

Statische und dynamische Evaluation von Prognosen

Weber, René

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Weber, R. (1998). Statische und dynamische Evaluation von Prognosen. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 43, 111-123. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-199822>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Statische und dynamische Evaluation von Prognosen

von René Weber¹

Zusammenfassung

Es werden die grundlegenden Möglichkeiten zur Quantifizierung der Prognosegüte dargestellt. Für die Evaluation von horizontalen Prognoseverteilungen wird ein „Gütemaße-Mix“ empfohlen, der unterschiedliche Qualitäten von Prognosen abbilden kann und sich dadurch für differenzierte Prognosevergleiche in der Praxis eignet. Darüber hinaus können Prognosevergleiche informativer gestaltet werden, wenn Prognosen dynamisch evaluiert werden. Eine dynamische Evaluation erhält man, indem man den Prognosehorizont sukzessiv um eine Periode ausdehnt und nach jeder neu hinzukommenden Periode die Prognosegüte neu berechnet. Dadurch ist man in der Lage zu beurteilen, in welchen Zeiträumen ein Prognoseverfahren bessere Prognoseergebnisse erzielt als ein anderes. Anhand von Beispielen wird gezeigt, daß sich die Einschätzung der Güte einer Prognose auf Basis einer statischen und dynamischen Betrachtung unterscheiden kann. Für die statische und dynamische Berechnung der Gütemaße wird ein SAS/IML-Programm mitgeliefert.

Abstract

The fundamental methods to quantify the accuracy of forecasts are presented. A "mix of measures" for the evaluation of horizontal distributions of forecasts is recommended, which can indicate different characteristics of forecasts, and is therefore suitable for detailed comparisons of forecasts in practical applications. Evaluating forecasts dynamically makes comparisons more informative. This can be done by gradually expanding the forecast horizon by one period at a time, which yields a dynamic measure of forecast accuracy. It is thereby possible to assess in what periods one method of forecasting performs better than another. Examples are given to demonstrate that static and dynamic evaluations of forecast accuracy do differ. A SAS/IML-program is supplied that calculates the recommended "mix of measures" statically and dynamically.

¹ **René Weber** ist Wissenschaftlicher Mitarbeiter am Institut für Psychologie (Lehrstuhl Professor **Jürgen Bortz**). Fachgebiet Methodenlehre und Statistik. TU Berlin. Franklinstrasse 28/29 - Postfach FR3-8, 10587 Berlin. E-mail: R.Weber@TU-Berlin.De

Quantifizierung der Prognosegüte

Es gibt zwei Möglichkeiten, die Prognosegüte einer Methode zu quantifizieren - entweder auf Basis der vertikalen oder auf Basis der horizontalen Verteilung der Prognosewerte.

Bei der Betrachtung der vertikalen Prognoseverteilung leitet man in analytischer Form Prognoseintervalle (Konfidenzintervalle) her und schätzt diese im allgemeinen anhand der Kovarianzmatrix der Modellresiduen. Je kleiner das Prognoseintervall ist, desto sicherer ist die Prognose bzw. desto besser ist die Prognosegüte. Ist die analytische Herleitung komplex, können Prognoseintervalle mit Hilfe von „Resampling-Methoden“ approximiert werden. Die Güte von Ex-ante-Prognosen kann nur durch die Betrachtung der vertikalen Prognoseverteilung quantifiziert werden.

Die zweite Möglichkeit, die Prognosegüte einer Methode zu quantifizieren, ergibt sich aus der Berechnung eines Kennwertes der horizontalen Prognoseverteilung. Hierbei werden alle resultierenden Prognosewerte mit den tatsächlich beobachteten Werten (die nicht zur Modellidentifikation verwendet wurden!) verglichen und die Unterschiedlichkeit durch eine Maßzahl ausgedrückt. Da hierfür die tatsächlich beobachteten Werte bekannt sein müssen, eignet sich diese Art der Evaluation ausschließlich für eine Ex-post-Prognose. Wegen der einfacheren Berechnung und Interpretation im praktischen Einsatz konzentrieren sich die folgenden Erläuterungen auf die horizontale Prognoseverteilung, d.h. auf die Berechnung von Gütemaßen.

Faktoren der Prognosegüte

Betrachtet man die horizontale Prognoseverteilung, dann beeinflussen hauptsächlich die Faktoren Prognosehorizont, Beschaffenheit der Daten und das gewählte Prognosegütemaß die angegebene Genauigkeit von Prognosen.

Langfristige Prognosen sind in der Regel ungenauer als kurzfristige Prognosen, da die Zeitinvarianzannahme² bei einer langfristigen Prognose weniger wahrscheinlich ist als bei einer kurzfristigen Prognose. Unterscheiden sich die zu prognostizierenden Daten z. B. hinsichtlich der Varianz, der Saisonalität oder der Komplexität der Einflußfaktoren, so ist es plausibel, daß die Genauigkeit einer Prognose davon abhängt, wie gut ein Prognosemodell dazu in der Lage ist, die spezifischen Eigenschaften einer Datenbasis zu erfassen.

