Verstoß- und Vergleichszeitraum gibt. Betrachten wir nun isoliert die Schätzung des Verstoßeffektes (d.h. des Koeffizienten der Verstoß-Indikatorvariablen), so erinnern wir daran, dass dieser eine residuale Größe darstellt: Im einfachsten Verfahren stellt der so geschätzte Verstoßeffekt den Unterschied zwischen den Mittelwerten im Verstoß- und Vergleichszeitraum dar, nachdem über die aufgenommenen Kontrollvariablen andere Unterschiede zwischen den beiden Zeitperioden herausgerechnet wurden (so zumindest die Idealvorstellung). Auf einen solchen Mittelwertvergleich haben unsystematische Messfehler keinen systematisch verzerrenden Einfluss. Dies bedeutet natürlich nicht, dass gerade bei einer geringen Anzahl von Beobachtungen die Fehler dennoch das Ergebnis mit einer gewissen Wahrscheinlichkeit beeinflussen, allerdings nicht systematisch. Ist die Anzahl der Beobachtungen hinreichend hoch, so ist mit hoher Wahrscheinlichkeit der Einfluss unsystematischer

³⁴ Es ist daher sinnvoll wenn, ausgehend von Rohdaten, die weiteren Schritte der Datenarbeit nicht etwa an einer Excel-Tabelle durchgeführt werden, da die Änderungen nicht nachgehalten werden können, sondern im Rahmen eines statistischen Softwarepakets geschehen und somit im Programmcode dokumentiert sind. Oft ist es nicht der Fall, dass die entsprechenden Codes zwischen den Parteien ausgetauscht werden – und jede Partei mag sich im Nachteil wähnen, wenn sie einseitig eine solche Transparenz herstellt.

Messfehler auf die Höhe des geschätzten Verstoß-Koeffizienten gering. Die nachfolgend diskutierte Signifikanz gibt Auskunft über diese Schätzunschärfe.³⁵

In Hinblick auf die verwendete Methode unsystematische Messfehler: Auch hier betrachten wir wieder ein Beispiel. Der Gutachter mag für die Regressionsanalyse Daten des Warenwirtschaftssystems verwendet haben, in dem zwar nötige Informationen über die Beschaffenheit des gelieferten Produkts enthalten sind, aus dem allerdings nur ein Rechnungsbetrag ohne Skonto hervorgeht. Der Gutachter hat allerdings hinreichende Ansatzpunkte dafür, dass sich weder die Höhe des Skontos noch die Häufigkeit des tatsächlichen Abzugs zwischen den beiden Zeitperioden systematisch unterscheiden, und er kann dies entsprechend begründen. Führt der Gutachter nun die Regressionsanalyse so durch, dass ein prozentualer Aufschlag geschätzt wird, so ist zunächst davon auszugehen, dass trotz der Verwendung nominal zu hoher Beträge, d.h. der Rechnungsbeträge ohne Skonto, der Aufschlag nicht systematisch verzerrt ist.³⁶ Würde man aber einen räumlichen Vergleich durchführen anhand von Rechnungsbeträgen, bei denen der Skonto noch nicht enthalten ist, und hätte man keine Anhaltspunkte dafür, dass dieser zwischen den betrachteten Regionen (oder Ländern) gleich ist, so würde dieser Datenfehler bei dieser Vergleichsmethode zu einer Verzerrung führen.

Dieses Beispiel soll lediglich illustrieren, dass auch nicht rein zufällige, daher eigentlich systematische Messfehler, nicht zwangsläufig die Schätzung des Verstoßeffektes verzerren müssen. Dies hängt von der verwendeten Methode ab und von den sonstigen Informationen über beispielsweise nicht hinreichend in den Daten enthaltene Preiskomponenten. Auch hier wäre es problematisch, ohne die Berücksichtigung dieser zusätzlichen Faktoren die Analyse allein aufgrund solcher Messfehler zu verwerfen. Die Berücksichtigung der Analyseergebnisse verlangt dann aber von dem Gericht, Klarheit darüber zu erzielen, welcher Art der Messfehler ist und inwiefern eine Beeinflussung des Ergebnisses stattgefunden haben bzw. ausgeschlossen werden kann.

Systematische, verzerrende Messfehler: Es ist schwer, im Hinblick auf die abhängige Variable, hier den Preis, Ungenauigkeiten oder Messfehler ausfindig zu machen, die unabhängig von der verwendeten Methode stets zu einer Verzerrung des geschätzten Verstoßeffektes führen. Solches wäre aber der Fall,

³⁵ Daher: Messfehler wirken sich auf die Signifikanz des Ergebnisses aus. Intuitiv führen diese zufälligen Variationen in den Preisen dazu, dass die statistische Signifikanz sinkt und das Ergebnis weniger belastbar wird. Betreffen die Messfehler nur die abhängige Variable, so sind auch die Koeffizienten der anderen Kontrollvariablen unverzerrt. Dies ist nicht mehr der Fall, sofern selbst unsystematische Messfehler die Kontrollvariablen betreffen. In Hinblick auf den Verstoß mag man hier auch an Messfehler in der Indikatorvariablen denken, d.h. an eine zeitliche Fehlklassifikation. Selbst wenn der Bußgeldbescheid den jeweiligen Verstoßzeitraum anzeigt, mag dieser entweder bereits vor dem beußten Zeitraum existiert haben oder aber in Teilen noch darüber hinaus angehalten haben. Hier bietet es sich allerdings an, ähnlich wie bei der Behandlung von Nachwirkungen, ggf. ein Zeitfenster nach Ende des Verstoßes aus der Betrachtung auszuschließen.

³⁶ Formal ergibt sich dies daraus, dass mittels der Logarithmierung der konstante prozentuale Abschlag zu einer separierbaren additiven Größe wird.

wenn der entsprechende Messfehler sich systematisch zwischen dem Verstoß- und dem Vergleichszeitraum unterscheidet. In der Praxis dürfte es indes häufiger vorkommen, dass eine solche systematische Verzerrung im Hinblick auf die Kontrollvariablen auftritt. Eine solche Gefahr kann insbesondere bei Verwendung von interner Information der am Verstoß beteiligten Unternehmen auftreten. Gerade bei einem vielschichtigen und länger anhaltenden Verstoß können neben den Preisen auch andere Variablen vom Verstoß beeinflusst sein. Wir illustrieren dies im folgenden Abschnitt am Beispiel von Kosten.

III.iv Systematische, verstoßbedingte Beeinträchtigungen der Kontrollvariablen

Die vorherige Problematik überschneidet sich mit der Frage nach der Korrektheit der verwendeten Methode. Insgesamt dürfte aus den bisherigen Ausführungen zu den Daten deutlich geworden sein, dass die Frage der Verlässlichkeit der Datengrundlagen mit der Frage der Wahl und damit der Korrektheit der Methode eng verwoben ist. Dies gilt auch in Hinblick auf die Kontrollvariablen.

Wir beschränken uns nachfolgend wieder auf ein Beispiel. Hierfür betrachten wir die Aufnahme von Kosten als eine Kontrollvariable. Kosten stellen in der Tat eine naheliegende Kontrollvariable dar, sofern die für den Gutachter messbare Kostenkomponente tatsächlich eine ökonomisch hinreichende Rolle spielt.³⁷ Es ist naheliegend, soweit verfügbar, Informationen aus der unternehmensinternen Kostenrechnung zu verwenden. Hier können sich allerdings systematische Probleme ergeben, die ein Gutachter ausschließen muss. Hier zeigt sich exemplarisch, dass auch die Verlässlichkeit von fehlerfrei vom Gutachter erhobenen bzw. übernommenen Daten zu prüfen ist.

So mag es zwar möglich sein, aus der unternehmensinternen Datenbank quasi fehlerfrei die intern den betroffenen Produkten zugeschriebenen Kosten zu extrahieren. Diese reflektieren allerdings stets unternehmensinterne Entscheidungen, beispielsweise wie mit Gemeinkosten umgegangen wird. Dies gilt ebenso für Forschungs- und Entwicklungskosten. Sofern diese Kostenbestandteile nicht preisbestimmend waren, schmälert dies die Rolle als Kontrollvariable. Wesentlich ist wieder die Frage, ob es zwischen den beiden Zeiträumen systematische Unterschiede gibt. Zu prüfen wäre damit zumindest, ob nicht zwischen dem Verstoß- und dem Vergleichszeitraum Änderungen in der entsprechenden Kostenrechnung stattgefunden haben. Aber selbst die tatsächlichen Kosten können durch den Verstoß beeinflusst sein. So können sich Kartelle insgesamt negativ auf die Effizienz auswirken aufgrund der verringerten Preisdisziplin. Diese Feststellungen gelten ebenso für Variablen der Nachfrage, durch die Preisänderungen erklärt werden sollen. Auch diese können durch den Verstoß verzerrt sein.³⁸ Selbst fehlerfreie Daten (hier zu den tatsächlichen Kosten) mögen deshalb nicht Bestandteil einer verlässlichen Datengrundlage sein.

