

Ergebniszusammenfassung
**Empirie zu personalisierten Preisen im
E-Commerce**



research
an der Universität
Regensburg



trinno

ibi research an der Universität Regensburg GmbH
Galgenbergstraße 25
93053 Regensburg

trinno GmbH
Franz-Mayer-Straße 1
93053 Regensburg

1 Einleitung

Preise für Waren und Dienstleistungen ändern sich immer schneller. Im E-Commerce sind diese Preisänderungen von besonderer Bedeutung. Denn einerseits herrscht online eine höhere Preistransparenz vor: Preise sind zumindest theoretisch für Kunden jederzeit einsehbar und vergleichbar. Andererseits ist es Händlern durch technische Maßnahmen möglich, ihre Preise kurzfristig zu verändern und beispielsweise auf Basis höherer Nachfrage oder der Tageszeit zu differenzieren – deutlich einfacher jedenfalls, als dies im stationären Bereich der Fall ist. Im Extremfall könnte es mit entsprechendem technischem Aufwand sogar möglich sein, die Preise kunden- und nutzungsspezifisch zu individualisieren.

Dabei ist grundsätzlich zwischen zwei Arten von Preisdifferenzierung zu unterscheiden: der dynamischen Preisänderung als Anpassung auf Basis objektiver, kundenunabhängiger Faktoren und der personalisierten Preisänderung als Anpassung auf Basis individuell kundenabhängiger Faktoren. Insbesondere die personalisierte Preisanpassung kann aus Verbrauchersicht problematisch sein. Es könnte zum Beispiel die Gefahr bestehen, wegen der Zugehörigkeit zu einer bestimmten Nutzungsgruppe diskriminiert zu werden. Kritisch zu sehen wären solche Anpassungen insbesondere dann, wenn die individuelle Preissetzung für den Kunden nicht erkennbar ist bzw. deren Gründe nicht transparent sind.

Es gab in Fachpresse und Öffentlichkeit immer wieder Vermutungen, dass personalisierte Preisdifferenzierung in manchen Marktbereichen bereits betrieben wird. Die in anderen Bereichen des Internets erkennbare immer stärkere Individualisierung – etwa bei der Online-Werbung oder der Anzeige von Suchmaschinenergebnissen – könnte sich also auch in der individuellen Preisgestaltung niederschlagen.

Meist beruhten diese Vermutungen auf anekdotischen Berichten oder allenfalls stichprobenartigen Untersuchungen, die einer wissenschaftlichen Überprüfung nicht standhalten. Eine umfangreiche empirische Studie war nötig, um den Sachverhalt belastbar zu belegen, war aber – soweit öffentlich bekannt – noch nicht durchgeführt worden. Zu diesem Zweck wurde vom BMJV dieses Forschungsvorhaben in Auftrag gegeben, dessen Ergebnisse nun vorliegen.

Im Folgenden werden zunächst Methodik und Vorgehen der Untersuchung skizziert, anschließend die wesentlichen Ergebnisse zusammengefasst.

2 Methodik und Projektvorgehen

Die interessierenden übergeordneten Forschungsfragen der Untersuchung waren:

- Wird im Markt personalisierte Preisgestaltung verwendet?
- Von welchen Kriterien hängt – für den Fall der Verwendung – die personalisierte Preisgestaltung ab und welches Ausmaß haben ihre Auswirkungen?

Hervorzuheben ist dabei die Differenzierung zwischen üblichen, auf Marktvorgängen oder auf objektiven Gegebenheiten basierenden (auch dynamischen) Preisänderungen einerseits und solchen, die auf subjektiven personenbezogenen Daten des Kunden beruhen, andererseits. Erstere sind aus Sicht des Verbraucherschutzes nicht zu beanstanden, während letztere für den Fall ihres Auftretens in dieser Hinsicht problematisch wären.

Das Design der Untersuchung orientierte sich an der formulierten Zielsetzung und dem geforderten Umfang. Es konkretisiert sich wie folgt:

Über einen Zeitraum von drei Monaten wurden die Preise erhoben, die auf unterschiedlichen Anbieter- und Preisvergleichsseiten für jeweils ein und das gleiche Produkt zum jeweils gleichen Zeitpunkt verlangt werden. Dabei wurden insbesondere die umsatzstärksten Online-Shops und -Plattformen im deutschen E-Commerce sowie die beliebtesten Preisvergleichsportale berücksichtigt. Zur Bestimmung der zu untersuchenden Shops bzw. Plattformen wurden öffentlich verfügbare Umsatzinformationen herangezogen.

Es wurde darauf geachtet, eine möglichst breite Palette an Produktkategorien einzubeziehen. Zur Begrenzung des Datenvolumens wurde nicht jede Kategorie in jedem Shop abgefragt, sondern es fand eine Zuordnung entsprechend des typischen Angebots der Shops statt.