Ein Irrtum wäre es zu glauben, daß unterschiedliche Prognosegütemaße zu einer übereinstimmenden Beurteilung der verschiedenen Prognosetechniken führen. Fehlprognosen

2 Definition der Zeitinvarianzannahme: Annahme, daß sich die Bedingungen, unter denen ein Prognosemodell identifiziert wurde, im Prognosezeitraum nicht verändern.

können sich durch ein bestimmtes Ausmaß an systematischen und unsystematischen Abweichungs-, Niveau- und Wendepunktfehlern unterscheiden. Die meisten gebräuchlichen und einfach interpretierbaren Gütemaße bewerten jedoch in erster Linie Abweichungsfehler und nur indirekt Niveau- und Wendepunktfehler (diese induzieren Abweichungsfehler). Eine Betrachtung mehrerer Gütemaße, die möglichst unterschiedliche Merkmale aufweisen, um die verschiedenen Arten von Fehlprognosen sensitiv erfassen zu können, würde hier eine differenziertere Betrachtung erlauben. Werden vielfältige Gütemaße verwendet, hat der Anwender die Möglichkeit, durch eine subjektive Gewichtung der Gütemaße diejenige Prognose als die beste zu identifizieren, die seinen eigenen Präferenzen am besten entspricht.

Gütemaße

In der Literatur gibt es eine nahezu unüberschaubare Menge an Gütemaßen [vgl. z. B. *Khosrawi-Rad, M.D.* 1991]. Um so überraschender ist es, daß bis heute nur wenige Studien existieren, die intensiver die Merkmale und Charakteristiken von Gütemaßen analysieren [wie z.B. *Cicarelli, J.* 1982; *Wright, D.L.* 1986 und besonders *Dammers, E.* 1993].

Eine optimale Auswahl an Gütemaßen sollte sich in erster Linie an den Eigenschaften Sensitivität, Art der Verlustfunktion, Fehlergewichtung, Homogenität, Monotonie, Interpretierbarkeit und Bekanntheit der Maße orientieren. Im einzelnen bedeutet das: Es sollten Maße enthalten sein, die jeweils speziell auf Abweichungs-, Niveau- und Wendepunktfehler sensitiv reagieren. Es sollten symmetrische (Unter- und Überschätzungen werden gleich bewertet) und asymmetrische Verlustfunktionen berücksichtigt werden. Fehlprognosen sollten linear, quadratisch (größere Abweichungen werden stärker gewichtet) und abhängig vom Niveau der beobachteten Datenwerte in die Berechnung unterschiedlicher Gütemaße eingehen. Es sollten homogene und monotone Gütemaße [vgl. *Dammers, E.* 1993] vertreten sein, damit Prognosen mit unterschiedlichen Einheiten und auf einer Kardinalskala miteinander verglichen werden können. Die einfache Interpretierbarkeit erleichtert die Anwendung. Bekannte Gütemaße ermöglichen darüber hinaus Prognosevergleiche über verschiedene Studien hinweg.

Sämtliche Eigenschaften können nur in einem „Gütemaße-Mix“ berücksichtigt werden. In Tabelle 1 werden einige wenige Gütemaße wiedergegeben, die für einen Prognosevergleich zu empfehlen sind (P_t : Prognosen zum Zeitpunkt t ; O_t : Beobachtete Werte zum Zeitpunkt t ; s : Streuung; M : Prognosehorizont; A : Häufigkeit „Wendepunkt (WP) prognostiziert und vorhanden; B : Häufigkeit „WP prognostiziert und nicht vorhanden“; C : Häufigkeit „WP nicht vorhanden und prognostiziert“; D : Häufigkeit „WP nicht vorhanden und nicht prognostiziert“).

Der mittlere Fehler (ME): Der ME zeichnet sich in der Auswahl durch Monotonie aus. Damit können Prognosevergleiche auf einer Kardinalskala erfolgen. Halbiert sich beispielsweise bei zwei zu vergleichenden Prognosen der ME, so kann gesagt werden, daß eine der Prognosen doppelt so gut ist wie die andere. Der ME bringt zum Ausdruck, in welchem Ausmaß die vorhergesagten Werte von den beobachteten Werten abweichen, wobei sich positive und negative Abweichungen gegenseitig ausgleichen können (asymmetrische Verlustfunktion).

Tabelle 1: Gütemaße des „Gütemaße-Mix“.

Gütemaß	Formel
Mittlerer Fehler	$ME = \frac{1}{M} \sum (P_t - O_t)$
Mittlerer absoluter prozentualer Fehler	$MAPE = \frac{1}{M} \sum \left \frac{(P_t - O_t)}{O_t} \right \cdot 100\%$
Wurzel des mittleren quadr. Fehlers	$RMSE = \sqrt{\frac{1}{M} \sum (P_t - O_t)^2} = \sqrt{MSE}$
Bias-Anteil des MSE (MSE=RMSE ²)	$BIAS = \frac{(\bar{O}_t - \bar{P}_t)^2}{MSE}$; für $MSE \neq 0$
Varianz-Anteil des MSE	$VAR = \frac{(s_o - s_p)^2}{MSE}$; für $MSE \neq 0$
Kovarianz-Anteil des MSE	$KOV = \frac{2 \cdot (1 - R) \cdot s_o \cdot s_p}{MSE}$; für $MSE \neq 0$
Lineare Korrelation	$R = \frac{1}{M} \frac{\sum (P_t - \bar{P}) \cdot (O_t - \bar{O})}{s_p \cdot s_o}$
Wurzel des mittleren proz. Fehlers	$RMSP = \sqrt{\frac{1}{M} \sum \frac{(P_t - O_t)^2}{O_t^2}} \cdot 100\%$
Ballsches Q ²	$Q^2 = 1 - \frac{\sum (O_t - P_t)^2}{\sum (O_t - \bar{O}_{(A)})^2}$
Überzufällige Trefferquote	$\ddot{U}TQ = \frac{1}{M} \cdot \left((A + D) - \frac{(A + B)^2 + (C + D)^2}{M} \right)$

Der mittlere absolute prozentuale Fehler (MAPE): Der MAPE wird in erster Linie wegen seiner Fehlergewichtung und seiner anschaulichen Interpretation ausgewählt. Prognosefehler bewertet der MAPE als einziges Maß in der Auswahl linear und zugleich abhängig

vom Niveau der beobachteten Daten. Der MAPE ermittelt, um wieviel Prozent die Prognosen im Mittel von den beobachteten Werten abweichen. Wie alle folgenden Maße besitzt der MAPE eine symmetrische Verlustfunktion.