³⁷ Ist dies nicht der Fall, kann dies, wie stets bei der Aufnahme von ökonomisch nicht relevanten Kontrollvariablen, zu sogenannten Scheinkorrelationen führen, die die Schätzung des Verstoßeffektes verzerren können.

³⁸ Dies gilt selbstverständlich nicht mehr, wenn die entsprechenden Variablen auf makroökonomischer Ebene ansetzen, so etwa an der Entwicklung des Bruttoinlandsproduktes oder aber der Konsumbereitschaft der Verbraucher

Damit schließen wir unsere Ausführungen zur Prüfung der Verlässlichkeit der Datengrundlage ab. Unsere Ausführungen hierzu sind einerseits rein exemplarisch, andererseits zeigen sie auch eine Systematik auf, so im Hinblick auf die Unterscheidung von Datenfehlern und Ungenauigkeiten, die eher systematisch oder unsystematisch sind. Wird die Regressionsanalyse mit hinreichend vielen Daten durchgeführt, sollte der Fokus vor allem darauf liegen, inwieweit solche Messfehler zu einer systematischen Verzerrung führen können. Bei wenigen Daten können dagegen auch unsystematische Fehler das Ergebnis mit größerer Wahrscheinlichkeit verzerren – allerdings ist in diesem Falle wohl auch eine eingehendere Prüfung der geringeren Anzahl an Beobachtungen vertretbar. Wir haben auch herausgestellt, dass ein Gutachter den für die Regression verwendeten Datensatz erst konstruieren muss, so zum Beispiel durch die Selektion der aufgenommenen Produkte. Ob dies eine verlässliche Datengrundlage darstellt, kann nicht an einfachen Gütekriterien abgelesen werden, sondern ergibt sich unter Berücksichtigung des vorliegenden Sachverhalts und in diesem Falle insbesondere aus einem Verständnis der betrachteten Produkte. Anhand von Kosten wie auch Nachfragevariablen haben wir abschließend dargestellt, dass auch eine Prüfung der Verlässlichkeit der Kontrollvariablen ein solches ökonomisches Verständnis des Sachverhalts, einschließlich des Verstoßes, voraussetzt. Die Prüfung der Verlässlichkeit der Datengrundlage kann sich damit gerade nicht im Überprüfen von Datenfehlern erschöpfen.

IV Korrekte Methode

IV.i Zur Bedeutung der Methode

Wie bereits einleitend ausgeführt, können wir uns in diesem Artikel nicht umfassend zur Wahl der korrekten Methode auslassen. Diese Frage hat viele Facetten. Während wir zuvor stets vom Textbuch-Fall gesprochen haben, weichen die tatsächlich verwendeten Methoden - insbesondere in der Wissenschaft, aber mittlerweile auch in der relevanten gutachterlichen Praxis - oft davon ab, indem sie dem Sachverhalt besser gerecht werdende statistische Annahmen treffen. Dies betrifft insbesondere Annahmen zur statistischen Abhängigkeit der einzelnen Beobachtungen. Ein Beispiel: Werden in einer Regressionsanalyse Daten unterschiedlicher Käufer verwendet, so ist davon auszugehen, dass die vom selben Käufer ausgehandelten Preise auch über die Zeit bestimmten eigenen Einflüssen unterliegen. Diese Abhängigkeit kann bei der statistischen Umsetzung der Regressionsanalyse berücksichtigt werden. Um zu prüfen, ob die hierbei gemachten (statistischen) Annahmen erfüllt sind, können entsprechende statistische Tests verwendet werden.³⁹ Eine fehlerhafte Wahl führt hier oft zu einer fehlerhaften Schätzung der nachfolgend

insgesamt. Denn der Verstoß dürfte auf solche Variablen keinen Einfluss gehabt haben. Allerdings ist hierbei wiederum die crux, dass diese nur sehr bedingt Nachfrageveränderungen für das betroffene Produkte einfangen.

³⁹ Grob gesprochen setzen diese Annahmen daran an, wie die zuvor erwähnten Stör- oder Fehlerterme beschaffen (genauer: verteilt) sein sollen.

vorgestellten Signifikanz und damit womöglich zu einer Überschätzung der statistischen Belastbarkeit des Ergebnisses. Darauf gehen wir später noch ein.

Wir stellen nachfolgend bei der Diskussion der Korrektheit der Methode allerdings die Frage in den Vordergrund, ob die Schätzung des Verstoßeffektes verzerrt ist. Dies bedeutet, dass eine fehlerhafte Schätzung nicht lediglich an einer ungenügenden Schätzgenauigkeit liegt, die etwa durch eine geringe Anzahl von Beobachtungen bedingt ist. Auch hier gehen wir wieder so vor, dass wir zwar einzelne Beispiele in den Vordergrund stellen, dabei aber auch eine darüber hinausgehende Systematik anwenden. Wir nehmen bereits wieder ein zentrales Ergebnis vorweg: Gerade die Korrektheit der Methode lässt sich nicht (allein) an ausgewählten Maßen ablesen und insbesondere nicht mittels einer Checkliste bestimmen. Solche (statistischen) Gütemaße können lediglich eine letztlich ökonomische Gesamtbeurteilung ergänzen und unterstützen. Die von uns gewählten Beispiele zeigen insbesondere auch, dass in die Bewertung der Korrektheit der Methode stets auch der vorhandene Sachverhalt und damit auch die Natur des Verstoßes einzufließen hat. Die Vorstellung, es gäbe eine Methode, die quasi agnostisch, d.h. ohne weitere Kenntnisse des Sachverhalts und des Verstoßes stets korrekte Ergebnisse liefern würde, wäre nicht überzeugend.

Hierzu gehen wir nachfolgend zunächst auf die Wahl der Kontrollvariablen ein sowie dann auf die Spezifizierung der Beziehung zwischen Kontrollvariablen und abhängiger Variable (hier dem Preis). Hier gestalten wir unsere Ausführungen derart, dass sie allgemein verständlich bleiben und, wie bereits dargestellt, die Notwendigkeit einer ökonomischen statt lediglich statistischen Betrachtung betonen. Abschließend gehen wir noch kurz auf die erwähnten (statistischen) Gütemaße ein. Nochmals: Eine ökonomische und eine allein auf statistische Gütemaße fokussierte Betrachtungsweise stellen hier keineswegs einen Widerspruch dar, sondern sie ergänzen sich. Allerdings kann ein erhebliches Fehlerpotential darin liegen, dass sich ein Gutachter lediglich auf die statistischen Gütemaße stützt, insbesondere wenn er nicht die Möglichkeit hat, sich hinreichend mit den Spezifika der Industrie, der betroffenen Produkte und dem Verstoß zu befassen.

IV.ii Ökonomische Prüfung der Sinnhaftigkeit der (geschätzten) Koeffizienten der Kontrollvariablen

Wir beginnen wieder mit einem Beispiel. Hierfür nehmen wir an, dass es im Verstoßzeitraum einen erheblichen Anstieg der Energiekosten gegeben hat. Im selben Zeitraum stiegen die Preise der betroffenen Produkte deutlich an. Nimmt man nun die Energiekosten als Kontrollvariable mit auf, und ist der betrachtete Zeitraum insgesamt relativ kurz, d.h. einschließlich des Vergleichszeitraums, so kann dies dazu führen, dass der Koeffizient der Kontrollvariablen Energiekosten sehr hoch ausfällt – eben da dies das gleichlaufend ansteigende Verhalten von Energiekosten und Transaktionspreisen im Verstoßzeitraum widerspiegelt. Damit nimmt diese Kontrollvariable einen (ggf. erheblichen) Teil des Preisanstiegs mit auf. Verglichen mit einer Regressionsanalyse, bei der die zugrundeliegende Regressionsgleichung keine Energiekosten berücksichtigt, würde dann im ersten Fall der geschätzte Verstoßeffekt (ggf. deutlich) niedriger ausfallen.

