

Rationalität und Qualität von Wirtschaftsprognosen

Dissertation

*Zur Erlangung des Doktorgrads
der Wirtschaftswissenschaftlichen Fakultät
der Georg-August-Universität Göttingen*

*vorgelegt von
Dipl.-Kfm. (FH) Johannes Scheier
aus Freiburg (Breisgau)*

Braunschweig, den 27.09.2014

Erstgutachter:	Prof. Dr. Kilian Bizer
Zweitgutachter:	Prof. Dr. Markus Spiwoks
Weiteres Mitglied im Prüfungsausschuss:	Prof. Dr. Stefan Dierkes
Tag der mündlichen Prüfung:	28.04.2015

Abstract

Wirtschaftsprognosen sollen die Unsicherheit bezüglich der zukünftigen wirtschaftlichen Entwicklung mindern und Planungsprozesse von Regierungen und Unternehmen unterstützen. Empirische Studien bescheinigen ihnen jedoch in aller Regel ein unbefriedigendes Qualitätsniveau. Auf der Suche nach den Ursachen hat sich in Form der Rationalität eine zentrale Grundforderung an die Prognostiker herausgebildet. So müssten offensichtliche und systematische Fehler, wie bspw. regelmäßige Überschätzungen, mit der Zeit erkannt und abgestellt werden. Die erste Studie der Dissertation übt Kritik am vorherrschenden Verständnis der Rationalität. Dieses ist zu weitreichend, weshalb den Prognostikern die Rationalität voreilig abgesprochen wird. Anhand einer neuen empirischen Herangehensweise wird deutlich, dass die Prognosen aus einem anderen Blickwinkel heraus durchaus als rational angesehen werden können. Der zweite Aufsatz zeigt auf, dass in Form von Befragungsergebnissen öffentlich verfügbare Informationen bestehen, die bei geeigneter Verwendung zu einer Verbesserung der Qualität von Konjunkturprognosen beitragen würden. Die Rationalität dieser Prognosen ist daher stark eingeschränkt. Im dritten Papier erfolgt eine Analyse von Prognoserevisionen und deren Ursachen. Dabei zeigt sich überraschend, dass es keinen Zusammenhang zwischen der Rationalität und der Qualität der untersuchten Prognosezeitreihen gibt. Die vierte Studie dient der Präsentation der Ergebnisse eines Prognoseplanspiels, welches den Vergleich der Prognosen von Amateuren und Experten zum Ziel hatte. Es stellt sich heraus, dass die Prognosefehler bemerkenswerte Übereinstimmungen aufweisen.

Inhalt

Kapitel I

Einführung und Zusammenfassung	1
--------------------------------------	---

Kapitel II

Zur Beurteilung von Konjunkturprognosen – Eine Auswertung von Prognosen zur Entwicklung des BIP, der Industrieproduktion und der privaten Konsumausgaben in zwölf Industrienationen	15
---	----

Co-Autoren: Markus Spiwoks und Oliver Hein

Diskussionsbeitrag der Sonderforschungsgruppe Institutionenanalyse (sofia), Nr. 11-1, Darmstadt 2011.

Auch erschienen als:

On Assessing Economic Forecasts: An Evaluation of Predictions on GDP, Industrial Production and Private Consumer Spending Trends in Twelve Industrial Nations, in: Advances in Management & Applied Economics, Bd. 4, Nr. 2, S. 49-68, London 2014.

Kapitel III

Verbesserung von Konjunkturprognosen mittels des OECD Composite Leading Indicators und den Ergebnissen der ifo World Economic Survey – eine empirische Untersuchung für die G7-Staaten.....	43
---	----

Diskussionsbeitrag der Sonderforschungsgruppe Institutionenanalyse (sofia), Nr. 11-6, Darmstadt 2011.

Auch erschienen als:

The Improvement of Annual Economic Forecasts by Using Non-Annual Indicators: An Empirical Investigation for the G7 States, in: Wolfsburg Working Papers (WWP), Ostfalia Hochschule für angewandte Wissenschaften, Nr. 12-01.

Kapitel IV

Prognoserevisionen bei Konjunkturprognosen – Empirische Ursachenforschung anhand von Konsensprognosen für das Bruttoinlandsprodukt der G7-Staaten.....83

Wolfsburg Working Papers (WWP), Ostfalia Hochschule für angewandte Wissenschaften, Nr. 14-01.

Kapitel V

Planspiel Kapitalmarktprognose – Ein empirischer Vergleich der Prognosekompetenz von Amateuren und Experten..... 107

Co-Autoren: Kilian Bizer und Markus Spiwoks

Studie der Sonderforschungsgruppe Institutionenanalyse (sofia), Nr. 13-2, Darmstadt 2013.

