

Hipster versus Quant

Wie anwendbar ist maschinelles Lernen im Asset-Management?

Dr Bernd Scherer*

7. Januar 2020

Zusammenfassung

Hipster versus Quant, Daten versus theoretische Modelle, induktive versus deduktive Methode? Die Frage nach dem Primat der Methodik ist so alt wie die Wissenschaften selbst. Ausgelöst durch Erfolge von maschinellem Lernen in anderen Industrien hat der von Breiman (2001) ausgerufene Kulturkampf zwischen daten- und modellgetriebener Statistik nun auch das Asset-Management erreicht. Welche Aufgabe (n) kann maschinelles Lernen im Asset-Management übernehmen? Dieser Frage wird im folgenden kurzen Aufsatz nachgegangen. Ich beschränke mich dabei auf den Bereich des Portfoliomanagements und klammere andere Anwendungen („business analytics“) aus. Hier unterscheidet sich die Asset-Management Industrie (bis auf regulatorische Auflagen) nur unwesentlich von anderen Industrien. Mögliche Anwendungsgebiete sind das automatisierte Lesen von Lebensläufen, legalen Dokumenten und Rechnungen oder die Klassifizierung von Kunden und die Berechnung eines „customer lifetime value“. Darüber hinaus werden die spezifischen Eigenschaften und die damit einhergehenden Herausforderungen bei der Analyse von ökonomischen Daten im Hinblick auf die Methodenwahl diskutiert. Um das Thema neben Quant und Hipster auch einer möglichst breiten Zuhörerschaft zugänglich zu machen, habe ich weitgehend auf formale Darstellungen verzichtet und eine Vielzahl von Beispielen eingearbeitet. Bis auf wenige zitierte Ausnahmen verarbeitet dieser Aufsatz sowohl meine persönlichen Erfahrungen und Ideen, als auch die Ergebnisse vieler Gespräche der letzten Monate mit „peers“ in London, Frankfurt und New York.¹

*Research Associate at EDHEC Risk, Nice (bernd.scherer@edhec.edu), Associate Editor für das Journal of Asset-Management und das Journal of Systematic Investment Strategies, Geschäftsführer im Asset-Management einer deutschen Privatbank.

¹Ich danke den vielen inhaltlichen Kommentaren von Heinz Kasten, Kristian Kersting, Tobias Klein, Christopher Krauss und Thomas Zwirner.

Wissenschaftstheorie: Induktivismus Maximus

Maschinelles Lernen (ML) in Reinform² stützt sich auf einen radikalen Induktivismus. SAP definiert maschinelles Lernen auf seiner website als:

„Maschinelles Lernen nutzt ausgefeilte Algorithmen, um aus enormen Big-Data-Mengen zu lernen. Je größer die Datenmenge, auf die die Algorithmen zugreifen können, desto mehr lernen sie.“

Die Anhänger von ML sind davon überzeugt, mit minimaler Kenntnis des (ökonomischen) Problems, ohne Variablenvorauswahl, ohne theoretisches Modell, ohne Vermutungen über die funktionale Form des Zusammenhangs und mit alleiniger Sicht auf die Daten, komplexe Strukturen und Muster erkennen zu können. Alleiniger Zweck des Modells sind bessere Prognosen und der einzige zulässige Test liegt in der Feststellung der Prognosegüte. Tests zu funktionaler Form, Endogenität, Vorzeichen und Signifikanz von Parametern sind damit irrelevant. Man wird diese vergeblich in Lehrbüchern oder Veröffentlichungen zu ML suchen.

Dies ist in starkem Kontrast zur deduktiven Methode in der Ökonometrie. Für den Ökonomen ist ein Modell das Ergebnis der optimalen Entscheidungen rationaler Individuen. Parameter, Variablen und funktionale Form sind das Ergebnis eines theoretischen Modells. Die Theorie bestimmt das Modell, die Daten bestimmen nur noch die konkreten Werte der geschätzten Parameter. Die Theorie liefert testbare Hypothesen für das statistische Modell. Beurteilt wird das Modell und nicht nur dessen Prognosefähigkeit.³

Dort, wo Daten im Überfluss vorhanden sind, oder durch Simulationen und Experimente geschaffen werden können, hat ML in den letzten Jahren besondere Fortschritte erzielt. Bahnbrechende Anwendungen finden sich im Bereich der Computervision (Bild- und Personenerkennung), Datensicherheit, Betrugserkennung, Personalisierung von Botschaften, Empfehlungssystemen oder „online“-Suche. Dabei liegt der Fokus auf konsumentenbezogenen Anwendungen, d.h. auf der Lösung konkreter, kommerziell vielversprechender Probleme. Lassen sich diese Erfolge auch auf die Wirtschaftswissenschaft⁴ übertragen?

Wie viel (Wirtschafts-) Wissenschaft steckt in maschinellem Lernen? Korrelation versus Kausalität!

Ich beginne mit einem Beispiel aus den Naturwissenschaften. Nehmen wir an, wir hätten ein neuronales Netz trainiert, das alle Planetenbewegungen im Sonnensystem perfekt prognostiziert. Wäre das wissenschaftlicher Fortschritt? Unser Verständnis des Sonnensystems hätte sich nicht verbessert. Wir würden immer noch nicht die Ursache der Planetenbewegungen kennen. Schlimmer noch: das ML Modell wäre nicht in der Lage, mit kontrafaktischen Fragen umzugehen. Welche Flugbahn würde ein plötzlich auftauchender Meteor nehmen? Welchen Einfluss hat ein schwarzes Loch auf die Planetenbewegungen? Ein ML Modell könnte uns nicht

²Anhänger maschinellen Lernens verweisen gerne darauf, dass es diese Reinform in der angewandten Praxis gar nicht gibt. Dennoch bezieht sich dieser Beitrag in großen Teilen auf diesen Referenzpunkt. Zum einen, weil es einen Methodenvergleich erleichtert und zum anderen, weil ich in meiner Tätigkeit als Editor und Gutachter akademischer Zeitschriften immer wieder auf eingereichte Aufsätze gestoßen bin, in denen ein völliger Mangel an Kenntnis der zugrundeliegenden ökonomischen Zusammenhänge von der ersten Zeile an evident war, so dass maschinelles Lernen in seinen extremen Ausprägungen eher die Regel als die Ausnahme ist. Auch in Kaggle Wettbewerben ist es nicht unüblich einen Datensatz mit 1000 Variablen ohne sinnvolles Label und ohne Datenbeschreibung zu erhalten. Einen sinnvollen, für Laien verständlichen Einstieg in ML (aus Sicht eines erfahrenen Statistikers) liefert Spiegelhalter (2019).

³Sendhil and Spiess (2017) und Varian (2014) liefern einen guten Überblick über maschinelles Lernen aus Sicht der empirischen Wirtschaftswissenschaft (Ökonometrie).