Es wäre nun problematisch, wenn man in diesem Beispiel eine Auswahl zwischen den beiden Modellen, d.h. mit und ohne Energiekosten, allein auf Basis statistischer Kriterien der Modellselektion vornehmen wollte. Stattdessen ist es notwendig, diese Wahl (auch) nach ökonomischen Maßstäben zu treffen. Konkret bedeutet dies u.a., dass sich der Gutachter damit auseinandersetzen sollte, ob erstens Energiekosten überhaupt eine nicht vernachlässigbare Rolle für die Produktion der betroffenen Produkte spielen und ob zweitens die Höhe des Koeffizienten angesichts der tatsächlichen Rolle der Energiekosten im vorliegenden Fall ökonomisch sinnvoll ist. Gleiches gilt auch für andere wesentliche Kontrollvariablen. Beispielsweise mag die Regressionsgleichung Kontrollvariablen berücksichtigen, die Veränderungen der Nachfrage, die sich dann auf die Preise auswirken sollen, abbilden. Sofern die Schätzung diesen eine wesentliche Rolle zuweist, sollte sowohl geprüft werden, ob dadurch überhaupt die für die betroffenen Produkte relevante Nachfrage abgebildet wird, als auch ob die geschätzte Höhe des vermeintlichen Zusammenhangs ökonomisch plausibel ist.⁴⁰

Ein möglicher Einwand gegen ein solches Vorgehen könnte indes lauten, dass im Falle der Energiekosten ein ggf. zunächst ökonomisch unplausibel hoher Koeffizient möglicherweise andere Effekte mit aufnimmt, beispielsweise da mit den Energiekosten gleichzeitig andere, in der Regression vernachlässigte (Material-)Kosten mit gestiegen sind. Beispielsweise werden Modelle der Zeitreihenanalyse für kurzfristige Prognosen eingesetzt, wobei dann die Modellwahl oft allein auf Basis der kurzfristigen Prognosegüte erfolgt losgelöst von einer sinnhaften Interpretation aller aufgenommenen Variablen. Die Vorstellung, das vom Gutachter bei Kartellschadenschätzungen gewählte Modell wäre in der Regel in der Lage, den Preisbildungsprozess über einen ggf. langen Zeitraum mit hoher Präzision abzubilden, erscheint allerdings problematisch. Mit Ausnahme weniger Industrien dürften selbst die beteiligten Unternehmen oder aber sie beratende Spezialisten (die in der Regel nicht deckungsgleich mit den wettbewerbsökonomischen Beratungsfirmen sind) nicht über solche Modelle verfügen – es sei denn, die Preise sind beispielsweise eng an eine zentrale Kostengröße gebunden, oder aber sie unterliegen nur geringfügigen Schwankungen.⁴¹ Diese Aussage

⁴⁰ Diese Ausführungen sollen allerdings kein Plädoyer dafür sein, eine Regressionsanalyse allein deshalb zu verwerfen, weil nicht alle Koeffizienten ökonomisch sinnvoll sind. Beispielsweise können zwei Kostengrößen sich stark gleichlaufend verändern (d.h. hoch korreliert sein). Dann mag in den Daten ggf. nicht genug Variation in jeder der beiden Variablen vorhanden sein, um den jeweiligen Koeffizienten, d.h. den partiellen Effekt, *ceteris paribus* hinreichend präzise zu schätzen. Wichtig ist stets, ob eine Kontrollvariable mit ggf. einem ökonomisch nicht-sinnhaften Koeffizienten das interessierende Ergebnis, hier die Schätzung des Verstoßeffektes, wesentlich beeinflusst. Robustheitsanalysen, bei denen eine Teilmenge der Kontrollvariablen betrachtet werden, leisten hier einen wesentlichen Beitrag. Wir gehen nachfolgend auch auf die Rolle statistischer Gütemaße ein, die aber eine ökonomische Bewertung nicht ersetzen sollten.

⁴¹ Problematisch können hierbei im Rahmen von (Panel-)Analysen über verschiedene Produkte vorgetragene Behauptungen sein, das Modell könne beispielsweise 90 % der Preisvariationen „erklären“. Mitunter ist dies allein darauf zurückzuführen, dass sogenannte „Produkt-Fixeffekte“ in die Regressionsgleichung aufgenommen wurden, die lediglich den Niveauunterschied zwischen höher- und niedrigpreisigen Produkten erfassen. Was im zeitlichen Verlauf allerdings interessiert, ist eben die Variation über die Zeit, die diese Produkt-Fixeffekte dann mitunter nicht „erklären“.

beeinträchtigt allerdings nicht die grundsätzliche Zweckmäßigkeit der Schätzung, deren alleiniges Ziel die Ermittlung des Verstoßeffektes ist. Daher sollte zuvörderst darauf geachtet werden, dass die Wahl der Methode, hier der Regressionsgleichung und damit der Aufnahme verschiedener Kontrollvariablen, nicht von vorneherein Gefahr läuft, die Schätzung des interessierenden Verstoßeffektes zu verzerren. Dies verlangt ein ökonomisches statt einem rein statistischen Herangehen an die Modellselektion sowie an die Prüfung der geschätzten Koeffizienten.

IV.iii Einbezug des konkreten Verstoßes

Wir kommen noch einmal auf das Beispiel der Kosten als Kontrollvariable zurück. Dies ist nicht an einen konkreten Fall gebunden. Die Problematik von durch den Verstoß selbst verzerrten Kosten stellt stattdessen ein oft gewähltes Beispiel dar. Bei der Überprüfung der Verlässlichkeit der Datengrundlage haben wir u.a. ausgeführt, dass selbst eine fehlerfreie Übernahme von unternehmensinternen Kosten kein Garant dafür ist, dass damit die Kontrollvariable „Kosten“ eine verlässliche Datengrundlage erhalten hat (u.a. da die Kosten bzw. die interne Kostengröße selbst vom Verstoß beeinflusst sein können). Und im unmittelbar vorausgehenden Abschnitt haben wir am Beispiel von (Energie-)Kosten dargestellt, wie gerade bei kürzeren Zeitreihen (im zeitlichen Vergleich) der entsprechende Koeffizient auch fälschlicherweise, d.h. allein aufgrund einer zeitgleichen Veränderung, Preisänderungen „wegerklärt“ – aber auch genauso zu einem falschen Aufschlag führen kann. In diesem Falle ist, wie wir ausgeführt haben, eine Kenntnis des Sachverhalts wesentlich, da damit begründet werden kann, welcher Zusammenhang zwischen den betrachteten Kosten und den Preisen überhaupt ökonomisch plausibel ist. Nachfolgend greifen wir wieder auf das Kostenbeispiel zurück, um zu begründen, warum insbesondere auch eine Kenntnis des jeweiligen Verstoßes wesentlich ist für eine korrekte Spezifikation des Modells (der Regressionsgleichung) und damit insgesamt für die Gewährleistung einer korrekten Methodenwahl. Wie bereits dargestellt, sprechen wir uns damit gegen die Vorstellung aus, (Regressions-)Analysen würden stets bereits einen Erkenntnisgewinn darstellen, sofern sie nur „statistisch korrekt“ durchgeführt wurden (und auf einer verlässlichen Datengrundlage).

Hierzu nehmen wir an, ein Kartell hätte (unter anderem) explizit die (höhere) Weitergabe bestimmter Materialkosten in die Preise bezweckt. Für solche Absprachen gibt es genug Beispiele, gerade dann, wenn es sich dabei um einen wesentlichen Kostenfaktor handelt. Wir nehmen der Einfachheit halber auch an, dass ein Lesen des Fallberichts bzw. des Bußgeldbescheids auch darüber Auskunft gibt, dass solche Preiserhöhungen nach einem Anstieg der Materialkosten in der Tat besprochen und umgesetzt wurden. Weiter nehmen wir an, dass ein Gutachter dies entweder nicht zur Kenntnis genommen hat oder aber schlicht ignoriert. Der Gutachter spezifiziert nun seine Regressionsgleichung weiterhin derart, wie wir es für den Standardfall angenommen haben: Er nimmt Kosten in die Regressionsgleichung so auf, dass er für den Verstoß- und den Vergleichszeitraum einen gleichbleibenden Zusammenhang zwischen Kosten und Preisen unterstellt. Dies bedeutet formal nichts anderes, als dass er nur einen einzigen Koeffizienten für die Kontrollvariable „Kosten“ mit aufnimmt und schätzt. Dies kann aber dann unzulänglich sein, wenn gerade

die Weitergabe bestimmter Materialkosten Gegenstand der Absprachen waren. Denn dann würde der geschätzte Zusammenhang zwischen Kosten und Preisen den Verstoßeffekt fälschlicherweise (teilweise) absorbieren. Anders ausgedrückt sollte stattdessen der Zusammenhang zumindest zwischen diesen Materialkosten und den Preisen im Verstoßzeitraum höher sein als im Vergleichszeitraum. Eine mögliche Vorgehensweise ist es, ein sogenanntes interagiertes Modell zu schätzen, wobei es sich hierbei intuitiv schlicht darum handelt, dass für die beiden Zeiträume ein unterschiedlicher Zusammenhang (Koeffizient) möglich ist.⁴² Damit müssen allerdings auch zwei Koeffizienten geschätzt werden, was gerade bei kurzen Vergleichszeiträumen oder aber wenn dort die entsprechenden Materialkostenänderungen weniger ausgeprägt waren ggf. schwierig ist. Der Unterschied zwischen den Koeffizienten ist dann ebenfalls dem Verstoß anzulasten und kann bei der Berechnung des Aufschlags mit berücksichtigt werden.