Die Preise sollten daraufhin untersucht werden, ob die Preisgestaltung von nutzerindividuellen Merkmalen beeinflusst wird. Dazu zählen besonders, aber nicht ausschließlich, „technische“ Faktoren wie die Nutzung unterschiedlicher Endgeräte, Betriebssysteme, Browser oder das Setzen unterschiedlicher Datenschutzeinstellungen. In der Summe ergaben sich hier mindestens 100 relevante Kombinationen. Die Preise wurden zudem abhängig vom Standort des Nutzers untersucht. Hierfür wurden IP-Adressen aus unterschiedlichen Regionen Deutschlands genutzt. Die Preise wurden auch in Abhängigkeit der Surfhistorie und/oder Social-Media-Aktivitäten der Nutzer untersucht.

In einem zahlenmäßig breiten Strang erfolgte die Datenerhebung automatisiert über eine selbst programmierte Software-Lösung, die für eine große Zahl virtueller Nutzer regelmäßig bei den untersuchten Seiten und Produkten die Preisinformationen erhob und zusammen mit den im jeweiligen Erhebungsszenario unterstellten Nutzungsfaktoren ablegte. Die oben dargestellten Eckwerte führten zu einer Zahl von rund 75.000 Datenpunkten pro Tag, das heißt rund sieben Mio. Datenpunkten im Erhebungszeitraum der drei Monate. Durch den entsprechend großen Stichprobenumfang wird gewährleistet, dass die in der anschließenden Auswertung gewonnen Aussagen statistisch belastbar sind. Außerdem kann so sichergestellt werden, dass die personalisierte und die rein dynamische nutzerunabhängige Preissetzung mit einem akzeptablen statistischen Konfidenzniveau voneinander unterschieden werden können.

In einem zahlenmäßig schmäleren Strang tätigten natürliche Personen manuelle Abfragen. Diese decken das Spektrum von „wenig kaufkräftig“ bis „sehr kaufkräftig“ ab und bringen sehr unterschiedliche Surf- und Kaufhistorien mit sich. Es konnte dabei aber nicht die volle Kombinatorik z. B. aller Endgeräte, Betriebssysteme, Browser und Standorte getestet werden.

Für die Auswertung wurde zunächst eine deskriptive Analyse der erhobenen Preisdaten vorgenommen. Im Zentrum der Analyse stand aber dann die multivariate Untersuchung der erhobenen Daten. Hierfür wurde die multiple lineare Regression herangezogen. Diese ermöglicht die Erklärung eines abhängigen Faktors (nämlich der relativen Preisdifferenz) durch mehrere unabhängige Faktoren (nämlich der zu untersuchenden Kriterien und deren Ausprägungen).

Dabei ist zu beachten, dass es nicht auf die Veränderung des Preises an sich ankommt. Entscheidend sind vielmehr Preisabweichungen. Daher wurde ein „Standardfall“ definiert, der als Grundlage für Preisvergleiche dient. Dieser entsprach einem Nutzer ohne Surf- und Kaufhistorie, der an einem Windows-Desktop-PC mit Google Chrome zum ersten Mal die entsprechende Seite besuchte.

Die verwendete relative Preisdifferenz berechnet sich aus dem prozentualen Preisunterschied des jeweiligen Preises zum jeweiligen Standardfall. Folglich weisen Standardfälle als Bezugspunkt per Definition keinen Preisunterschied auf.

$$\text{Relative Preisdifferenz} = \frac{\text{Preis}_{\text{produkt}} - \text{Preis}_{\text{standardfall}}}{\text{Preis}_{\text{standardfall}}}$$

Durch die Verwendung von relativen Differenzen ist zusätzlich die Vergleichbarkeit der berechneten Preisunterschiede über alle Produkte und möglichen Konfigurationen hinweg gewährleistet.

3 Wesentliche Ergebnisse

3.1 Umfang der erhobenen Daten

Vom 9. Juni bis zum 4. Oktober 2020, also über einen Zeitraum von fast vier Monaten, wurden durch die Preisabfragesoftware automatisiert Preise abgefragt und aufgezeichnet. Der manuelle Abfragestrang wurde zum 6. Juli 2020 gestartet und bis zum 4. Oktober 2020 durchgeführt. Insgesamt waren 21 natürliche Personen mit der manuellen Abfrage beschäftigt.

Der gewählte Versuchsaufbau ergab umfangreiches Datenmaterial für die Auswertung. Durch die beiden Erhebungsstränge (automatisiert/manuell) sowie den langen Erhebungszeitraum konnten insgesamt über 400.000 Datensätze mit rund sieben Millionen Datenpunkten erhoben werden. Rund 250.000 Datensätze entfielen auf die automatisierte Erhebung, rund 153.000 auf die manuelle Abfrage.