⁴Der vorliegende Beitrag beschäftigt sich fast ausschließlich mit den Anwendungen von ML im Asset-Management bzw. in den Wirtschaftswissenschaften.

helfen, ein theoriegeleitetes und an Daten überprüfetes Modell schon.⁵ Auch in den Wirtschaftswissenschaften sind Daten ohne Theorie für unser Verständnis der ökonomischen Welt weitgehend nutzlos. So schreibt der Träger der Turing Auszeichnung in seinem Buch:

„This is why deep-learning systems (...) will never be able to answer questions about interventions, which by definition break the rules of the environment the machine was trained in.“⁶

Maschinelles Lernen findet daher kaum Platz in der wirtschaftswissenschaftlichen Literatur. Der Wissenschaftler ist nicht an der Frage interessiert, ob es einen statistischen Zusammenhang (Korrelation) gibt, sondern an der Frage, warum es diesen Zusammenhang gibt (Kausalität).

Größtes Problem maschinellen Lernens in allen wirtschaftswissenschaftlichen Anwendungen sind endogene Variablen und die sogenannte Lucas-Kritik von Lucas (1976), da diese auch die Prognosefähigkeit von ML Modellen untergräbt. An dieser Stelle macht es Sinn formaler auf Endogenität (dem „ground zero“ für ML) einzugehen. Gegeben sei ein einfaches lineares Modell der Form

$$y = \alpha + \beta x + \varepsilon \tag{1}$$

Um die Koeffizienten konsistent schätzen zu können (die Schätzung konvergiert in großen Datenmengen gegen den wahren Wert), ist die Kernannahme des Modells, dass sich neue Daten nur in den Werten für y (und damit in ε) unterscheiden. Die Werte für x bleiben gleich, d.h. sind nicht stochastisch, sondern exogen gegeben (werden außerhalb des Modells bestimmt). In empirischer Arbeit ist Endogenität allgegenwärtig und die größte Sorge des Ökonometrikers. Hier ein einfaches Beispiel. Der Preis für Eiskrem (x) ist ein Prediktor für die Nachfrage nach Eiskrem (y). Aus unserer Kenntnis der Mikroökonomie würden wir erwarten, dass ein höherer Preis zu einer geringeren Nachfrage führt, d.h. wir würden erwarten, dass $\beta < 0$ ist. Das wird immer dann der Fall sein, wenn unsere Daten einem kontrollierten Experiment entstammen, d.h. ich beobachte für ein gegebenes, exogenes x (Preis von Eiskrem) das resultierende y (Nachfrage nach Eiskrem). Ökonomische Daten sind in den wenigsten Fällen exogen, sondern oftmals das Ergebnis strategischen (endogenen) Verhaltens. Wenn die Nachfrage nach Eiskrem vom Wetter abhängt, wird der gewinnmaximale Preis bei schönem Wetter (hohe Nachfrage) höher sein, als bei schlechtem Wetter. Der geschätzte Koeffizient wäre dann positiv, da in den realen Daten hohe Preise mit hoher Nachfrage einhergehen. Unsere Schätzung wäre inkonsistent und nicht für kausale Analysen (optimale Preispolitik) geeignet. ML Methoden vernachlässigen Endogenität meist komplett (hohe Preise sind nicht für hohe Nachfrage kausal).⁷ Endogenität entsteht regelmäßig durch strategisches Verhalten, latente oder ausgelassene Variable und Messfehler in den erklärenden Variablen. Die Ökonometrie hat Lösungen entwickelt (Instrumentvariablen und Two-Step Least Squares Schätzung), ML nicht. Noch härter ausgedrückt: ML identifiziert in dieser Situation kein verallgemeinerbares Wissen.

In die gleiche Kerbe schlägt die Lucas-Kritik. Sie besagt, dass sich der ökonomische Zusammenhang mit veränderten Erwartungen der Marktteilnehmer und deren endogen bestimmter Handlungen verändert. Ohne ein theoretisches Modell, dass diese Erwartungen explizit formuliert kommt man hier nicht weiter. Populärwissenschaftliches Beispiel für diese Verkürzung ist Bundeskanzlerin Angela Merkel. Sie hatte geäußert

⁵Ptolomäus schuf das geozentrische Weltbild (Erde im Mittelpunkt) mit Hilfe eines Modells (ähnlich einer Fourieranalyse), dass er fast perfekt an die beobachteten Planetenbahnen anpassen konnte. Sein rein empirische Modell war den Vorhersagen des heliozentrischen Weltbilds (Sonne im Mittelpunkt) des Kopernikus im 16. Jahrhundert, weit überlegen. Selbst ein besserer Fit ist nicht alles.

⁶Aus dem „Book of Why“ von Pearl and MacKenzie (2018), Seite 32, Mitte. Eine formālere Abhandlung liefert Pearl J., Glymour M. und N. Jewell (2016).

⁷Der Anhang beschreibt die Nachfrage nach Eiskrem in einem formaleren Modell und zeigt den Endogenitätsbias für maschinelles Lernen (eines „Datenwissenschaftlers“) in Reinform.

nicht zu verstehen, wie bloße Steuerankündigungen schon heute Arbeitsplätze kosten können. Dies hat dann einen deutschen Professor für Volkswirtschaft zu folgendem Kommentar verleitet; „Frau Merkel versteht das nicht. Sie ist Physiker, kein Ökonom. Sie hat gelernt: erst Stein, dann Welle. Eine Umkehrung existiert in der Physik nicht.“ Aber nicht alle Physiker sind gleich. Prof. Richard Feynman, einer der besten Physiker der Welt, hat einmal gesagt: „Stellen Sie sich vor wie schwierig die Physik wäre, wenn Elektronen Gefühle hätten.“⁸ Und noch ein letztes Beispiel zum Thema Endogenität. Der Sozialstatus eines Menschen ist ein Prediktor für seinen Gesundheitszustand. Auf den ersten Blick erschüttert dies, aber für den Wissenschaftler ist das alleine nicht interessant. Interessanter (um Politik sachlich zu begründen) ist es zu wissen, ob eine Veränderung des Sozialstatus, den Gesundheitszustand verbessert. Wenn der Sozialstatus durch Intelligenz oder Ausbildung bestimmt wird (latente Variable), muss man das bezweifeln. Ein Lottogewinner aus einem sozialen Umfeld mit schlechten Angewohnheiten mag durch den Gewinn seinen Sozialstatus erhöhen. Gesünder wird er dadurch nicht. Er hat jetzt nur mehr Geld um seine unreflektierten Angewohnheiten (Alkohol, Drogen und Solarium) auszuleben. Der Ökonometriker kann das testen, der Datenwissenschaftler kann es nicht testen (und der Politiker will es nicht testen).

Aus Sicht des Ökonomen stellt maschinelles Lernen ein einfaches „reduced form model“ dar. Diese Modelle sind nicht in der Lage Verhaltensänderungen von Wirtschaftssubjekten mit rationalen Erwartungen zu erfassen. „Reduced form“ Modelle finden sich zwar auch häufig in der Finanzmarktökonomie (Vektorautoregressionen), sind aber nur dann interessant (veröffentlichungswert), wenn ein theoretisches Modell existiert, dass testbare Restriktionen für Parameter(kombinationen) des „reduced form“ Modells enthält. Nur dann, kann ein theoretisches Modell getestet werden.