Wir haben dieses Beispiel lediglich als Illustration dafür gewählt, dass das Aufsetzen einer Regressionsanalyse und speziell der Regressionsgleichung ohne eine Kenntnis zur Wirkungsweise des Verstoßes zu Fehlern führen kann. Im vorliegenden Fall wäre es auch nicht möglich, bei einer Beschränkung auf Modelle, die einen einheitlichen Zusammenhang zwischen Kosten und Preisen über den gesamten Zeitraum zugrunde legen, den Fehler allein über (statistische) Gütemaße zu erkennen. Wenn der Vergleichszeitraum kurz ist bzw. wenig Variation in den für die Absprachen relevanten Materialkosten aufweist, dann mag es außerdem sogar so sein, dass auch ein rein statistischer Vergleich zwischen dem richtigen (interagierten) Modell und dem falschen, wesentliche Punkte des Verstoßes ignorierenden Modells sich nicht eindeutig für ersteres ausspricht.⁴³ Es sollte aber deutlich geworden sein, dass in diesem Beispiel eine Korrektur geboten wäre, sofern es zu einer solch unterschiedlichen Weitergabe von Kosten gekommen ist. Die Gebotenheit einer solchen Korrektur erschließt sich dagegen erst dann, wenn die Annahmen des Modells abgeglichen werden mit der Wirkungsweise des konkret vorliegenden Verstoßes.

IV.iv Zur Rolle statistischer Gütemaße

Wie bereits zuvor betont, wollen wir nicht die Rolle statistischer Gütemaße für die Modellselektion bzw. damit einhergehend für die Bewertung der Korrektheit der Methode in Frage stellen. Solche Maße unterstützen unter anderem die Auswahl der Kontrollvariablen oder aber die Spezifikation des funktionalen Zusammenhangs zwischen einer bestimmten Kontrollvariablen und der abhängigen Variablen (wobei wir in unserem Standard-Modell schlicht einen einfachen, linearen Zusammenhang unterstellt haben, der dann durch einen einzigen Koeffizienten beschrieben ist).

⁴² *Inderst/Milde*, Concurrences 2-2019.

⁴³ Beispielsweise könnte ein Gutachter testen, ob die beiden Koeffizienten verschieden sind. Liegen aber beispielsweise wenige Beobachtungen im Vergleichszeitraum vor, so mag selbst ein deutlicher Unterschied nicht statistisch signifikant sein. Der Gutachter mag dann feststellen, dass die Hypothese eines einheitlichen Koeffizienten nicht „verworfen“ wurde, und am ursprünglichen Modell festhalten. Ein solches Vorgehen würde aber wesentliche Aspekte des Sachverhalts ignorieren.

Wir wollen hier abschließend kurz an einem Beispiel die Rolle, aber auch die Beschränkungen, dieser Maße darstellen (sofern dies nicht bereits hinreichend erfolgt ist). Hierfür kehren wir zu unserem Salatbeispiel zurück und nehmen an, wir hätten nun für den Verstoß- und Vergleichszeitraum jeweils Beobachtungen für den Niederschlag und den Preis gesammelt, aber keine weiteren Informationen. Wir ignorieren nun zunächst den Verstoß. Somit haben wir lediglich Paare aus Niederschlag und Preis. Unsere Aufgabe soll es nun sein, aus diesen Paaren (oder aber, graphisch dargestellt, aus der entsprechenden Punktwolke) einen Zusammenhang abzuleiten. Bisher haben wir uns hierbei auf einen rein linearen Zusammenhang beschränkt. Stellt man sich diese Punktwolke nun so vor, dass auf der horizontalen Achse der Niederschlag abgetragen ist und auf der vertikalen Achse der Preis, und verbindet man diese Punkte, so erhält man einen ggf. extrem volatil verlaufenden Graphen, welcher aber, wenn unsere ökonomische Theorie richtig ist, insgesamt eine fallende Beziehung darstellt: Mehr Niederschlag soll zu einer höheren Menge und in der Folge zu einem geringeren Preis führen. Eine Linie mag aber diesen Zusammenhang und insbesondere den durch andere Einflussfaktoren verursachten volatilen, nicht geradlinigen Verlauf schwerlich erfassen. Man könnte nun stattdessen annehmen, dass es eine linear-quadratische Beziehung zwischen Niederschlag und Preis gibt, was uns einen zusätzlichen Freiheitsgrad einräumt. Dies mag in der Tat sinnvoll sein, da ggf. ab einer bestimmten Niederschlagsmenge der Salat nicht mehr schneller wächst. Wir könnten aber den Niederschlag nicht nur quadratisch eingehen lassen, sondern auch mit höherer Potenz. In mathematischer Hinsicht stellen dies alles sogenannte Polynome dar⁴⁴. Je mehr Terme wir nun zulassen, umso genauer kann sich unsere gewählte (polynomische) Funktion dem tatsächlichen Zusammenhang zwischen Niederschlag und Preis annähern.

Erlauben wir gar eine sehr hohe Anzahl an Termen, so „erklärt“ unsere hochgradig komplexe Funktion den Zusammenhang zwischen Niederschlag und Preis vollständig. In einer Regressionsanalyse könnten wir nun all diese Terme mit einführen und den Preis als Funktion des Niederschlags, des quadrierten Niederschlags etc. darstellen. Diese Regressionsgleichung ergäbe dann einen sehr guten (bzw. gar den besten) „fit“ bzw. sie kann jegliche Variation in den Preisen vermeintlich „erklären“. Allerdings passen wir dabei die Regressionsgleichung lediglich an die Idiosynkrasien der vorhandenen Beobachtungen an. Auskunft über den wirklichen Erklärungsgehalt der verschiedenen Terme erhalten wird dabei nicht. Die statistischen Gütemaße helfen nun u.a. dabei, dass man kein solches „overfitting“ betreibt.⁴⁵

Es soll rein illustrativ zusätzlich angenommen werden, dass es, da wir nur einen kurzen Zeitraum betrachten, innerhalb von Verstoß- und Vergleichszeitraum geringe Variationen im Niederschlag gab, allerdings einen ausgeprägt höheren Niederschlag im Vergleichszeitraum insgesamt. Beispielsweise könnten sich die Beobachtungen auf lediglich zwei Jahre beschränken, wobei das zweite (Vergleichs-)Jahr ein sehr

⁴⁴ Ein Polynom bezeichnet die Summierung der Vielfachen von Potenzen einer Variablen.

⁴⁵ Maße wie das oft verwendete AIC-Kriterium (ausgeschrieben: Akaike Information Criterion) haben dabei eine bestimmte Zielfunktion vor Augen, die einerseits einen besseren „fit“ belohnt und andererseits die Aufnahme zusätzlicher Kontrollvariablen „bestraft“.

verregnetes Jahr war. Unterstellt man einen einfachen linearen Zusammenhang zwischen Niederschlag und Preis, so wird dieser sowohl auf die geringen Niederschlagsvariationen in den beiden Zeiträumen als auch auf den Vergleich zwischen den beiden Jahren „angewandt“. Dadurch mag ein erheblicher Teil des Preisabfalls (d.h. der unterschiedlichen Mittelwerte) „wegerklärt“ werden. Lässt man stattdessen eine andere funktionale Beziehung zu, durch die beispielsweise der Mehrwert eines höheren Niederschlags ab einer gewissen Menge beschränkt ist, so mag ein deutlich höherer Anteil des Preisabfalls dem Ende des Verstoßes zugeschrieben werden. Die statistische Modellselektion und die ökonomische Begründung und Auswahl der Kontrollvariablen und ihrer Beziehung zur abhängigen Preisvariablen können sich hier sinnvoll ergänzen. So können Gespräche mit Industrieexperten oder aber andere Studien Aufschluss dazu geben, welcher Zusammenhang zwischen Niederschlag und beispielsweise der geernteten Menge insgesamt ökonomisch plausibel ist. Es sollte aber in dem uns interessierenden Zusammenhang eben nicht die Wahl ausschließlich einem statistischen Gütemaß überlassen werden.