Die Grundlage der Auswertung bildeten nicht die erhobenen absoluten Preise, sondern die relative Preisdifferenz zum definierten Standardfall, bei dem die Plattformen mit hoher Sicherheit keine Preispersonalisierung betreiben. Deshalb wurde in jedem einzelnen Erfassungsvorgang dieser Standardfall erhoben und im Rahmen des Preprocessing die relativen Preisdifferenzen berechnet. Die Anzahl der ausgewerteten Datensätze betrug damit rund 286.000. Von diesen wiesen rund 9.500 eine Preisdifferenz auf (rund 3,3 Prozent).

3.2 Deskriptive Analyse

Bereits die erste Betrachtung der Daten ergab, dass ein Großteil dieser Preisabweichungen auf nur ein einziges Produkt bzw. eine einzige Plattform zurückzuführen war, nämlich der Buchung eines Hotels über Booking.com.

Die übrigen Preisdifferenzen waren extrem gering und bewegten sich in einem Spektrum von -0,04 Prozent bis +0,03 Prozent. Überwiegend waren diese Differenzen negativ. Es wurden also – im Fall des Auftretens einer Differenz – geringere Preise veranschlagt als im Standardfall. Auch bei der Betrachtung der relativen Preisdifferenzen in Abhängigkeit von einzelnen Einflussfaktoren änderte sich dieser Eindruck nicht.

Die Abweichungen bei der Buchung des Produkts „IntercityHotel Hamburg“ über die Plattform „Booking.com“ sind erläuterungsbedürftig. Sie konnten darauf zurückgeführt werden, dass das Hotel bei der Buchung über mobile Endgeräte einen Sondertarif, die sogenannte „Handy-Rate“, gewährte, um die Konversionsrate zu erhöhen. Offenbar wurde in der Mehrzahl der mobilen Buchungen ein Rabatt von 10 Prozent gewährt.

3.3 Regressionsanalyse

Als Hauptbestandteil der Auswertung kam die multivariate lineare Regressionsanalyse in Form der Kleinste-Quadrat-Methode zum Einsatz. Um ein möglichst aussagekräftiges Ergebnis zu erzielen, wurde der vorliegende Datensatz mehrfach sinnvoll beschnitten und jeweils ausgewertet.

Zunächst wurde eine Regression mit allen zur Verfügung stehenden rund 286.000 Datensätzen gerechnet. Die dabei berechneten Koeffizienten waren nahezu durchgehend nur unwesentlich von null verschieden. Das bedeutet, dass der Einfluss der betrachteten unabhängigen Variablen auf den abhängigen Faktor „relative Preisdifferenz“ extrem gering ausfiel. In zwei Feldern gab es Hinweise auf systematische Differenzen. Zum einen ergab sich für das Produkt „IntercityHotel Hamburg“ eine Abweichung von über 3 Prozent, zum anderen zeigte sich ein signifikanter (negativer) Einfluss bei Zugriffen über mobile Endgeräte.

Es zeigte sich bei der deskriptiven Analyse, dass es Produkte gab, bei denen überhaupt oder nahezu keine Preisdifferenzen auftraten. Um die Qualität der Analyse zu verbessern, wurde diese in einem nächsten Schritt aus dem Datensatz entfernt. Es ergaben sich insbesondere bei den Zugriffen über Mobilgeräte Koeffizienten zwischen -0,0152 und -0,0114, also Abweichungen zwischen -1,52 und -1,14 Prozent. Dabei lag auf Grundlage der Ergebnisse der deskriptiven Analyse zu vermuten, dass hier das Produkt „IntercityHotel Hamburg“ einen hohen Einfluss auf die Modellergebnisse ausübte.

Um die Modellqualität weiter zu verbessern, wurde daher der Datensatz getrennt und zunächst nur das Produkt „IntercityHotel Hamburg“ betrachtet, anschließend der restliche Datensatz ohne das Produkt. Hier wurde auch jeweils der (Teil-)Datensatz des manuellen Erhebungsstrangs zusätzlich ausgewertet, da nur in diesem Strang die weichen qualitativen Faktoren wie der Login in ein Nutzerkonto oder Login in Social-Media-Konten abgefragt wurden.

Bei der Betrachtung des Produkts „IntercityHotel Hamburg“ wurden die Eindrücke aus der deskriptiven Analyse bestätigt. Es ergaben sich Koeffizienten zwischen -6 und -11 Prozent; diese waren auch signifikant. Nutzerkonto- oder Social-Media-Logins hatten dagegen offenbar keinen Einfluss.