Wann hat ML Methodik wesentliche Vorteile gegenüber traditioneller Ökonometrie? Immer dann, wenn der Fokus auf Prognose, statt auf Kausalität liegt und immer dann, wenn große Datenmengen mit einer Vielzahl von erklärenden Variablen (aber kein theoretisches Modell) vorliegen (oder wenn Daten unstrukturiert und gemischt d.h. nominal, ordinal und metrisch sind). Die Vorteile reichen hier von numerischen Vorteilen (Matrix Algebra versus „stochastic gradient search“), über Vorteile bei der Bestimmung wichtiger Variablen, Regularisierung von Schätzungen (Strafterme für Komplexität), Möglichkeiten mehr Variablen als Datenpunkte zuzulassen, bis zu Methoden der Kreuzvalidierung, die bislang in der Ökonometrie wenig Beachtung finden.

Der Ökonom dagegen, wird immer auf der Suche nach einem kausalen ökonomischen Modell bleiben, um kontrafaktische Szenarien (noch nicht gesehenen Daten außerhalb des bisherigen Wertebereichs) zu bewerten. Daten ohne Wissenschaft sind aus Sicht des Ökonomen im günstigsten Fall schlechte Soziologie. Die Ökonometrie als Kombination von Ökonomie und Statistik hat sich damit schon vor Jahrzehnten beschäftigt. Domainspezifisches Wissen und theoretische Modelle wurden in das statistische Modell integriert. Damit ist der Ökonometriker schon lange „Datenwissenschaftler“ in ökonomischen Fragen.

Zugang und Aufbereitung neuer Daten im Asset-Management

Die sogenannte „Quantkrise“ im August 2007 löste das Interesse an alternativen Daten und alternativen Modellen aus. Damals hatte die Positionsreduktion eines großen Marktteilnehmers mit für quantitative Investoren typischen Holdings einen kaskadenartigen, kombinierten Abverkauf quantitativer Strategien erzwungen. Alternative Modelle hatten sich nach kurzer Zeit als weniger zielführend erwiesen. Quantitatives Investieren ist durch Daten getrieben, d.h. alternative Modelle haben oft die gleichen Anomalien mit unterschiedlichen

⁸Berichtet von Lo (2017) in seinem Buch „Adaptive Markets“.

Ansätzen gefunden.

Vielversprechender scheint es daher kollektive aktive Positionierungen durch die Verwendung alternativer Daten zu vermeiden. Diese dienen beispielsweise dazu Informationen früher als andere Investoren zu erhalten („nowcasting“ von Unternehmensgewinnen durch höherfrequente Daten wie Kreditkartenabrechnungen oder Lokationsdaten von Kunden) oder zusätzliche Dimensionen eines Unternehmens wie Kundenfranchise und Mitarbeiterzufriedenheit (erfasst durch Textanalysen aus Blogs und sozialen Medien) abzuschätzen. Zur Verfügung stehen hier Textdaten (Blogs, Nachrichten, Unternehmensberichte), Lokations- und Klickdaten (Kundenkontakte), visuelle Daten (Beladung von Schiffen, Besetzung von Parkplätzen), Sprachdaten (Analystencalls, Pressekonferenzen). Der Kreativität des Analysten sind wenig Grenzen gesetzt.⁹ Es bietet sich beispielsweise an durch ML bessere „peer groups“ verwandter Aktien (traditionell als Industrie bezeichnet) zu finden. Man kann die Quartalsberichte nach gemeinsamen ökonomischen Themen untersuchen oder alle von einem Analysten verfolgten Aktien bündeln. Im Handel kann der Nachrichtenfluss als Prediktor von Handelsvolumen verwendet werden. Momentum kann ergänzt werden, indem Unternehmen nach Nachrichtenflüssen klassifiziert werden. Wir können Klimahedgeportfolios bauen, indem wir Portfolios konstruieren, die Innovationen im Nachrichtenfluss zu Umweltthemen möglichst eng „tracken“ (sogenannte „mimicking portfolios“). Einzig echte Grenze sind Datenbudget, Methodenkenntnis und Kreativität des Analysten.

Im Bereich von Datenbeschaffung und Interpretation kann eine starke AI („artificial intelligence“) Abteilung, eine Armee von Analysten ersetzen. ML liefert letztlich unendlich viele digitale Analysten. Die Produktivitätsgewinne sind enorm und der Mehrwert von ML durch die Fähigkeit des computergestützten Sehen, Hören und Lesen kann gar nicht überschätzt werden. Wir befinden uns hier sicher in einer radikal veränderten Welt, in der große Vermögensverwalter massive Größenvorteile (hohe Fixkosten eines guten Research-Teams) beim Sammeln, Testen und Evaluieren von neuen Datensets besitzen.

Die Herausforderungen an die entsprechenden Teams sind enorm. Neue Datenquellen, die oftmals gar nicht für den Zweck des Investierens gesammelt wurden, müssen evaluiert werden. Dazu verwendet man typischerweise nicht aggregierte Daten (Rohdaten), da jede Aggregation Informationen vernichtet und der Aggregator den Verwendungszweck der Daten bestenfalls ungenau kennt. Wer in Ermangelung von Ressourcen auf aggregierten Daten aufsetzt hat schon den ersten Nachteil. Die Daten sind oftmals unrein, fehlerhaft und unstrukturiert. Genaue Zeitstempel der Verfügbarkeit ohne „backfill“ (sogenannte „point in time“ Daten) sind ebenfalls nicht immer klar. Zudem sind Historien kurz und dadurch schwer zu normalisieren. Man kann nicht beurteilen ob das Weihnachtsgeschäft 2019 besonders gut war, wenn man nur zwei Weihnachten beobachtet hat.

Zusammengefasst: ML ist für Datenzugang und Datenaufbereitung (Informationsextraktion) nicht wegzudenken und hat sich bereits fest im Asset-Management etabliert. Das verfügbare Wissen ist unternehmensspezifisch und eher bei großen Asset-Managern vorhanden. Mit welchen Methoden die neu gefundenen Daten verarbeitet werden sollten, ist dagegen weniger eindeutig. Der nächste Abschnitt wird sich diesem Problem zuwenden.

⁹Siehe hierzu im Detail Guida (2019).

Bessere Prognosen durch ML im Asset-Management

Worin unterscheiden sich Finanzmarktdaten?