V Signifikanz

V.i Das statistische Konzept der Signifikanz

Das Konzept der Signifikanz hat eine eindeutige statistische Definition. Das Einfordern signifikanter Ergebnisse gehört auch zum Standardrepertoire der empirischen Forschung. Damit sollte das Kriterium „signifikanter Ergebnisse“ eigentlich hinreichend klar sein. Im vorliegenden Zusammenhang ist allerdings zu berücksichtigen, dass die Zielsetzung und die Möglichkeiten der Schadensschätzung in follow-on Klagen nicht mit denen eines wissenschaftlichen Hypothesentests in einer kontrollierten (Labor-)Umgebung gleichgesetzt werden können.

Wir gehen nachfolgend konkret wie folgt vor, um die folgenden Aussagen herzuleiten: Wir stellen zunächst abstrakt das Vorgehen bei einem (ökonomischen) Hypothesentest dar. Die Definition der Signifikanz eines Ergebnisses bzw. ein gefordertes Signifikanzniveau leiten sich unmittelbar aus einem solchen Hypothesentest ab. Unsere wesentliche Aussage dabei ist, dass dies einseitig nur auf sogenannte Fehler 1. Art abstellt, d.h. im vorliegenden Fall der Fehler einer Annahme eines Verstoßeffektes trotz tatsächlicher Unwirksamkeit. Der Fehler 2. Art, dass beispielsweise aufgrund einer ungenügenden Datenbasis kein statistisch signifikanter Effekt geschätzt werden kann, obwohl der Verstoß tatsächlich wirksam war, wird dabei nicht hinterfragt. Eine solche Betrachtung der verschiedenen tolerierbaren Fehler ist jedoch ein Umstand, der bei der Würdigung der Ergebnisse einer Regressionsanalyse Bedeutung haben kann, wenn es darum geht, die Wahrscheinlichkeit eines Schadenseintritts oder einer bestimmten Schadenshöhe zu bewerten. Daher ist zweifelhaft, ob ein Gericht diese Frage dem Gutachter allein überlassen kann – und damit auch die Beantwortung der Frage, ab welchem Signifikanzniveau ein Ergebnis hinreichend belastbar ist. Vielmehr sprechen gute Gründe dafür, dass die Belastbarkeit eines Ergebnisses im Rahmen der Schadensschätzung zuvörderst eine juristische Kategorie ist. Eine Regressionsanalyse kann dazu beitragen, diese Belastbarkeit in der Überzeugungsbildung des Gerichts herzustellen, was aber bedingt, dass das

Gericht seine eigenen Maßstäbe für die Überzeugungsbildung anlegt und die Ergebnisse der Regressionsanalyse daran misst und nicht umgekehrt.

Die Umsetzung eines solchen Hypothesentests mittels einer Regressionsanalyse muss zudem andere Informationen vernachlässigen. So kann das Ergebnis der Regressionsanalyse statistisch insignifikant sein, obwohl sich sogar aus dem Beweisbeschluss für bestimmte Zeitpunkte einzelne konkrete Absprachen ableiten lassen, die sich dann auch so in den Daten finden. Die statistische Signifikanz allein ist blind gegenüber solcher anderer Evidenz. Andererseits sagt auch ein signifikantes Schätzergebnis eines Aufschlages von beispielsweise 30 % in keiner Weise aus, dass der tatsächliche Verstoßeffekt ebenfalls 30 % (mit einer gewissen Wahrscheinlichkeit) beträgt oder aber mit einer bestimmten Wahrscheinlichkeit in ein gewisses Intervall fällt. In der Tat kann sogar ein signifikanter Aufschlag, den der Kläger geschätzt hat, hinreichend vereinbar sein mit dem Ergebnis der Beklagten, dass kein signifikanter Aufschlag vorliegt. Dies liegt mit an der inhärenten Schätzunsicherheit. In anderen Fällen können diese Unterschiede aber nahelegen, dass zumindest eines der beiden Modelle „nicht korrekt“ ist, was in der Regel daran liegen dürfte, dass die verschiedenen Gutachter trotz ggf. der gleichen Anwendung eines zeitlichen Vergleichs mittels einer Indikatorvariablen-Methode unterschiedliche Kontrollvariablen verwenden.

Das Zusammenbringen von Fehlern 1. Art und 2. Art, von Regressionsergebnis mit anderer Evidenz des Falles sowie von unterschiedlichen, (scheinbar) widersprüchlichen (Regressions-)Ergebnissen verschiedener Gutachter ist unseres Erachtens nötig und es kann, so unser einheitlicher Tenor, nicht über Checklisten oder allein über anzuwendende Gütemaße erfolgen, sondern nur in einer ökonomischen fundierten Gesamtschau. Hierzu bieten sich auch andere (aber verwandte) statistische Instrumente als das der Signifikanz (ergänzend) an, worauf wir abschließend eingehen. Das alleinige Abstellen auf beispielsweise das Regressionsergebnis eines gerichtlichen Sachverständigen kann dem insbesondere dann nicht gerecht werden, wenn der Sachverständige auf die andere vorgelegte Evidenz und insbesondere auf mögliche Unterschiede zwischen den Regressionsanalysen nicht in erforderlichen Maßen eingeht. Die Vorstellung, es gäbe den einen richtigen Test (inkl. entsprechendem Datensatz), führt im vorliegenden Fall zu Fehlschlüssen.

V.ii Idealvorstellung eines Testverfahrens und Notwendigkeit der gemeinsamen Berücksichtigung von Fehlern 1. und 2. Art

Lediglich für ein Gedankenexperiment wollen wir nun annehmen, dass wir nicht etwa einen tatsächlichen Verstoß *ex post* untersuchen, sondern uns aus Forschungsperspektive die Frage stellen, ob ein bestimmter Typus eines Verstoßes die Preise erhöht. Hierfür entwickeln wir nun ein Testdesign. Dabei würden wir Unternehmen in einem bestimmten Markt die entsprechende Absprache erlauben (was, in der gängigen Terminologie, unsere sogenannte „treatment“-Gruppe wäre), während dies für Unternehmen in einem anderen Markt nicht der Fall wäre (was unsere Kontrollgruppe darstellt). Wir würden dabei auch sicherstellen wollen, dass sich die beiden Märkte möglichst vollends gleichen. Solche Umgebungen können in der experimentellen Wirtschaftsforschung quasi im Labor hergestellt werden – mit dem wesentlichen

Abstrich, dass in diesen Experimenten in der Regel Studierende hoch stilisierte Entscheidungen mit für sie geringen finanziellen Implikationen treffen. Wenn solche Experimente durchgeführt werden, sollte ebenfalls vorab bestimmt werden, wie viele Wiederholungen durchgeführt werden sowie mit wie vielen verschiedenen Teilnehmern. In anderen Worten: Die Struktur der letztendlich zur Verfügung stehenden Daten ist wesentlicher Teil des Testdesigns. Damit kann *a priori* bestimmt werden, wie aussagefähig der Test sein wird.

Um dies genauer darzustellen, müssen wir den Test etwas formaler beschreiben. Hierbei geht es im Wesentlichen um das Formulieren der sogenannten Ausgangs- oder Nullhypothese. Wir wenden uns nun kurz vom Kartellexperiment ab und bilden das Beispiel, dass mit dem aufzusetzenden Experiment die Wirksamkeit eines Medikaments getestet werden soll. Hier würde man als Nullhypothese formulieren, dass das Medikament gerade keine Wirkung hat. Mit dem Experiment wollen wir dann feststellen, ob wir diese Nullhypothese „keine Wirkung“ mit hinreichender Sicherheit ablehnen können, so dass das Medikament seine Zulassung erhält. Es geht dann vor allem darum, sogenannte Fehler der 1. Art zu vermeiden, d.h. die fehlerhafte Ablehnung der Nullhypothese (und damit die Zulassung eines unwirksamen Medikaments). Wir können diesen Fehler 1. Art auch beziffern mittels des Signifikanzniveaus des Tests. In der Wirtschaftswissenschaft wird dabei häufig ein Signifikanzniveau von 5 % verwendet. Dies bedeutet, dass wir gewillt sind, einen Fehler 1. Art in 5 % aller Fälle zu akzeptieren. Oder anders gesagt: Wenn wir viele Medikamente unabhängig voneinander testen, die aber allesamt tatsächlich unwirksam wären (was wir aber nicht wissen), so würden wir *a priori* im Durchschnitt eines von zwanzig Medikamenten fehlerhaft als wirksam klassifizieren.