Die Wurzel aus dem mittleren quadratischen Fehler (RMSE): Der RMSE wird im wesentlichen wegen seiner quadratischen Fehlergewichtung, seiner einfachen Interpretation und seinem hohen Bekanntheitsgrad ausgewählt. Der RMSE liefert die Information, inwieweit die vorhergesagten Werte im Durchschnitt den beobachteten Werten entsprechen, ohne daß sich dabei positive und negative Abweichungen gegenseitig ausgleichen können. Dabei werden größere Abweichungen stärker gewichtet als kleinere Abweichungen.

Bias-, Varianz- und Kovarianz-Anteil (BIAS, VAR, KOV): Diese drei Gütemaße zerlegen den RMSE (bzw. MSE, siehe Tabelle 1) in systematische und unsystematische Fehleranteile. Darüber hinaus besitzt der Bias-Anteil als einziges Maß der Auswahl eine hohe und spezielle Sensitivität gegenüber Niveau-Fehlprognosen. Weicht der Bias-Anteil von Null ab, so ist dies als eine systematische Unter- bzw. Überschätzung der tatsächlich beobachteten Werte durch die Prognosen zu interpretieren (Niveau-Fehlprognose). Weicht der Varianz-Anteil von Null ab, so können die Schwankungen der Beobachtungswerte um ihren mittleren Wert zwar durch die Prognosen erfaßt werden, das Ausmaß dieser Erfassung ist jedoch zu stark bzw. zu schwach (systematische Abweichungs-Fehlprognosen). Prognosen, die weder systematische Niveaufehler noch Abweichungsfehler aufweisen, können sich von den beobachteten Werten nur noch unsystematisch unterscheiden. Diese unsystematischen Abweichungsfehler erfaßt der Kovarianz-Anteil. Bias-, Varianz- und Kovarianz-Anteil sind beschränkt (normiert) zwischen den Werten Null und Eins. Je kleiner der Bias- sowie Varianz-Anteil und je näher der Kovarianz-Anteil bei Eins liegt, desto besser ist die Prognose.

Die Korrelation (R): Die exklusive und hohe Sensitivität der Korrelation bezüglich quantitativer Wendepunkt-Fehlprognosen ist der Hauptgrund für die Auswahl. Von Vorteil bei der Interpretation ist die Beschränktheit und der hohe Bekanntheitsgrad. Je größer R ist ($-1 \leq R \leq 1$), desto mehr Wendepunkte werden in korrekter Richtung und mit korrektem Betrag prognostiziert.

Die Wurzel des mittleren quadr. prozentualen Fehlers (RMSP): Der RMSP wird ausgewählt, da er Prognosefehler sowohl quadratisch als auch in Abhängigkeit vom Ausgangsniveau der Daten gewichtet. Die inhaltliche Interpretation ist deckungsgleich mit der des MAPE. Es wird ermittelt, um wieviel Prozent die Prognosen im Mittel von den beobachteten Werten abweichen.

Das Ballsche (Q^2): Das Ballsche Q^2 liefert als einziges quantitatives Gütemaß der Auswahl einen direkten Vergleich mit einer Trivialprognose (arithmetischer Mittelwert der

beobachteten Werte). Je näher Q^2 bei einem Wert von Eins liegt, desto besser ist die Prognose im Vergleich zur gewählten Trivialprognose. Negative Q^2 -Werte indizieren eine überlegene Trivialprognose.

Die überzufällige Trefferquote (ÜTQ): ÜTQ erfaßt exklusiv qualitative Wendepunkt-Fehlprognosen bzw. ausschließlich die Richtung (das Vorzeichen) der Fehlprognosen. Dabei wird ÜTQ um zufällig richtige Prognosen bereinigt. Je größer ÜTQ, desto besser ist die Prognose. Bei gleichwahrscheinlichen Kategorien kann ÜTQ nicht größer als 0.5 bzw. 50% werden. Eine negative überzufällige Trefferquote würde bedeuten, daß eine Zufallsprognose (Trivialprognose) gemäß den Kategorien-Randwahrscheinlichkeiten bessere Ergebnisse liefert als eine zu evaluierende Prognosemethode.