Den Investmentpraktiker wird die wissenschaftstheoretische Diskussion über die Grenzen des Induktivismus wahrscheinlich langweilen. Das Versprechen auf bessere Prognosen klingt für ihn gut genug. Wissenschaft nimmt für den Praktiker nur einen geringen Wert ein. Reich sein, ist besser als Recht zu haben. Verfolgt man den Hype um ML, ausgelöst durch selbstfahrende Autos und Go spielende Rechner oder automatisierte Waffensysteme, fragt man sich, warum nicht schon längst alle Portfolios mit Hilfe von ML verwaltet werden? Die Prognose von Finanzmarktdaten unterscheidet sich von anderen ML Anwendungen im wesentlichen durch drei Problemfelder: *Nichtstationarität, begrenzte Anzahl verfügbarer Datenpunkte und Messfehler in den Zustandsvariablen.*

Beginnen wir mit der zeitlichen Instabilität (Nichtstationarität) des unbekanntem statistischen Zusammenhangs in Finanzmarktdaten. Stationarität bedeutet, dass meine Schätzung des unbekanntem Zusammenhangs mit zunehmender Anzahl von Datenpunkten immer besser wird. Der Zusammenhang wartet nur darauf entdeckt zu werden. Seine Entdeckung lässt den Zusammenhang unberührt. Für viele ML Aufgaben außerhalb des Asset-Managements ist diese Voraussetzung gegeben. Ein Algorithmus, der Bilder mit Katzen finden soll und im Jahr 2013 trainiert wurde, wird auch in 2019 unverändert gut arbeiten. Gleiches gilt auch für selbstfahrende Autos (es sei denn sie wurden in England trainiert und fahren zum ersten Mal in Kontinentaleuropa). Für Finanzmarktdaten ist diese Stabilität nicht vorstellbar (allein die Entdeckung des Zusammenhangs hat Auswirkungen auf seine zukünftige Existenz).

Das zweite Problem ist der erforderliche Freiheitsgrad (hohe Anzahl an Variablen, Parametern und Hyperparametern und deren nichtlineare Interaktion) von ML Algorithmen, der auf eine begrenzte Anzahl von Daten trifft. Für den S&P500 selbst, existieren etwa 23850 tägliche Datenpunkte (90 Jahre mit 250 Beobachtungen pro Jahr). Das ist für ML vergleichsweise wenig, vor allem weil es der einzige Datensatz ist den der Ökonom hat. Er darf ihn nicht für die Entwicklung seiner Theorie nutzen, weil nachfolgende Tests auf dem gleichen Datensatz bedeutungslos wären. In den Naturwissenschaften, wo viele Daten im Überfluss vorhanden (Genetik) oder einfach experimentell erzeugt werden können, gilt diese Einschränkung nicht. Der folgende Abschnitt befasst sich mit der Frage, welche Anwendungen von ML im Asset-Management zielführend sind und welche nicht. Die Stärke von ML Algorithmen ist der Umgang mit großen Datenmengen. Für ML Anwendungen ist es kein Problem wenn die Anzahl der Variablen größer als die Anzahl der Datenpunkte ist. Es stehen eine Reihe von Methoden zur Identifikation der einflussreichsten Variablen zur Verfügung, Kreuzvalidierung und Prognosediversifikation über Modelle sind weit fortgeschritten und in den meisten Modellen kann Modellkomplexität bestraft werden. Es ist aber auch richtig dass extremer Induktivismus, kombiniert mit Modellen, die eine hohe Anzahl an Freiheitsgraden besitzen, viele Daten benötigt. ML Anwendungen sind daher dort erfolgsversprechend, wo eine große Zahl von Daten existiert. Dies ist bei Aktienselektionsmodellen mit Querschnittsdaten (großes Universum von 5000 Einzelaktien oder mehr) oder im intradaytrading (Tausende von Transaktionen pro Tag) der Fall.

Das dritte Problem betrifft Messfehler in den Zustandsvariablen. Die Stellung auf dem Schach- oder Go-Brett (Zustandsvariable) kann mit Sicherheit festgestellt werden. Anders ist das mit Finanzmarktdaten. Der Zustand einer Volkswirtschaft (Stellung im „business cycle“) ist nur mit Rauschen zu messen, Viele Variablen sind eigentlich latent und werden nur mit Hilfe von Proxies gemessen oder durch dynamische latente Faktormodelle gefiltert.

Querschnittsdaten

Zarte akademische Evidenz existiert vor allem im Bereich der Aktientitelselektion, wie beispielsweise Kelly et al. (2019) zeigen. Dies kommt ML in methodischer Sicht entgegen, da viele ML Methoden für Querschnittsdaten entwickelt wurden. In der Literatur verwendet wurden „neural nets“, „support vector machines“, „regression trees“ oder „random forests“ mit zum Teil sehr guten Erfolgen. Modelle des ML haben sich traditionellen Methoden zum Teil deutlich überlegen gezeigt. Wie „sample“ spezifisch diese Ergebnisse sind, wird zur Zeit von vielen Praktikern und Wissenschaftlern untersucht. Dennoch ist die Zahl der erfolgreichen Einzelaktien ML Fonds (mit echter live, d.h. out of sample Performance) bislang sehr überschaubar. Aus deutscher Sicht erinnert man sich noch an Catana Capital, die den deutschen Fintech Preis gewannen und dann 2011 nach einer aktiven Rendite (relativ zu ihrer Benchmark) von -41% aus dem Markt gegangen sind. Daher scheint eine Reihe von Warnungen angebracht.

Die Anwendung von ML Methoden durch sogenannte Datenwissenschaftler ist (rein) von Daten getrieben. Dieses Vorgehen entfernt unser ökonomisches Vorwissen (prior) und kommt fast wörtlich vom Mars. Das ist nur solange ein Vorteil, wie unser Vorwissen falsch ist, oder wir unendlich viele Daten besitzen. Ein prominentes Beispiel ist das „clustering“ von Aktien. Würde man sich nur auf Kurszeitreihen (PCA oder autoencoder zur Identifikation der wichtigsten Faktoren) verlassen, würde man viele Jahre benötigen um zwischen einem Pharmaunternehmen und einer Biotechnologieaktie zu unterscheiden. Je geringer die Anzahl der verfügbaren Daten umso wichtiger ist ein theoriebestimmter prior.

Ein Faktorportfolio setzt sich bewusst Risiken aus die langfristig entlohnt werden. Das Eingehen von Positionen, die dem Investor ein stark verändertes oder bisweilen negatives Exposure zu Faktoren geben, ist nicht konsistent mit dieser Sicht. Typischerweise werden ML Methoden traditionelle Methoden nicht schlagen, wenn der Zeithorizont steigt. Bei einem Prognosehorizont von 3-12 Monaten wird die Investmentwelt sehr linear. Dies entspricht der „buy and hold“ Charakteristik von Faktoren. Es stellt sich also die Frage: Wieviel Faktortiming steckt im ML Erfolg dokumentierter Strategien mit einem Zeithorizont von einem Monat? Hier ist substantielles Faktortiming zu vermuten, doch nicht nur das. Das Anlageergebnis von Faktorportfolien hängt nicht nur von der eingegangenen Faktorintensität ab, sondern von einer Vielzahl entlohnter (Aktienmarktbeta) und unentlohnter (Faktoren aus einem Risikomodell wie etwa Industriezugehörigkeit). Es ist nicht klar wie sich das Anlageergebnis in veröffentlichten Studien aufteilt, da fast alle Studien sehr nachlässig mit Portfoliokonstruktion und Risikomanagement umgehen.