Wir kommen nun zu unserem Kartellexperiment zurück, in dem wir (mittels studentischer Teilnehmer in einem Experimentallabor) den Effekt einer bestimmten Absprache auf den Preis testen. Stellen wir nun im Durchschnitt über alle Versuche einen höheren Preis bei einer solchen Absprache fest, so können wir mittels einfacher statistischer Verfahren auch die Signifikanz dieses Ergebnisses bestimmen. Auch hier gilt: Wenn wir feststellen, dass der Effekt auf dem Niveau von 5 % signifikant ist, so können wir aufgrund der vorhandenen Informationen davon ausgehen, nur mit 5 % Wahrscheinlichkeit einen Fehler zu machen, wenn wir die Nullhypothese „kein Effekt“ verwerfen. Genau dies besagt das Signifikanzniveau. Es sagt beispielsweise nichts darüber aus, mit welcher Wahrscheinlichkeit der tatsächliche Effekt einen bestimmten Wert annimmt.

Die Signifikanz eines ermittelten Effektes ist damit untrennbar mit dem Testen einer Hypothese verbunden. Und diese Hypothese ist im vorliegenden Fall die Nullhypothese „kein Effekt“. Wie wir dargestellt haben, folgt ein solcher Testaufbau der Logik, dass wir die Fehler 1. Art beziffern wollen, d.h. den Fehler, fälschlicherweise davon auszugehen, dass es einen positiven Verstoßeffekt gab. Das Signifikanzniveau sagt allerdings nichts über den Fehler 2. Art aus, d.h. den Fehler, die Nullhypothese nicht abzulehnen, obwohl tatsächlich ein Verstoßeffekt vorliegt. Fordert man bei einem Testdesign ein hohes Signifikanzniveau, um den Fehler 1. Art zu verringern, so erhöht man zwangsläufig *a priori* den Fehler 2. Art. Gerade bei einer

geringen Anzahl von Beobachtungen oder aber wenn aufgrund einer unzureichenden Modellspezifikation die Schätzung mit erheblicher Unsicherheit behaftet wird, kann der Fehler 2. Art erheblich sein und insbesondere ein Vielfaches des geforderten Signifikanzniveaus. Wenn aber der Verstoß bereits festgestellt wurde, so wäre es fraglich, warum für eine Schadensschätzung allein auf den Fehler 1. Art abgestellt wird. Dies wäre aber die zwangsläufige Konsequenz, wenn allein darauf geachtet würde, ob ein vorgelegtes Regressionsergebnis signifikant oder insignifikant ist.

Mit Sicherheit kann dies kein Plädoyer dafür sein, die statistische Signifikanz eines Ergebnisses bzw. das Fehlen einer solchen zu vernachlässigen. Ergebnisse mit geringer Signifikanz können aber sowohl solche sein, bei denen der Verstoßeffekt sehr niedrig geschätzt wurde, oder solche, bei denen die erhebliche Schätzunsicherheit (so im Falle weniger Beobachtungen) den Ausschlag gegeben hat trotz einem hohen geschätzten Verstoßeffekt. Auch ignorieren wir nicht, dass das Gericht seine Schätzung stets auf eine hinreichend belastbare Grundlage stellen muss. Wie bereits dargestellt, kann es beispielsweise für die Wirksamkeit des Verstoßes aber auch Anhaltspunkte geben, die durch die vorgelegte Schätzung nicht erfasst sind. Ferner mag es sein, dass ein anderer Ansatz, der beispielsweise enger an den Verstoß ansetzt (und dabei beispielsweise nur bestimmte Zeitintervalle betrachtet, die etwa auch im Bußgeldbescheid markiert sind), zu Schätzungen mit geringerem Fehler 1. Art und damit einer höheren Signifikanz führen. Wir halten es für wichtig, nachfolgend noch einmal auf die beschränkte Aussagefähigkeit der Signifikanz einzugehen sowie auch auf die Fehleranfälligkeit, wenn allein dieses Kriterium angewandt wird – und dies insbesondere auch mit einem allein vom Gutachter, nicht allerdings vom Gericht, festgesetztem notwendigem Signifikanzniveau.

V.iii Die Beschränktheit der Aussagekraft signifikanter und insignifikanter Ergebnisse

Folgendes wollen wir zunächst noch einmal herausstellen: Die Signifikanz eines Ergebnisses hat keinen Wert, wenn das Modell nicht adäquat ist, d.h. wenn es zu verzerrten Ergebnissen führt. Dies betrifft insbesondere die Auswahl der verwendeten Kontrollvariablen sowie, wie dargestellt, die in der Regressionsgleichung formulierten Beziehungen zwischen diesen und der abhängigen Variablen. Nachfolgend setzen wir dies aber voraus und stellen allein auf die (verbleibende) Schätzunsicherheit ab.

Wir betrachten zunächst die Möglichkeit, dass ein gerichtlich bestellter Gutachter lediglich einen statistisch insignifikanten Verstoßeffekt findet. Ein solches insignifikantes Ergebnis darf nach dem Vorgesagten nicht aus sich heraus bereits als hinreichende Evidenz gegen einen Verstoßeffekt gewertet werden. Die Aussage eines insignifikanten Effektes bedeutet nämlich lediglich, dass man die Nullhypothese „kein Effekt“ nicht mit einer vom Gutachter aufgrund von Gepflogenheiten in der Wissenschaft als notwendig erachteten Sicherheit verwerfen kann. Wenn aber der Test sehr ungenau ist, da wenig Beobachtungen zur Verfügung stehen oder aber wenig Kontrollvariablen, so kann gleichsam ein geschätzter Effekt von 2 % wie auch ein deutlich höherer Effekt noch insignifikant sein. Es ist daher zumindest darauf zu achten, aus welchen Gründen das Ergebnis insignifikant ist. Und gleichsam ist daran zu erinnern, dass die Feststellung der Signifikanz zwingend von dem vom Gutachter angewandten Signifikanzniveau abhängt. Ein Gutachter

wird in der Regel das bei Ökonomen übliche Signifikanzniveau von 5 % anwenden – und dies ohne jegliche Berücksichtigung der damit einhergehenden Fehler 2. Art oder aber anderer Evidenz, so etwa zur Natur des Verstoßes. Auf Letzteres gehen wir nachfolgend noch separat ein.

Nehmen wir nun alternativ an, der Gutachter habe ein signifikantes Ergebnis festgestellt. Wie bereits dargestellt, darf aber auch ein signifikantes Ergebnis nicht mit Aussagen zur Wahrscheinlichkeit eines bestimmten Verstoßeffektes und insbesondere des geschätzten Effekts verwechselt werden. Ein geschätzter Verstoßeffekt von 20 % ist keine Evidenz dafür, dass der tatsächliche Verstoßeffekt genau 20 % ist. Selbst wenn der Gutachter einen Aufschlag von 20 % als „bestmögliche Schätzung“ betrachtet, ist dieses Caveat dennoch wichtig. Denn die Feststellung der „besten Schätzung“ bezieht sich nur auf die dafür verwendeten Daten und das dafür verwendete Modell. Ein anderer Gutachter mag mit einem anderen Datensatz oder einer anderen Regressionsgleichung einen geringeren Aufschlag geschätzt haben. Die beiden Ergebnisse sind allerdings aufgrund der inhärenten Schätzunsicherheit nicht notwendigerweise inkompatibel zueinander. Anders ausgedrückt: Die Signifikanz eines Effektes ist keine Aussage dazu, dass dieser Wert dem wahren Wert entspricht oder aber dass mit Sicherheit andere Werte ausgeschlossen sind. Die Signifikanz eines Ergebnisses rechtfertigt insbesondere nicht, andere Evidenz zu ignorieren.