Die Regressionsanalyse des nun final bereinigten Datensatzes ergab keine auffälligen Koeffizienten. Bei einigen Variablen traten Preisunterschiede auf, zum Teil waren diese auch signifikant. Bei der Betrachtung der Daten des manuellen Erhebungsstrangs ergab sich ein nahezu identisches Ergebnis. Den größten signifikanten Einfluss hatte hier die Verwendung von Cookies, die zu einer Preisdifferenz von -0,2 Prozent führte.

Mit Bestimmtheitsmaßen (R^2) von 0,1 bzw. 0,3 Prozent ergab sich, dass nur ein verschwindend geringer Anteil der vorhandenen Preisdifferenzen durch das Regressionsmodell erklärbar sind. Die Schlussfolgerung liegt nahe, dass die betrachteten unabhängigen Variablen keinen systematischen Einfluss auf die relative Preisdifferenz haben.

Von Interesse war zuletzt noch die Betrachtung des vollständigen manuell erhobenen Datensatzes, denn nur dort wurden auch die genannten „weichen“ Faktoren erhoben. Dabei hatte aber weder der Login in ein Nutzerkonto der Plattform noch der Login in soziale Medien einen merklichen Einfluss auf die Preisdifferenzen.

3.4 Fazit

Die zu Beginn der Untersuchung formulierten Forschungsfragen waren: Wird im Markt personalisierte Preisgestaltung verwendet? Von welchen Kriterien hängt – für den Fall der Verwendung – die personalisierte Preisgestaltung ab und welches Ausmaß haben ihre Auswirkungen? Die vorliegende Untersuchung hat sich dem Thema empirisch genähert und mittels einer breit angelegten Datenerhebung im deutschen E-Commerce erstmals systematisch analysiert.

Personalisierte Preise sind dabei zu unterscheiden von in der Regel auf Marktgegebenheiten zurückzuführenden Preisänderungen. Personalisierte Preisgestaltung liegt dann vor, wenn ein Kunde auf Basis von persönlichen Daten einen individuellen Preis angezeigt bekommt, der anderen Kunden nicht angeboten wird.

Die durchgeführte statistische Auswertung der gesammelten Daten konnte nicht bestätigen, dass personalisierte Preisgestaltung im Markt vorliegt und von Anbietern aktiv eingesetzt wird.

Im Datensatz ließ sich ein Einfluss des verwendeten Endgeräts auf die relative Preisdifferenz signifikant belegen. Durch eine vertiefte Regressionsanalyse ließ sich dieser Effekt auf ein Produkt bzw. eine Plattform im Datensatz eingrenzen. Die deskriptive Analyse wurde damit bestätigt. Beim fraglichen Sachverhalt handelt es sich offenbar um marktübliches Verhalten.

Nicht bestätigt werden konnte ein Einfluss der Do-not-track-Einstellung in Browsern auf den Preis, ebenso wenig der Einfluss des Setzens oder Nicht-Setzens von Cookies. Die Ergebnisse waren hier nur zum Teil signifikant und dann nur geringfügig von null verschieden.

Untersucht wurde ebenfalls, ob es Preisdifferenzen gibt, wenn der Nutzer in ein Nutzerkonto beim jeweiligen Anbieter bzw. der jeweiligen Plattform oder auf einem sozialen Netzwerk wie Facebook eingeloggt ist. Ein Einfluss dieser Faktoren konnte nicht belegt werden. Die Ergebnisse waren durchweg nicht signifikant.

Eine ganze Reihe weiterer potentieller Einflussfaktoren wurde untersucht, bei denen wiederum kein Einfluss auf die angezeigten Preise festgestellt werden konnte. Dazu zählten das verwendete Betriebssystem des für den Kauf benutzten Endgeräts, der verwendete Browser, der Wochentag oder die Uhrzeit des Kaufs und der auf Basis der IP-Adresse übermittelte Standort.

Impressum

Empirie zu personalisierten Preisen im E-Commerce – Ergebniszusammenfassung

ibi research an der Universität Regensburg GmbH:

Holger Seidenschwarz

Johann Faltermeier

Quirin Wierer

Nils Deichner

Dr. Georg Wittmann

trinnovalive GmbH:

Dr. Andreas Beer

Dr. Johann Schenkl

Kontakt

ibi research an der Universität Regensburg GmbH

Galgenbergstraße 25

93053 Regensburg

Tel.: 0941 943 1901

Fax: 0941 943 1888

E-Mail: info@ibi.de

Internet: www.ibi.de

trinnovalive GmbH

Franz-Mayer-Straße 1

93053 Regensburg

Tel.: 0941 604 889 450

Fax: 0941 604 889 459

E-Mail: info@trinnovalive.de

Internet: www.trinnovalive.de