Zeitreihendaten

Wie bereits angesprochen wird reines ML dort überlegen sein können, wo eine Vielzahl von Daten und bestenfalls ein vages theoretisches Vorverständnis existiert. Eine Möglichkeit ist das Lernen von Mustern zur Identifikation von „Trends“ oder „Mean Reversion“ in Einzelaktien. Fischer und Krauss (2017) trainieren ein Modell für S&P500 Aktien. Prognostiziert wird die Preisbewegung des nächsten Tages (Zielvariable ist ein 0/1 Dummy wobei alle Aktien, die besser als die Medianaktie performen den Wert 1 bekommen) Für jeden Tag und jede Aktie (nehmen wir an es seien immer genau 500) berechnen die Autoren 31 Momentumdefinitionen („features“). Nimmt man einen lookback von 500 Tagen, dann kommt man auf 250000 Datenpunkte für die Prognose der Bewegung des nächsten Tages. Implizite Annahme ist hierbei, dass das Muster für jede Aktie gleich ist und über 500 Tage stationär bleibt. Mit diesen Daten kann dann ein „deep neural net“ oder ein „random forest“ (als benchmark, da weitestgehend robust gegen „overfitting“) für Prognosezwecke geschätzt werden. Die Modellperformance der Autoren ist zum Teil überoptimistisch (simultane Signalerrechnung und

Implementierung zum Tagesende mit einer Sharpe-ratio von über 10) und bricht in den letzten Untersuchungsjahren stark ein, stellt aber im Vergleich zu anderen Studien einen interessanten Ausgangspunkt für mehr Research dar (zusätzliche features, unterschiedliche Zielvariablen und Prognosehorizonte, Portfoliokonstruktion).

Backtest Protokoll

ML Methoden haben eine potentiell hohe Zahl von Variablen und Parametern, die es erlauben sich den Daten sehr stark anzupassen. Zudem gibt es wenig theoretische Vorüberlegungen zur Überlegenheit einzelner Methoden für einen vorliegenden Datensatz. Alles wird zu einer empirischen Fragestellung. Leider findet man selbst in akademischen Veröffentlichungen keine Hinweise auf die Anzahl der getesteten und verworfenen Modelle. In vielen Fällen trifft das selbst auf das präsentierte Modell zu. Angaben zur minimalen „split size“ von „regression trees“ sind meist nicht zu finden. Das macht saubere „research governance“ zum wesentlichen Erfolgsfaktor von ML Methoden. Andererseits kann dies auch zur Stärke von ML werden. ML Anwendungen verlassen sich unter Umständen nicht auf schlechte Theorie, oder die fehlgeleitete Intuition eines Quants. Aus Sicht des Autors haben sich eine Reihe von Techniken bewährt.

- Theoretische Fundierung. Je weniger Daten zur Verfügung stehen, desto wichtiger wird ein theoretische Fundament. Dies liefert typischerweise Variablen und funktionale Form.
- Sinnvolle Variablenselektion. Eine kleine Anzahl von Variablen, für die ein theoretischer oder zumindest intuitiv plausibler Zusammenhang besteht, ist generell einem großen Variablensatz vorzuziehen. ML Algorithmen mit vielen freien Variablen werden auch in der Kombination vieler Variablen mit geringer „signal to noise ratio“ eine Struktur konstruieren können.
- Keine Modellselektion. Die Suche nach dem besten Modell ist zwecklos. Es gibt eine Vielzahl guter Modelle. Diese zu „ensembles“ zu kombinieren ist der bessere Weg.
- Adjustierung um das Testen multipler Hypothesen. Die Renditezeitreihen aller d.h. wirklich aller Modellspezifikationen werden gespeichert und anschließend verwendet um die kritischen t-Werte (nach oben) zu korrigieren. Je unkorrelierter die Anlageergebnisse einzelner Modelle waren, desto mehr Möglichkeiten hatte der ML Analyst um nach guter Performance zu suchen.
- Research Kultur. Qualitätskontrolle in der Wissenschaft lebt von externen Validierung („peer review“, Replikation). Dieser Prozess sollte so weit wie möglich nachgeahmt werden, d.h. Analysten sollten für gute Forschung, nicht für gute (Anlage-) Ergebnisse entlohnt werden.
- Randomisierung der Originalzeitreihen. Addiert man zu den Inputvariablen Zufallsterme, so lässt sich feststellen, wie wichtig eine Variable ist. Wird sie bereits bei wenig Rauschen deselektiert ist es wahrscheinlicher dass die Signalgebung zufällig ist.
- Validierung über die Zeit und alternative Universen. Echte Validierung über die Zeit (Test mit wirklich Neuen, dem Modell unbekanntem Daten) ist nur für ökonomische Daten schwer möglich (Wiederholtes „out of sample“ Testen ist nicht mehr „out of sample“). Zurückhalten von relevanten Testassets beginnt schon beim Design des Research Prozesses).

- Hypothesenklarheit. Jede Hypothese (und der verwendete Datensatz) muss zu Beginn des Research formuliert werden. Einmal verworfen, darf diese nicht reaktiviert werden. Eine ex-post Rationalisierung auf Grund von Ergebnissen muss vermieden werden
- Occams Razor. Bevorzuge regularisierte (für Komplexität bestrafte) Modelle, um die Freiheitsgrade der Modelle zu kontrollieren.

Research Governance ist aus meiner Sicht einer der wesentlichen Stellgrößen für erfolgreiches ML. Es ist für Investoren in ML Fonds auch die weitestgehend einzige Quelle, um die Qualität eines Anbieters zu evaluieren.

Interpretierbarkeit

ML Methoden werden oft als Black-Box beschrieben. Diese Eigenschaft zwingt viele Portfoliomanager ihr Modell, bei ungewöhnlich schlechter Performance, abzuschalten. Die persönlichen und rechtlichen (Karriere-) Risiken sind oftmals zu groß. Ein einfaches „drübergucken“, wird aber nicht ausreichen, um ein „deep neural net“ zu interpretieren. Generell gilt zwar, dass die Interpretierbarkeit mit zunehmender Komplexität abnimmt, dennoch existieren eine Reihe von Methoden um ML Algorithmen zu befragen. Während „decision trees“ einfach zu interpretieren sind und sogar Methoden zu deren (weiterer) Vereinfachung („pruning“) existieren sind neuronale Netze mit hoher Komplexität schwierig zu „verstehen“. In der sich schnell entwickelnden Literatur zu XAI (erklärbares maschinelles Lernen), sind aber auch hierfür Werkzeuge vorhanden. Zweck der Interpretierbarkeit sind (neben regulatorischen Anforderungen) Verifizierung, Systemverbesserungen und die Fähigkeit von einem ML System zu lernen. Was sind die verallgemeinerbaren Gesetzmäßigkeiten, die ein Algorithmus gefunden hat? Interpretierbarkeit darf aber nicht dazu führen, ein Modell zu überstimmen, nur weil dem Anwender eine einzelne Entscheidung (im Rahmen einer lokalen Approximation) nicht interpretierbar erscheint.