V.iv Das (quasi zwangsläufige) Ignorieren anderer Informationen und Evidenz in der Regressionsanalyse

Wie bereits angeführt, mag auch ein positiver Verstoßeffekt vornehmlich deshalb statistisch insignifikant sein, weil der mittels der Regressionsanalyse durchgeführte Test mit einer relativ großen Unsicherheit behaftet ist (nach Maßgabe des angewendeten Signifikanzniveaus, d.h. des vom Gutachter als tolerierbar erachteten Fehler 1. Art). Wir halten es bei einer gerichtlichen Schadensschätzung daher nicht für ausreichend, wenn allein aufgrund eines solchen Ergebnisses bereits, ungeachtet der anderen vorliegenden Evidenz, geschlussfolgert würde, dass kein Verstoßeffekt vorläge. Denn eine solche Einschätzung würde mitunter die potentiell relevanten anderen Indizien für eine Schadensschätzung noch nicht ausschöpfen. Zum einen sei daran erinnert, dass die festgestellte mangelnde Signifikanz die Wahl eines bestimmten Signifikanzniveaus voraussetzt. In der Ökonomie ist es für das Testen wissenschaftlicher Hypothesen in der Tat üblich, dass das geforderte Signifikanzniveau auf 5 % gesetzt wird. Man ist demnach gewillt, einen Fehler 1. Art höchstens in einem von 20 Fällen zu tolerieren (bei 5 %). Wie bereits ausgeführt wurde, sagt dies aber nichts über den Fehler 2. Art aus, d.h. trotz eines Verstoßeffektes bei der Nullhypothese „kein Effekt“ zu bleiben. Dies gilt, erneut, umso mehr, wenn dem Gericht auch andere Informationen, etwa zur Natur des Verstoßes vorliegen, z.B. aus einem Bußgeldbescheid. Das kann etwa die Frage betreffen, ob es sich um ein „Hardcore“-Kartell handelte oder aber um einen losen und sehr sporadischen Informationsaustausch. Auch mögen direkte Beweismittel, etwa durch Zeugen, dazu vorliegen, dass es zu konkreten Absprachen und deren Umsetzung kam. Gerade dann, wenn das Aufsetzen der Regressionsgleichung nicht die Natur des Verstoßes, so beispielsweise die spezifischen Absprachen, berücksichtigt, blieben diese zusätzlichen Informationen unberücksichtigt, sofern man sie nun nicht direkt

in die Schadensschätzung einfließen lässt. Es ist zweifelhaft, ob sich diese Zusatzinformationen ohne Weiteres durch einen „Zu- oder Abschlag auf die Signifikanz“ in eine statistisch fundierte Gesamtbewertung integrieren lassen. Diese Informationen dürfen allerdings trotzdem nicht ignoriert werden.

Diese Bemerkungen gelten allerdings nicht nur in eine Richtung, d.h. für den Fall, dass der gerichtlich bestellte Gutachter andernfalls lediglich ein insignifikantes Ergebnis feststellt. Wie bereits dargestellt, darf das Signifikanzniveau bei einem signifikanten Verstoßeffekt nicht verwechselt werden mit einer Aussage zur Wahrscheinlichkeit eines solchen Aufschlags oder aber eines Aufschlags in einem bestimmten Intervall.⁴⁶ Sei nun beispielsweise der geschätzte Aufschlag 20 %. In einzelnen Fällen mag dies durchaus plausibel sein; in anderen Fällen allerdings gerade nicht, so angesichts beispielsweise der sich daraus angeblich ergebenden kontrafaktischen Gewinnmargen. Wiederum mögen einem Gericht auch andere Informationen vorliegen, so dazu, dass es sich hier lediglich um einen unspezifischen, sporadischen Informationsaustausch gehandelt hat. Eine Partei mag auch Evidenz dazu angeführt haben, dass sich die Preise der am Verstoß beteiligten Unternehmen sehr unterschiedlich und damit alles andere als zumindest augenscheinlich koordiniert entwickelt haben. Auch hier gilt wieder, dass diese Informationen üblicherweise nicht in die vorgelegte Regressionsanalyse integriert werden. Der Gutachter sollte allerdings gerade wenn er vom Gericht bestellt ist diese Informationen nicht ignorieren, sondern sie zum Anlass nehmen, um die Ergebnisse und damit auch den gewählten Ansatz zu hinterfragen Und dies gilt erst recht für das Gericht.

Die Beispiele sowohl der Interpretation eines signifikanten wie insignifikanten Ergebnisses haben deutlich gemacht, dass eine Regressionsanalyse nicht nur durch die verwendeten Daten und die Methode, insbesondere die aufgesetzte Regressionsgleichung, zwingend beschränkt ist, sondern dass es in der Regel nicht möglich ist, umfassend die gesamte vorliegende Evidenz mit aufzunehmen. Die Regressionsanalyse stellt lediglich die Umsetzung eines bestimmten Testverfahrens dar, nicht aber eine Gesamtbewertung aller Evidenz.

Es ließe sich möglicherweise einwenden, dass selbst wenn sich aus dem Sachverhalt sogar eindeutige Evidenz für die Umsetzung einer bestimmten Absprache ergibt, es letztlich auf die gesamte Quantifizierung ankäme, was eben eine Regressionsanalyse leiste. So könnte es sich ggf. um die einzige erfolgreiche Absprache gehandelt haben, deren Auswirkung insgesamt dann aber vernachlässigbar gering war. Ein insignifikantes Ergebnis „im Durchschnitt“ kann allerdings schlicht nicht beispielsweise konkreten Belegen der tatsächlichen Umsetzung einer Absprache widersprechen. Das insignifikante Ergebnis würde daher in diesem Fall gerade nicht hinreichende Evidenz zur Frage des „Ob“ eines Schadens darstellen. Zum anderen

⁴⁶ Zwar gehen wir darauf nicht weiter ein, allerdings dürfen auch die manchmal angegebenen sogenannten Konfidenzintervalle nicht in diesem Sinne interpretiert werden. Siehe dazu genauer, auch anhand von Simulationsbeispielen, *Frank/Inderst/Oldehaver*, ZWeR 2019, 39 ff.

übersähe dieses Argument sowohl die generelle Schätzunsicherheit wie auch die erhebliche Gefahr einer falschen Spezifikation der Regressionsgleichung, d.h. im weiteren Sinne einer fehlenden Korrektheit der Methode.

V.v Der Umgang mit widersprüchlichen „signifikanten“ Ergebnissen

Üblicherweise werden dem Gericht verschiedene Gutachten vorliegen. Die Kläger werden wohl mit ihrer Regressionsanalyse, d.h. mit ihrem Datensatz und ihrer Regressionsgleichung, einen signifikanten Verstoßeffekt gefunden haben. Die Beklagten werden auf ihrem Datensatz eine Regressionsgleichung geschätzt haben, die zu keinem signifikanten Verstoßeffekt führt. Beide Parteien mögen dabei einen zeitlichen Vergleich über eine Indikatorvariablen-Methode angewendet haben. Das Gericht mag nun versucht sein, sich *a priori* nicht mit diesen Parteigutachten auseinander zu setzen und einen gerichtlichen Sachverständigen benennen. Wenn die beiden Parteigutachter im Grunde dieselbe Methodik (hier eines zeitlichen Vergleiches mittels einer Indikatorvariablen-Methode) angewendet haben, dann wird in der Regel auch der gerichtlich bestellte Sachverständige das so halten. Denn die Wahl der Methode mag schlicht den Umständen des Falles geschuldet sein, so den fehlenden Informationen zu Vergleichsmärkten. Damit mögen drei Gutachten vorliegen, die auf derselben fundamentalen Methodik fußen, so beispielsweise einem zeitlichen Vergleich mittels einer Regressionsanalyse.

Es wäre nun allerdings problematisch, für das Urteil lediglich auf das Gutachten des bestellten Sachverständigen abzustellen, insbesondere dann wenn dieser relativ agnostisch eine eigene Schätzung durchführt, ohne zu erörtern, warum ggf. seine Resultate (deutlich) von denen eines oder beider Parteigutachten abweichen. Ein alleiniges Abstellen auf das Gutachten des bestellten Sachverständigen würde dann möglicherweise vorhandene Evidenz nicht hinreichend beachten.