Dynamische Evaluation von Prognosen

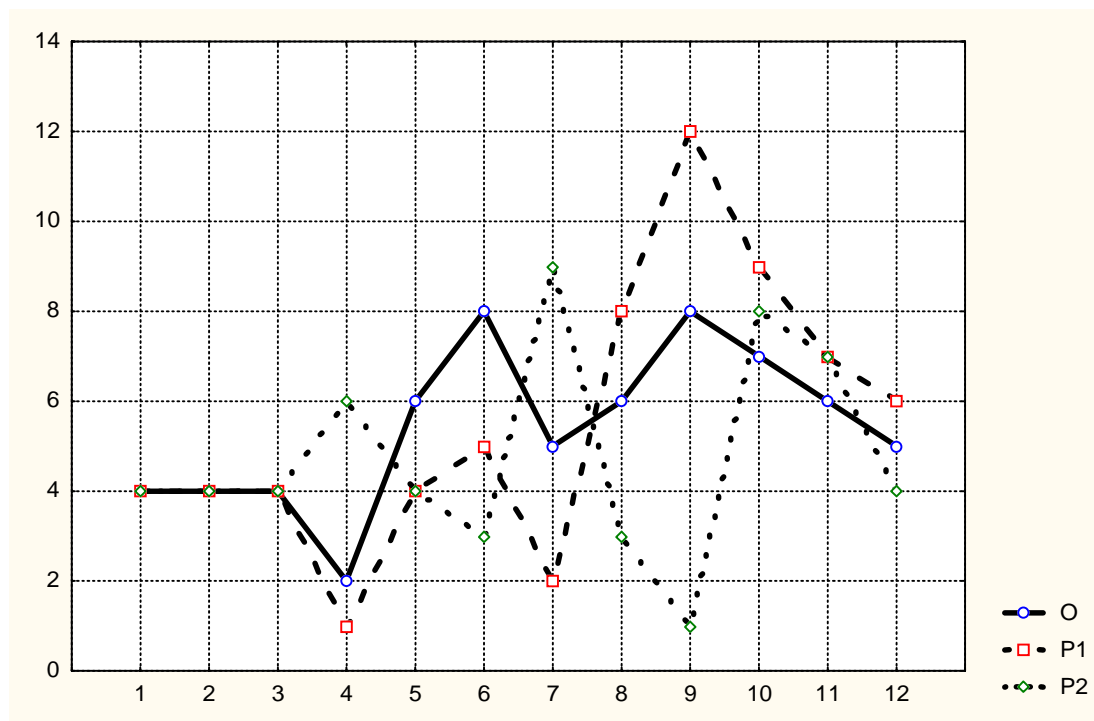
Anhand der empfohlenen Gütemaße ist es möglich, die horizontale Prognoseverteilung in statischer Form zu evaluieren, d.h. sämtliche Prognosewerte werden in einem Schritt bei der Berechnung der Gütemaße einbezogen. Prognosevergleiche in der Praxis werden üblicherweise auf diese Weise vorgenommen. Informativere Vergleiche sind jedoch möglich, wenn man die Prognosegüte in dynamischer Form darstellt. Diese erhält man, wenn man den Prognosehorizont sukzessiv um eine Periode ausdehnt und nach jedem neu hinzukommenden Datum die Prognosegüte neu berechnet. Eine dynamische Prognosegüte kann demnach nur berechnet werden, wenn den zu prognostizierenden Daten eine ordnende Struktur (z. B. die Zeit) zugrundegelegt werden kann. Bei einer größeren Fallzahl kann eine Neuberechnung auch nach einer jeweils festzulegenden Anzahl an Perioden erfolgen, so daß die wesentliche Information einer dynamischen Evaluation komprimierter dargestellt werden kann.

Durch die Dynamisierung kann die Prognosegüte in Abhängigkeit vom Prognosezeitraum betrachtet werden, d.h. man ist in der Lage zu beurteilen, in welchen Zeiträumen ein Prognoseverfahren bessere Prognoseergebnisse liefert als ein anderes. Ein derartiger Vergleich ist insbesondere für die Formulierung eines optimalen Prognosehorizontes einer Methode bei gegebenen Daten geeignet. Zur weiteren Analyse der dynamischen Prognosegüte kann man erneut statistische Verfahren (z. B. zeitreihenanalytische Methoden) heranziehen. Werden beispielsweise Gütemaße, die Abweichungs-Fehlprognosen sensitiv erfassen, dynamisch berechnet, kann darüber hinaus eruiert werden, ob eine schlechte Prognosegüte nur durch singuläre Ausreißer (Strukturbrüche in der dynamischen Darstellung), durch andauernde Fehlprognosen (Trends in der dynamischen Darstellung) oder durch Fehlprognosen nur zu bestimmten Zeitpunkten des Prognosezeitraumes (Trends und Plateaus in der dynamischen Darstellung) bedingt sind. Ebenfalls ist denkbar, daß Prognosen, die statisch evaluiert werden, als gleich gut eingestuft werden, während ein dynamischer

Prognosevergleich deutliche Unterschiede offenbart. Anhand eines Daten-Beispiels und der dynamischen Berechnung des ME, RMSE sowie des BIAS-, VAR- und KOV-Anteils (siehe Tabelle 1) soll dies verdeutlicht werden.

Die Prognosen „P1“ und „P2“ sind bezüglich der beobachteten Werte „O“ miteinander zu vergleichen (siehe Abbildung 1). Aus Gründen der Übersichtlichkeit werden lediglich 12 Beobachtungswerte verwendet. Bei dieser geringen Menge an Beobachtungen könnten grundlegende Prognosequalitäten auch per Augenschein erkannt werden. Bei Prognosen mit einer größeren Menge an Daten ist dies i.d.R. nicht möglich.

Abbildung 1: Prognosebeispiele.



Berechnet man die Prognosegüte statisch, dann erhält man zur Evaluation das Ergebnis in Tabelle 2. Man würde interpretieren, daß P2 sich durch stärkere Abweichungs-Fehlprognosen auszeichnet. Hinsichtlich einer Niveau-Fehlprognose unterscheiden sich beide Prognosen nicht. Insgesamt sind systematische Fehler stärker bei P1 als bei P2 ausgeprägt (siehe Abschnitt „Gütemaße“).

Berechnet man die Prognosegüte dynamisch (siehe Abbildung 2 bis Abbildung 6), dann erhält man zur Evaluation ein differenzierteres Bild der Prognosegüte. Der wesentlich größere mittlere Fehler von Prognose P2 entsteht offenbar erst ab Periode 9 (bedingt durch eine extreme Fehlprognose). Anhand des dynamischen RMSE erkennt man, daß der Unterschied der zwei Prognosen hinsichtlich der Intensität von Abweichungs-Fehlprognosen im

gesamten Prognosezeitraum ungefähr gleich bleibt (Abbildung 3). P2 leistet sich größere Abweichungsfehler als P1.