Portfoliokonstruktion

Portfoliokonstruktion ist das Thema, das aus Sicht des Autors am geringsten von maschinellem Lernen betroffen sein wird. Probleme der Portfoliokonstruktion basieren auf mikroökonomisch fundierter, normativer (Entscheidungs-) Theorie. Es existiert eine große Bandbreite von Modellen, die alle wesentlichen Fragestellungen adressieren: unterschiedliche Anlegerpräferenzen, Verteilungsannahmen, Risikomodelle, Transaktionskostenmodelle, Verfahren zum Umgang mit Schätzfehlern in Renditeprognosen und Risikomodelle. Das theoretische Vorverständnis innerhalb der Portfoliokonstruktion ist so groß, dass ML keine ernsthafte Alternative darstellt. Das ML Modell kann bestenfalls das theoretische Modell Lernen, aber eben nur für die trainierten Fälle. Ein potentielle Anwendung für diesen Fall sind Berechnungen die in Realzeit zu langsam sind aber von einem ML Algorithmus hinreichend genau approximiert werden können. Ein Beispiel ist die Bewertung von Optionen deren Berechnung in Realzeit zu zeitaufwendig ist (komplexe Monte Carlo Simulationen). Für dieses Problem könnte über Nacht ein Grid vieler möglicher Preise für eine Vielzahl von Parameterkonstellationen (Laufzeit, „moneyness“, Volatilität, „strike“) errechnet werden. ML berechnet dann für diesen Wertebereich eine Approximation an die Bewertungsfunktion, die am nächsten Tag fast in Realzeit abgerufen und berechnet werden kann.

Die einzige potentielle Ausnahme beim Thema Portfoliokonstruktion, besteht für das „reinforcement learning“. Hier werden optimale Strategien mit Hilfe von „trial und error“ und Ideen der dynamischen Optimierung gefunden. Das funktioniert sehr gut bei Spielen (Go), bei denen die Spielregeln bekannt und stabil sind. Das

ist auch bei der Portfoliokonstruktion denkbar und zwar immer dort wo das Portfolioproblem noch nicht gelöst oder nur sehr schwer theoretisch zu lösen ist. So kann „reinforcement learning“ (mit Hilfe simulierter Daten, d.h. unter den Prozessannahmen eines geschätzten Modells für den datengenerierenden Prozess) für Probleme der dynamischen Optimierung angewandt werden. Das optimale Rebalancing eines Portfolios mit Signalen unterschiedlichen Zeithorizonts und Transaktionskosten ist ein schwer zu lösendes Problem. Hier kann „reinforcement learning“ helfen.

Schlussfolgerungen

Das Schwierigste im Umgang mit maschinellem Lernen ist der Umgang mit seinen Advokaten. Die aggressiv vorgetragenen Behauptungen über neue überlegene Methoden zu verfügen, die mittelfristig die (Wirtschafts-) Wissenschaft revolutionieren und andere Methoden verdrängen werden, sind aus Sicht des Autors nicht gerechtfertigt. Realistischerweise gilt vielmehr: Wer sich im Portfoliomanagement gegen ML sperrt, wird den Wettkampf um bessere Performance verlieren. Das Gleiche gilt aber auch für den, der naiv sein Heil in ML sucht.

In der Wissenschaft wird es keine Rolle spielen, ob sich jemand als Datenwissenschaftler oder Ökonometriker fühlt. Es gibt nur gute oder schlechte Wissenschaft. Gegeben der wissenschaftstheoretischen Nachteile von ML wird sich ML in den Methodenkanon der Ökonometrie einordnen. Die Methoden sind für den quantitativen Wirtschaftswissenschaftler leicht zu erlernen und einfach zu implementieren. Es gibt umfangreiche und schnell wachsende Bibliotheken in Python und R. Mit dem vereinfachten Zugang steigt allerdings auch die Gefahr unreflektierter Anwendungen. Der Aufsatz ist daher keine Methodenkritik am ML, sondern die Warnung ML im Asset Management nicht in die Hände eines „Datenwissenschaftlers“ zu geben.

Die Voraussetzung für erfolgreiche, durch rein maschinelles Lernen erstellte Prognosen im Asset-Management sind

- große Datenmengen (existierend oder experimentell erzeugbar)
- stationärer Zusammenhang
- wenig existente Theorie oder ökonomischer „prior“
- keine Endogenität (latente Variablen, strategisches Verhalten)
- Fokus liegt ausschließlich auf Prognose / Klassifikation

Wo diese Voraussetzungen nicht vorliegen (beispielsweise bei vielen Datensätzen für Finanzmärkte), wird es deutlich schwieriger werden ML erfolgreich anzuwenden. Wie schwierig, ist eine empirische Frage, die Wissenschaft und Praxis gerade zu beantworten suchen. Die geringe Anzahl an erfolgreichen ML Fonds legt aber nahe, dass die Anwendung von ML zur Prognose von Wertpapierrenditen, keinesfalls einfach ist. Selbst für den erfolgreichsten existierenden ML Fonds (Renaissance) war es ein (Jahrzehnte) langer Weg, wie es der Lebensgeschichte von Jim Simons zu entnehmen ist.¹⁰

ML im Asset Management wird den Weg jeder neuen Idee gehen. Nach einer Welle großer Begeisterung wird Ernüchterung eintreten. Überoptimistische Behauptungen müssen zurückgenommen werden und man wird sich auf den Kern zurückbesinnen: Wo machen die Methoden Sinn und wo nicht. Portfolio Management ist sicher kein einfacher „use case“ für maschinelles Lernen.

¹⁰Zuckerman (2019) beschreibt in der Biographie von Jim Simons wie viele Irrwege dieser im Verlauf seiner Suche nach Mustern in Finanzmärkten gegangen ist und wie oft er kurz vor dem Scheitern war.

Appendix: Drei ? für ML - Verkauf von Eiskrem am Rocky Beach

Dieser Anhang schließt an unsere Beschreibung des Endogenitätsproblems für die Bestimmung der optimalen Preispolitik aus dem Haupttext an. Optimale Preispolitik benötigt eine verlässliche Schätzung, wie viel Nachfrage sich am Markt, für einen gewählten Preis, einstellt. Verzerrte Schätzungen für die Preis-Absatzfunktion machen die wichtigste Unternehmensentscheidung zum Glücksspiel.