Um festzustellen, ob die Unterschiedlichkeit der Ergebnisse an verschiedenen Daten liegt, bietet es sich unmittelbar an, die entsprechenden Regressionsanalysen mit den unterschiedlichen Datensätzen zu wiederholen. Das Gericht kann einen Sachverständigen instruieren, alternative Gutachten unter Zugrundelegung verschiedener Anknüpfungstatsachen zu erstellen⁴⁷ Hier kann eingewendet werden, dass dies zu aufwändig sei. Da die Analysen allerdings auf bereits von den jeweiligen Gutachtern aufbereiteten Daten erfolgen, was oftmals den wesentlichen Teil der Arbeit ausmacht, und da bei besonders umfangreichen Datensätzen ggf. sogar stichprobenhafte Auszüge zur Überprüfung im genannten Kontext ausreichen, wiegt dieses Argument weniger schwer. In der Regel werden allerdings die verschiedenen Gutachten auf anderen Regressionsgleichungen aufsetzen, d.h. insbesondere auch unterschiedliche Kontrollvariablen verwenden. Das Weglassen bzw. Hinzufügen dieser Variablen sollte allerdings bereits

⁴⁷ Zimmermann, in: Münchener Kommentar ZPO, 6. Auflage 2020, § 404a Rn. 6; Scheuch, BeckOK ZPO, BeckOK ZPO, Vorwerk/Wolf 42. Edition, Stand: 01.09.2021, § 404a Rn. 6. Zur gebotenen Abstimmung zwischen Gutachter und Gericht in Kartellschadensersatzverfahren siehe Burrichter, in: FS Canenbley 2012 (Hrsg. Studienvereinigung Kartellrecht e.V.), Ökonomische Gutachten in Kartellschadensersatzprozessen, S. 111, 113.

eine wesentliche Robustheitsanalyse darstellen, und das Gericht sollte sich mit diesen Einflüssen auf die Ergebnisse der Analysen daher auseinandersetzen.

Insgesamt sollte sich deshalb beim Vorliegen verschiedener Gutachten ein Urteil nicht ohne weitere Überlegungen auf ein einziges Ergebnis stützen, ungeachtet dessen, ob dieses statistisch signifikant ist oder kein signifikantes Ergebnis aufweist. Gerade da es, wie zuvor dargestellt, nicht die eine einzig korrekte Methode gibt, die sich zudem anhand (statistischer) Gütemaße abschließend ermitteln ließe, ist auch hier eine (ökonomische) Gesamtschau nötig. Zusätzlich wird im Schrifttum vorgeschlagen, dass bei der Bewertung statistischer Evidenz auch andere Kriterien und insbesondere auch unterstützende (graphische) Instrumente herangezogen werden. In mehreren aktuellen Beiträgen wurde das Konzept der „Schwere“ der Evidenz gerade auch für die Anwendungen in Schadensersatzklagen ausgearbeitet.⁴⁸

VI Schluss

Dieser Artikel setzte sich mit der Einordnung und Handhabarmachung von Regressionsanalysen im Lichte der durch den BGH aufgestellten Kriterien auseinander. Zu allen drei Kriterien – der verlässlichen Datengrundlage, der methodischen Korrektheit sowie der Signifikanz – haben wir an Beispielen zweierlei gezeigt. Zum einen können diese Kriterien auf Basis allgemeiner Einsichten zur Regressionsanalyse aber auch auf Basis praktischer Erfahrungen systematisch ausgefüllt werden. Zum anderen kann dies allerdings nicht zu einer abschließenden Checkliste führen, nach die Fragen eines Schadensersatzanspruchs ohne Hinterfragen der einzelnen Elemente der vorgelegten Regressionsanalyse „abgearbeitet“ werden können.

Die Bewertung der Datengrundlage sollte sich nicht in der Frage nach der Fehlerhaftigkeit erschöpfen. In der Regel muss der Gutachter erst das entsprechende Datengerüst erstellen und dabei wesentliche Annahmen zum Sachverhalt treffen, so wenn er beispielsweise bestimmte Produkte auswählt. Fehler in den Daten sind insbesondere dann problematisch, wenn sie nicht lediglich die Schätzgenauigkeit beeinflussen, da dieser Einfluss über statistische Maße erfasst werden kann, sondern wenn sie auch bei vielen Beobachtungen zu verzerrten Ergebnissen führen. Ob dies der Fall ist, hängt mit an der verwendeten Methode. Die Beurteilung der Korrektheit der Methode sollte wesentlich an der Regressionsgleichung ansetzen, die den unterstellten Zusammenhang zwischen verschiedenen Kontrollvariablen (sowie dem Verstoß) und der abhängigen Variablen, so dem Transaktionspreis, beschreibt. Eine solche Beurteilung sollte nicht allein durch statistische Gütemaße erfolgen, sondern nur in einer ökonomischen Gesamtschau, die den Sachverhalt mit einbezieht und damit wesentliche Charakteristika der Industrie sowie des Verstoßes.

⁴⁸ Für einen kurzen Überblick siehe *Bönisch/Inderst*, *Journal of European Competition Law & Practice* 2019, 499–504 sowie ausführlicher in *Bönisch/Inderst*, *ZWeR* 2020, 52. In einem weiteren Beitrag wird gezeigt, wie das Konzept genutzt werden kann, um scheinbar widersprüchliche Evidenz verschiedener Gutachten gemeinsam zu betrachten, *Bönisch/Inderst*, *Using the Statistical Concept of “Severity” to Assess the Compatibility of Seemingly Contradictory Statistical Evidence (With a Particular Application to Damage Estimation)*, *Journal of Competition Law & Economics* (im Erscheinen).

Weiterhin ist die Feststellung der Signifikanz eines Ergebnisses nur dann sinnvoll, wenn von der Korrektheit der Methode und damit insbesondere der zugrundeliegenden Regressionsgleichung ausgegangen werden kann. Allein die Signifikanz als Gütemaß in den Vordergrund zu stellen, birgt ein Potential für Fehleinschätzungen, wie wir dargestellt haben. So sollte ein insignifikantes Ergebnis nicht per se bereits als hinreichende Evidenz gegen einen Verstoßeffekt gewertet werden, und ein hoher signifikanter Verstoßeffekt schließt umgekehrt keineswegs geringere tatsächliche Verstoßeffekte aus. Auch aus diesem Grunde kann eine Gesamtbewertung der Evidenz sich nicht beispielsweise darin erschöpfen, allein das signifikante oder insignifikante Ergebnis eines gerichtlich bestellten Sachverständigen als Basis der gerichtlichen Schätzung zu verwenden, d.h. ohne den Einbezug der ggf. dazu widersprüchlichen Schätzungen der Parteigutachter oder aber vielfältiger anderer Evidenz, die sich in der Regel, wie wir dargestellt haben, gerade nicht in das Korsett einer Regressionsanalyse pressen lässt. In anderen Worten: Die Durchführung einer Regressionsanalyse ersetzt keine Gesamtbetrachtung aller verfügbaren Evidenz nach dem Maßstab der freien richterlichen Beweiswürdigung.

Wie wir zuvor dargestellt haben, können für eine solche Gesamtbetrachtung zusätzliche (intuitive) Konzepte wie das der „Schwere“ der mittels der verschiedenen Analysen vorgebrachten Evidenz verwendet werden. Das Konzept der Signifikanz ist letztlich zugeschnitten auf einen Hypothesentest und damit weniger geeignet insbesondere für den Umgang mit verschiedener, teils quantitativer und teils qualitativer sowie auch teils widersprüchlicher, Evidenz. Alternative, praxisorientierte Konzepte können sich erst dann entwickeln, wenn sich die Tatsacheninstanzen stärker mit den Ergebnissen von Regressionsanalysen auseinandersetzen und hierbei auch den Anspruch an die Gutachter stellen, das Aufsetzen der Analyse sowie die Darstellung der Ergebnisse verständlich zu begründen.

English abstract: The German Federal Supreme Court (Bundesgerichtshof) has, in its judgment of 13 April 2021 (case KZR 19/20), defined criteria by which to assess the results of a regression analysis in antitrust damages cases. The court refers to the reliability of the data basis, the validity of the method, and the significance of the results. In our article, we outline the general principles of regression analysis in an antitrust context. We venture to demonstrate that the evaluation of the results of a regression analysis in this context cannot follow a strict checklist-approach. This bears on how the criteria as defined by the court must be construed. We argue that a court should engage in a deeper analysis of how the regression analysis was performed, and to read the results in the light of the other information that a case provides, such as the specificities of the industry and the infringement. Our main posit is that a regression analysis cannot substitute for a comprehensive judicial assessment of the entirety of evidence submitted in an antitrust damages case.