Tabelle 2: Gütemaße des „Gütemaße-Mix“.

Gütemaß	Prognose P1	Prognose P2
ME	-0.08	0.67
RMSE	2.02	3.19
BIAS	0.00	0.04
VAR	0.39	0.02
KOV	0.61	0.94

Der deutlichste Unterschied zwischen den zwei Prognosen, der anhand der statischen Prognosegüte nicht erkennbar war, besteht allerdings hinsichtlich einer systematischen Niveau-Fehlprognose. Bis zur Periode 8 zeichnet P1 einen erheblich höheren BIAS-Anteil aus, der sich erst ab Periode 9 dem von P2 gleicht (Abbildung 4). Des weiteren ist P1, im Unterschied zu P2, insbesondere zu Beginn und am Ende des Prognosezeitraums durch systematische Abweichungsfehlprognosen gekennzeichnet (Abbildung 5 und Abbildung 6).

Anwendungsbeispiele

Bei jeder empirischen Studie, die mit Prognosen operiert, stellt sich die Frage nach der Güte der Ergebnisse. Dabei sind die Ansprüche an den Informationsgehalt einer durchzuführenden Evaluation unterschiedlich. So mag für einen Wirtschaftswissenschaftler, der den Aktienmarkt bezüglich kurzfristiger Kauf- und Verkaufssignale analysiert, entscheidend sein, ob Prognosen in der Lage sind, Wendepunkte korrekt vorherzusagen. Ob sich Prognosen auch durch einen geringen Abweichungs- und/oder Niveaufehler auszeichnen, ist von untergeordneter Bedeutung. Durch eine dynamische Betrachtung der Prognosegüte ist es darüber hinaus möglich, optimale Prognosezeiträume, z. B. die optimale Vorlaufzeit eines kurzfristigen Handels zu ermitteln.

Abbildung 2: Dynamische Prognosegüte des ME.

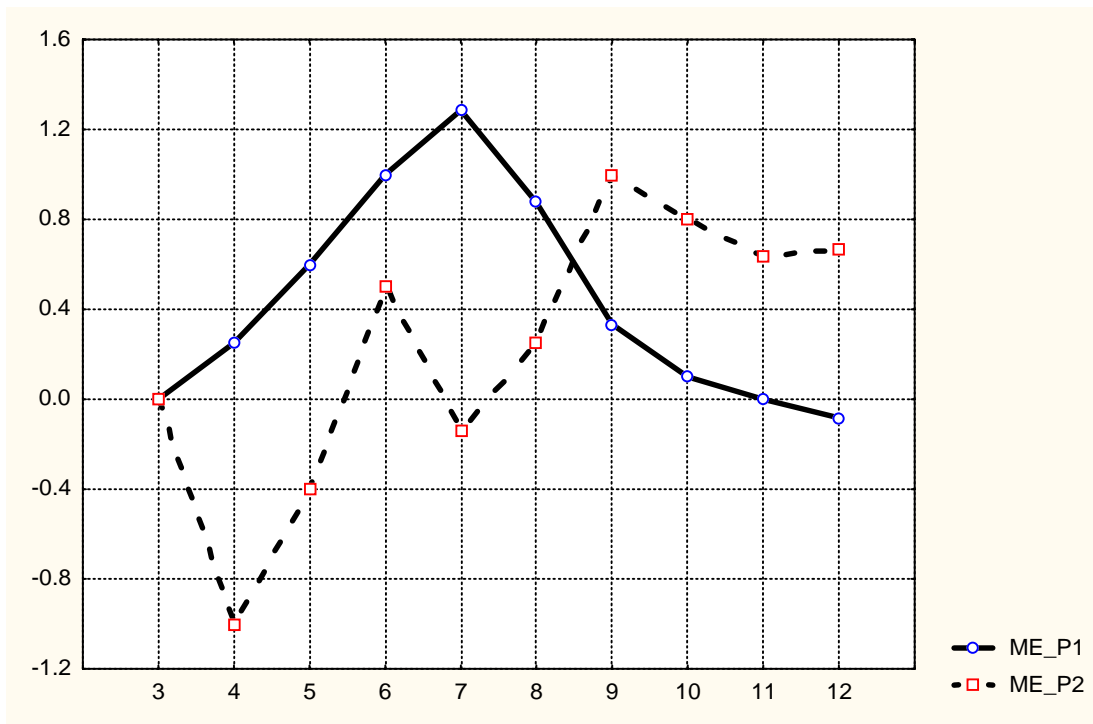


Abbildung 3: Dynamische Prognosegüte des RMSE.