Justus, Bob und Peter erben Giovanni's Eisdielen. Alles was Giovanni ihnen hinterlassen hat, ist eine Aufstellung der täglich verkauften Menge (Anzahl der Kugeln), des Preis pro Kugel und der Kosten pro Kugel Eis. Generiert wurden die Daten aus folgendem Standardmodell der Mikroökonomie. Gegeben sei eine Preis-Absatzfunktion der Form $q_t = \tilde{\alpha}_t - \beta p_t$, mit $\tilde{\alpha}_t = \bar{\alpha} + \tilde{\varepsilon}_t$. Der Zufallsterm ε_t beschreibt einen stochastischen Nachfrageschock, der bei der Preisentscheidung am Morgen des Tages t dem Entscheidungsträger bekannt ist. Wir können uns den Schock (Parallelverschiebung der Preisabsatzfunktion nach links unten oder rechts oben) als Temperaturvorsage aus dem Wetterbericht vorstellen. Bei schönem Wetter ist die Nachfrage nach Eis höher als bei schlechtem Wetter. Giovanni wusste dies und hat den gewinnmaximalen Preis p_t^* entsprechend seines Gewinnmaximierungsproblems gesetzt, wobei \tilde{c}_t die Kosten einer Kugel Eiskrem bezeichnet.

$$p_t^* = \underset{p_t}{\operatorname{argmax}} [p_t (\bar{\alpha} + \tilde{\varepsilon}_t - \beta p_t) - (\bar{\alpha} + \tilde{\varepsilon}_t - \beta p_t) \tilde{c}_t] \quad (2)$$

Die endgültige Nachfrage nach Eiskrem ergibt sich aus $q_t = \tilde{\alpha}_t - \beta p_t^* + \nu_t$, wobei ν_t idiosynkratische Nachfrageschwankungen darstellt, die Giovanni bei der Preisfestsetzung nicht beobachten konnte. Das Modell wurde mit folgenden Parametern kalibriert: $\bar{\alpha} = 500$, $b = 200$, $\varepsilon_t \sim N(0, 20)$, $\nu_t \sim N(0, 10)$, $c_t \sim \operatorname{unif}(0.5, 0.6)$. Giovanni's Daten für Justus, Peter und Bob finden sich in Abbildung 1. Man sieht einen klaren positiven Zusammenhang zwischen Preis und Menge. Je höher der Preis, desto höher die verkaufte Menge an Eiskrem. Für einen reinen „Datenwissenschaftler“ ist es leicht den falschen Schluss zu ziehen. Man könnte fabulieren das eine Hochpreispolitik die besondere Qualität von Giovanni's Eiskrem signalisiert. Für den Ökonomen ist ein anderer Zusammenhang wahrscheinlicher. Würden die Daten einem randomisierten Experiment (ein Preis für Eiskrem wird zufällig gewählt und die realisierte Nachfrage beobachtet) entstammen wäre eine Preis-Absatzfunktion mit negativer Steigung zu erkennen. Die Daten entsprechen aber keinem Experiment, sondern werden endogen bestimmt, d.h. der Preis wird gewinnmaximal gewählt. Wenn das Wetter schön (schlecht) und die Nachfrage nach Eis hoch (niedrig) ist, steigt der gewinnmaximale Preis. Über alle Datenpunkte sieht es so aus, als würde die Nachfrage mit dem Preis steigen. Der Zusammenhang ist direkt testbar indem die entsprechenden Variablen (Temperatur) in das Modell einbezogen werden. Es ergeben sich dann eine Vielzahl von Preis-Absatzfunktionen mit negativer Steigung. Diese sind in Abbildung 2 dargestellt. Die optimale Preispolitik ergibt sich aus

$$p_t^* = \frac{\bar{\alpha} + \varepsilon_t}{2\beta} + c_t \quad (3)$$

Jetzt kann man einwenden, dass ein ML Algorithmus, ergänzt um Temperaturdaten, auch in der Lage gewesen wäre, die Preis-Absatzfunktion zu identifizieren. Vielleicht. Der Ökonom kennt und versteht das Problem und kann alternative Modelle testen. Für einen Anwender ohne ökonomische Kenntnisse, ist es nicht sicher, ob er das statistische Problem löst. Es ist nicht unwahrscheinlich, dass er mit der Schätzung zufrieden ist (vor allem dann, wenn das Ergebnis plausibel ist). Ein Testinstrumentarium existiert für den Datenwissenschaftler nicht.

In jedem Fall wird der Datenwissenschaftler ohne Kenntnis der Mikroökonomie, die optimale Preispolitik nicht

finden, wie das nachfolgende Beispiel zeigt. Ich kalibriere auf den simulierten Daten einen sogenannten „regression tree“, einem Lieblingswerkzeug im maschinellen Lernen. Dabei werden Giovannis Preisaufzeichnungen mit den Variablen Temperatur (beobachteter Schock) und Kosten erklärt. Dies würde dann die optimale Preispolitik ergeben. Die Ergebnisse finden sich in Abbildung 3. Jeder Endknoten beinhaltet die durch Temperatur und Kosten erklärten Preise (Anova Zielfunktion, Standardeinstellung von `rpart` in `R`). Der beobachtete Nachfrageschock („obs.shock“) bestimmt maßgeblich den Preis. Je höher der Nachfrageschock, desto höher der Preis. Der „regression tree“ konnte diesen Zusammenhang ermitteln, obgleich er für lineare Regressionsprobleme eigentlich wenig geeignet ist. Man mag den Zusammenhang als einfache Preisregel in Form eines Entscheidungsbaumes stehen lassen und ihn selbst einer Aushilfskraft zur Preisfestsetzung überlassen. Dennoch wird der Nachteil von ML offensichtlich. Die Preisregel ist plump und die strukturellen Parameter bleiben unerkant. Fragestellungen nach dem Einfluss einer veränderten autonomen Nachfrage (α) oder der Preiselastizität (β) sind nicht zu beantworten.