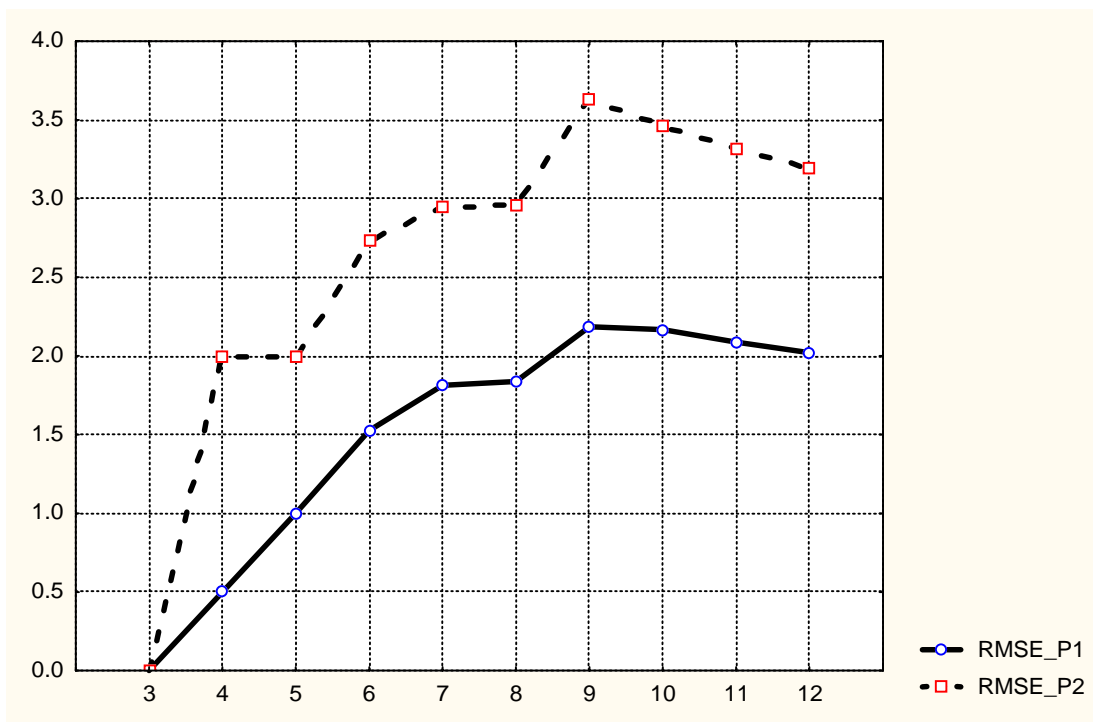


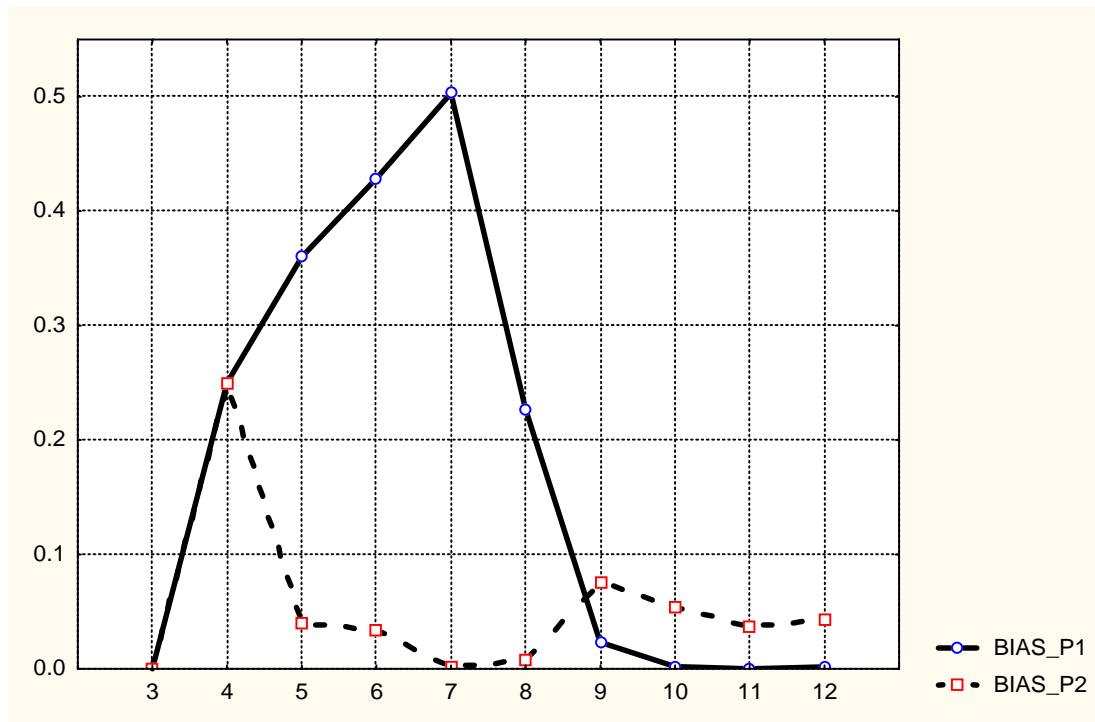
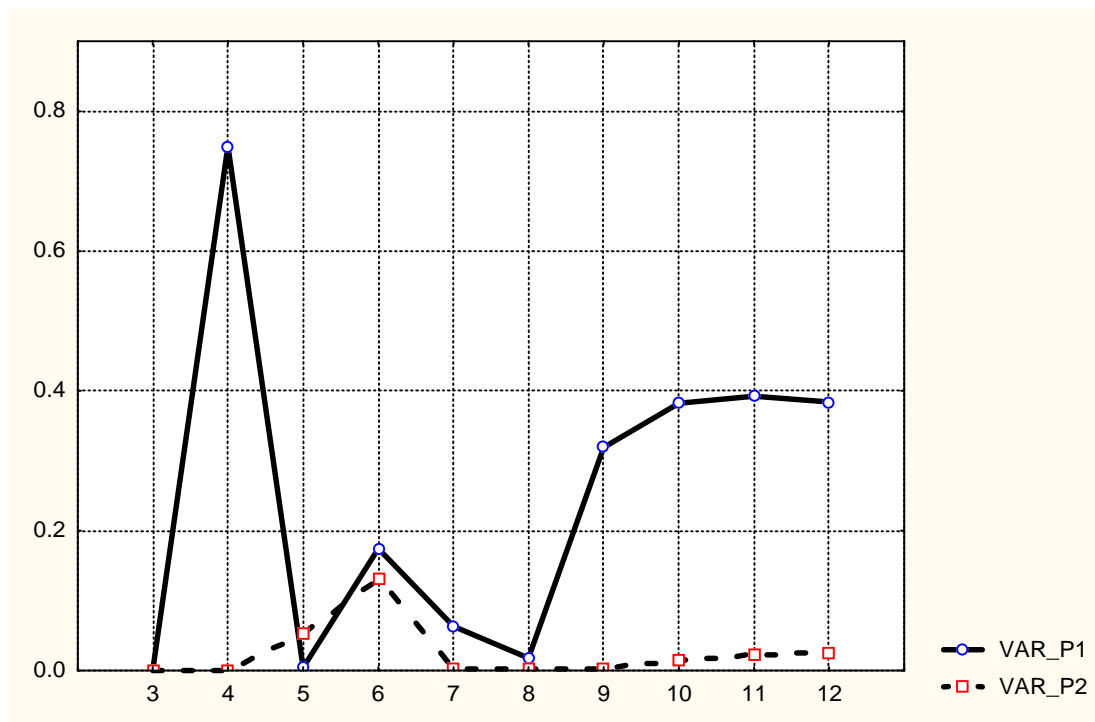
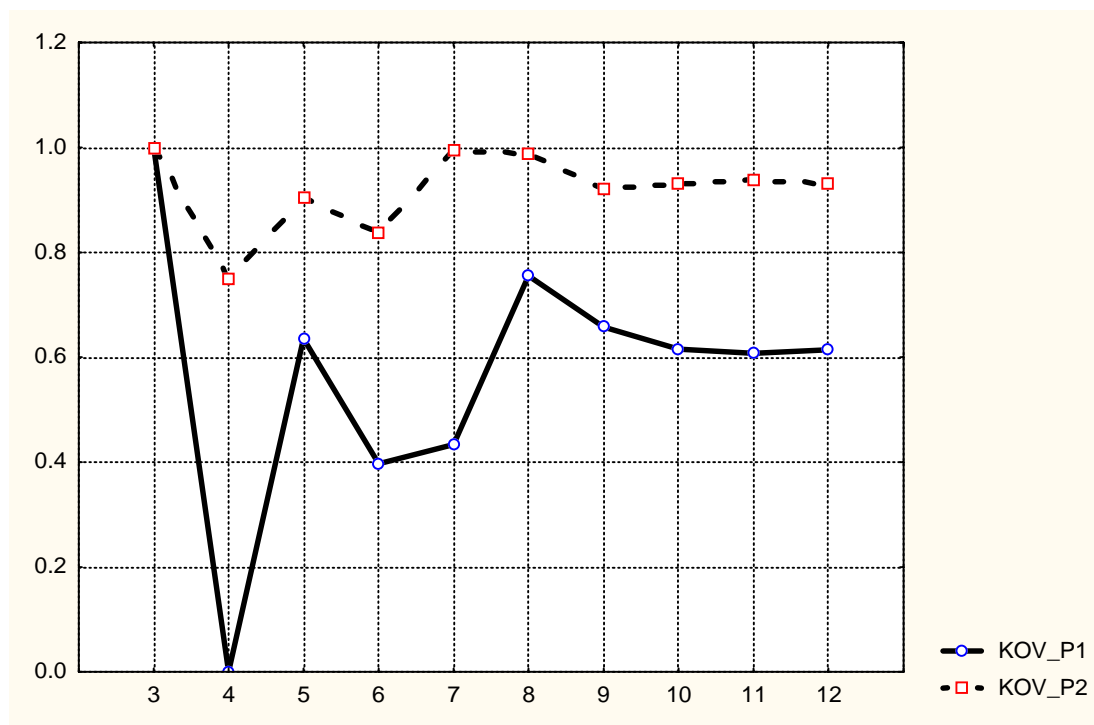
Abbildung 4: Dynamische Prognosegüte des BIAS-Anteils.**Abbildung 5:** Dynamische Prognosegüte des VAR-Anteils.

Abbildung 6: Dynamische Prognosegüte des KOV-Anteils.



Auf der anderen Seite könnte für einen empirischen Forscher (z. B. Soziologen, Psychologen, Kommunikationswissenschaftler), der den zukünftigen Verlauf der Fernsehnutzung analysieren und prognostizieren möchte, ein möglichst geringer Abweichungsfehler (eine möglichst gute Modellanpassung im Prognosezeitraum) im Fokus seiner Betrachtungen stehen. Zusätzlich wird hier eine Rolle spielen, ob die Prognose eher zu Unter- oder Überschätzungen der tatsächlichen Werte tendiert und eine Evaluation in der Lage ist, beide Arten der Fehlprognosen zu unterscheiden. Ein Forscher der Erziehungswissenschaften beispielsweise, der den Standpunkt vertritt, daß weniger fernsehen positiv zu beurteilen ist, wird vermutlich Überschätzungen weniger problematisch als Unterschätzungen beurteilen. Ein Forscher, der im Dienste der Programmanbieter arbeitet, und auf dessen Prognosen Preise für Werbezeiten berechnet werden, wird geneigt sein, seine Modelle so zu spezifizieren, daß unvermeidliche Fehlprognosen eher zu Unterschätzungen führen. Darüber hinaus liefert hier eine dynamische Betrachtung der Prognosegüte z. B. Anhaltspunkte dafür, für welchen Zeithorizont Preiszusagen vereinbart werden können bzw. ab welchem Zeitpunkt die Prognosegüte eines spezifizierten Modells Neuverhandlungen nahelegt.

Nicht zuletzt ist eine differenziertere statische und dynamische Evaluation von Prognosen für die Methodiker unter den empirischen Sozialforschern von Bedeutung. Ein Beispiel möge dies verdeutlichen: In den letzten Jahren ist zu beobachten, daß „moderne“ Modellierungs- und Prognosemethoden, die sich verschiedener Verfahren der künstlichen Intelligenz bedienen (z. B. neuronale Netze), zunehmend auch bei Analysen in der empirischen

Sozialforschung eingesetzt werden. Teilweise wird hierbei undifferenziert eine Überlegenheit bezüglich klassischer statistischer Verfahren behauptet, die heute, generell betrachtet, als Mythos einzuschätzen ist. In zahlreichen Analysen zeigt sich, daß der Einsatz derartiger Methoden nur bei bestimmten Datencharakteristiken zu empfehlen ist. Des weiteren offenbart eine differenzierte statische und dynamische Evaluation der Prognoseergebnisse (bzw. der Modellanpassung) in vielen Fällen, daß eine vorhandene Überlegenheit hinsichtlich einer Prognosequalität (siehe oben) durch eine Unterlegenheit in einer anderen Prognosequalität „erkauft“ wird. Eine dynamische Evaluation auf Basis eines „Gütemaße-Mix“ liefert hier dem Forscher ein Instrument, mit dessen Hilfe er entscheiden kann, ob sich für ihn die i.d.R. aufwendigere Anwendung neuer Modellierungs- und Prognosemethoden lohnt.

SAS - Programm

Tabelle 3 (siehe Anhang) liefert die dynamische Berechnung aller in Tabelle 1 angegebenen Gütemaße für die Daten-Vektoren „ o_t “ (observed) und „ p_t “ (predicted). Umfaßt die dynamische Prognosegüte den gesamten Prognosehorizont, so liefert das Programm zugleich die Prognosegüte in statischer Form (sämtliche Prognosewerte werden dann einbezogen). Die Berechnung der Gütemaße in Matrix-Schreibweise ermöglicht deren einfache Berechnung auch in einer anderen Software. Für andere Daten-Vektoren ist der Programmteil „Daten einlesen“ entsprechend zu modifizieren oder das Programm in Form eines SAS-Makros umzusetzen. Falls gewünscht, kann das Programm per e-mail (R.Weber@TU-Berlin.De) zugesandt werden.

Literatur

Khosrawi-Rad, M.D.: Probleme und Möglichkeiten bei der Definition, Klassifikation, Interpretation und Operationalisierung von Prognose, Prognosemodell, Prognosefehler und Prognosefehlermaße. Kovac-Verlag, Hamburg 1991.

Cicarelli, J.: New method of evaluating the accuracy of economic forecasts. In: Journal of Macroeconomics 4/1982. Seite 469-475.

Wright, D.J.: Evaluation of forecasting methods for decision support. In: International journal of forecasting Vol. 2/1986. Seite 139-152.

Dammers, E.: Measurement in the ex post evaluation of forecasts. In: Quality and Quantity 27/1993. Seite 31-45.

Anhang

Tabelle 3: SAS/IML-Programm zur dynamischen Berechnung der Gütemaße.

```

PROC IML;

* Daten einlesen;
* -----;
ot = {4,4,4,2,6,8,5,6,8,7,6,5}; * Beobachtete Werte (O);
pt = {4,4,4,1,4,5,2,8,12,9,7,6}; * Prognostizierte Werte (P1);

* Dynamische Berechnung der Masse;
* -----;
Guete = J(nrow(ot)-2,10+1,0);

DO i = 3 to nrow(ot);

o = ot[1:i]; p = pt[1:i]; m = nrow(o);
meano = o[:]; meanp = p[:];
sol = (o[##] - m*meano**2)/m;
spl = (p[##] - m*meanp**2)/m;

IF (sol < 0) THEN sol = 0;
IF (spl < 0) THEN spl = 0;

so = sqrt(sol);
sp = sqrt(spl);

IF (so = 0) THEN so = 0.0001;
IF (sp = 0) THEN sp = 0.0001;

res = o-p;
eins = J(m,1,1);
deltao = (o[2:m] - o[1:m-1])>=0;
deltap = (p[2:m] - p[1:m-1])>=0;
n = nrow(deltao);

ME = (eins`*res)/m;
MAPE = (eins`*abs(res/o))/m;
MSE = (res`*res)/m;
RMSE = sqrt(MSE);
R = (((p-meanp)`*(o-meano)/m)/(so*sp));

IF (MSE ^= 0) THEN BIAS = ((meano - meanp)**2) / MSE;
ELSE BIAS = 0;
IF (MSE ^= 0) THEN VAR = ((so - sp)**2) / MSE;
ELSE VAR = 0;
IF (MSE ^= 0) THEN KOV = (2 * (1 - R) * so * sp) / MSE;
ELSE KOV = 1;

RMSP = sqrt((res/o)`*(res/o)/m);
Q2 = 1 - (res`*res) / (m * so**2);
UETQ = ((deltao=deltap)[+] - (deltao[+]**2 + (n-deltao[+]**2)/n) / n;

Guete[i-2,] = i || ME || MAPE || RMSE || BIAS || VAR || KOV || R || RMSP || Q2 || UETQ;

END;

* Output produzieren;
* -----;
Namguete = {'Horizont' 'ME' 'MAPE' 'RMSE' 'BIAS' 'VAR' 'KOV' 'R' 'RMSP' 'Q2'
'UETQ'};
PRINT Guete[colname=namguete format=8.3];
CREATE GMASSE FROM guete [COLNAME=Namguete];
APPEND FROM guete;
QUIT;
RUN;

```