Abbildung 1: Giovannis Daten

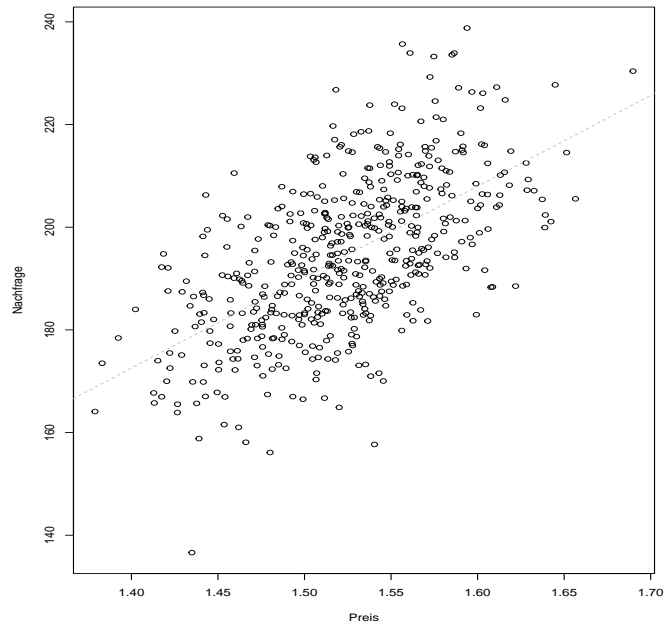


Abbildung 2: Versteckte Preis-Absatz Funktionen

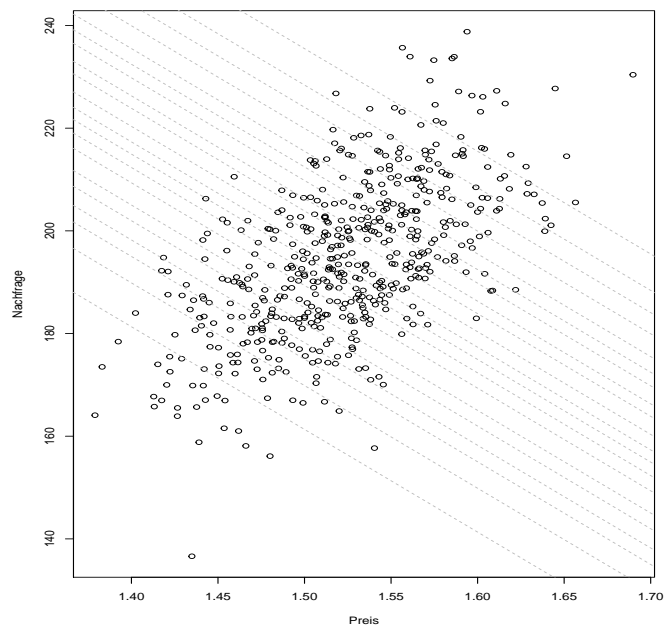
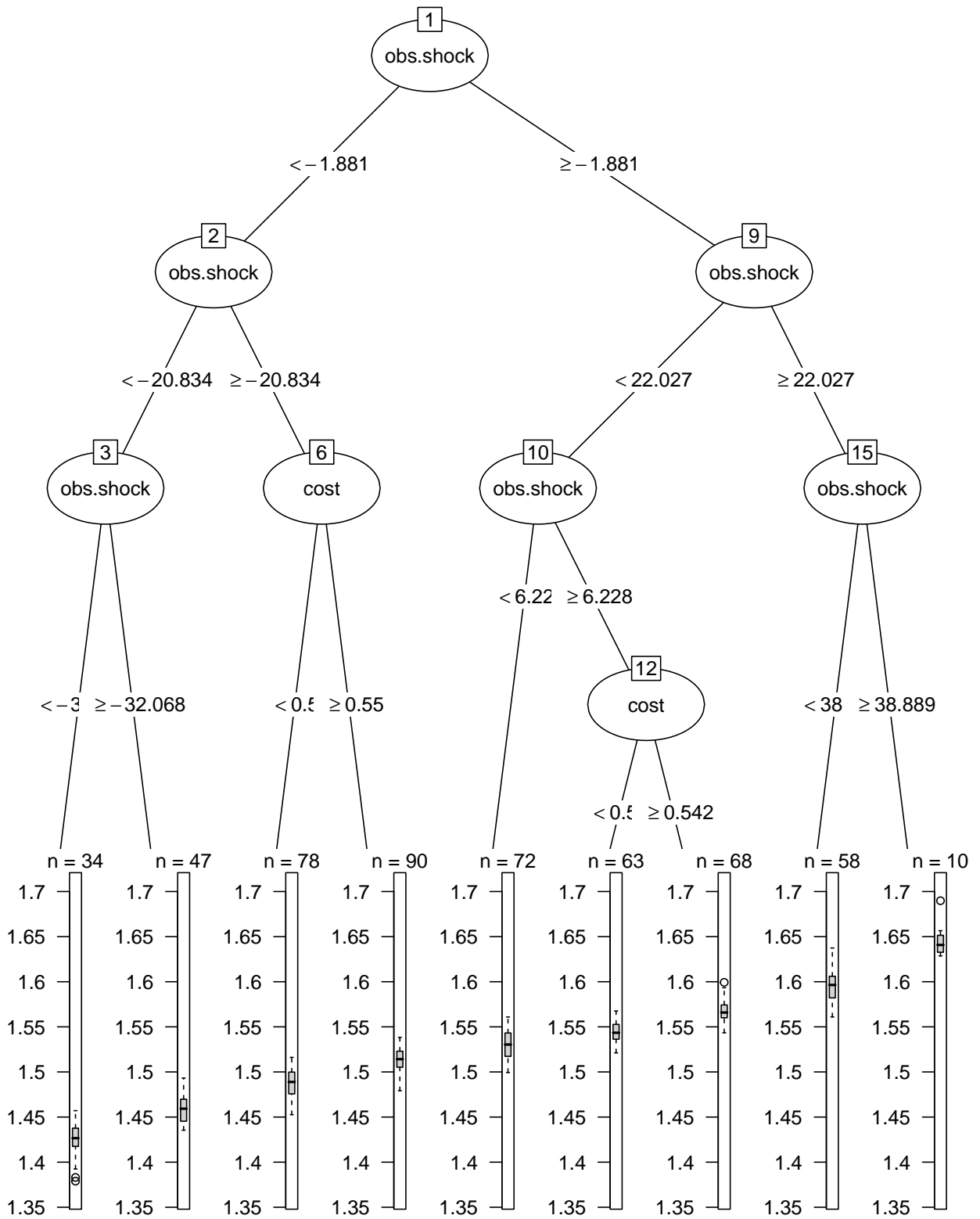


Abbildung 3: Optimal Preispolitik mit Hilfe eines „regression tree“



Literatur

- [1] Sendhil M. and J. Spiess (2017), Machine Learning. An Applied Econometric Approach, *Journal of Economic Perspectives*, v31(2), 87-1.
- [2] Breiman L. (2001), Statistical Modelling: The Two Cultures, *Statistical Science*, v16(3), 199-231.
- [3] Varian H.R. (2014), Big Data: New Tricks For Econometrics, *Journal of Economic Perspectives*, v28(2), 3-28.
- [4] SAP, Was ist maschinelles Lernen? <https://www.sap.com/germany/products/leonardo/machine-learning/what-is-machine-learning.html>
- [5] Lo A. (2017), *Adaptive Markets*, Princeton University Press
- [6] Zuckerman G. (2019), *The Man Who Solved The Market*, Penguin
- [7] Kelly, Bryan T, Gu, Shihao and Xiu, Dacheng, Empirical Asset Pricing via Machine Learning (2019). Chicago Booth Research Paper No. 18-04; 31st Australasian Finance and Banking Conference 2018; Yale ICF Working Paper No. 2018-09. Available at SSRN: <https://ssrn.com/abstract=3159577> or <http://dx.doi.org/10.2139/ssrn.3159577>
- [8] Guida T. (2019), *Big Data and Machine Learning in Quantitative Investment*, Wiley
- [9] Lucas R.E. (1976), *Econometric Policy Evaluation: A Critique*, Carnegie-Rochester Conference Series on Public Policy, volume 1, 19-46
- [10] Spiegelhalter D. (2019), *The Art of Statistics: Learning from Data*, Pelican
- [11] Fischer T. and C. Krauss (2017), Deep Learning with Long-Short Memory Networks For Financial Markets Prediction, *European Journal of Operational Research*
- [12] Pearl J and D. MacKenzie (2018), *The Book of Why*, Basic Books
- [13] Pearl J., Glymour M. und N. Jewell (2016), *Causal Inference in Statistics*, Wiley