Kapitel I

Einführung und Zusammenfassung

Prognosebegriff

Diese Dissertation setzt sich mit der Rationalität und Qualität von Wirtschaftsprognosen auseinander. Im weitesten Sinne lässt sich jede Aussage über die Zukunft als Prognose bezeichnen. Ohne die Erstellung von Prognosen auf Basis von Erfahrungswissen ist der Verlauf der Menschheitsgeschichte wohl nicht denkbar. Beispielhaft seien Prognosen bezüglich lohnender Jagd- und Ackerbaugelände sowie dem Anlegen von Vorräten auf Basis des vermuteten Verbrauchs genannt. Der Fortschritt bei Prognosen bezüglich der Statik ließen immer komplexere Bauwerke zu. Aus heutiger Sicht wenig erfolgsversprechende Prognosebemühungen erfuhren ebenfalls große Beliebtheit – wie der jahrhundertlange Erfolg des Orakels von Delphi zeigt.

Im Rahmen dieser Arbeit erfolgt eine Bezugnahme auf wissenschaftliche Prognosen. Diese grenzen sich von den zuletzt genannten Prophezeiungen durch folgende Eigenschaften ab:¹

- **Objektivität:** Die Methoden, Annahmen und etwaige zusätzliche Geltungsbedingungen, die zur Prognose führen, sollten überprüfbar sein.
- **Nichtrivalität:** Die Eindeutigkeit der Prognose muss gegeben sein, mehrdeutige Aussagen sind zu vermeiden.
- **Reliabilität:** Mittels des gleichen Prognoseprozesses sollte sich für den gleichen Prognosegegenstand² und – Gültigkeitszeitpunkt dieselbe Prognose ergeben.
- **Validität:** Die Prognose sollte genau für den vorher benannten Zeitpunkt und Prognosegegenstand gelten. Das Eintreffen muss verifizierbar sein.

Prognosen in den Sozialwissenschaften

Mit der Entstehung moderner Wirtschaftssysteme verstärkte sich der Wunsch, deren Systematik und darauf aufbauend deren zukünftigen Verlauf zu erkennen. In den Mittelpunkt des Interesses rückte in der Folge der Industrialisierung das Auf und Ab der allgemeinen Wirtschaftsleistung – die Konjunktur. Als Entdecker des Konjunkturzyklus bezeichnet Stavenhagen (1951) den französischen Wirtschaftsforscher Juglar, der bereits 1862 die drei Phasen Aufschwung, Krise und Stockung benennt und diesbezüglich eine Wellenbewegung von 8-10 Jahren feststellt.³ Die Erkenntnisse wurden kaum beachtet, bis Crum (1923) und Kitchin (1923) für die

¹ de Groot und Spiekerman (1969), S. 89ff.

² Als Prognosegegenstand wird dabei der Typ des prognostizierten Werts verstanden (Temperatur, Bevölkerungswachstum, Börsenkurse etc.).

³ Juglar (1862)

US-amerikanische Volkswirtschaft ebenfalls einen Konjunkturzyklus feststellten. Dieser fiel mit gut 3 Jahren deutlich kürzer aus. Kondratjev (1926) vermutet eine Überlagerung durch 40- bis 60-jährige langfristige Wellen – deren Existenz jedoch bis heute kontrovers diskutiert wird.

Mit der Identifikation der Zyklen schien die Prognose des Konjunkturverlaufs zunächst nicht besonders schwierig. In den 20er Jahren des letzten Jahrhunderts kam es zur Gründung zahlreicher Wirtschaftsforschungsinstitute. Die erste fundamentale Krise der Wirtschaftsforschung war dann auch die erste große Krise der noch jungen Prognosezunft. Der Beginn der Weltwirtschaftskrise ist weder für Amerika⁴ noch für Deutschland⁵ prognostiziert worden. Dies gilt sowohl für die Konjunktur als auch für die sich ebenfalls in der Entwicklung befindlichen Kapitalmarktprognosen.

Als Folge der ersten „Prognosekrise“ kam es zu einer Grundsatzdebatte. Ausgangspunkt war die bereits vor der Krise von Morgenstern (1928) formulierte These der „Unmöglichkeit von Wirtschaftsprognosen“. Im Kern zielt die Kritik an Wirtschaftsprognosen auf Besonderheiten von sozialwissenschaftlichen Prognosen im Allgemeinen ab:

- Der Verlauf des Prognosegegenstands wird von Akteuren determiniert, deren individuelles Verhalten nicht prognostizierbar sei,
- deren Verhaltensmuster, so sie identifiziert werden können, abrupten Änderungen unterworfen sind (Strukturbrüche, evolutionärer Wissenschaftsgegenstand)
- und die darüber hinaus durch die Kenntnisnahme der Prognose in ihrem Verhalten beeinflusst werden können (self-fulfilling-prophecy und self-destroying-prophecy)⁶.

Nach einer Unterbrechung wird diese theoretische Diskussion hierzulande nach dem zweiten Weltkrieg mit erhöhter Intensität wieder aufgenommen.⁷ Mit der Zeit rückten jedoch neue Wirtschaftstheorien und ökonometrische Prognosemethoden in den Vordergrund, zumal sich die

⁴ Dominguez u. a. (1988)

⁵ Krengel (1984)

⁶ Schöpf (1966, S. 169ff) spricht von der „Selbstaufhebung“ der Prognose, Popper (1957, S. 13) prägt den Begriff „Ödipus-Effekt“ und Merton (1995, orig. 1949) führt den Begriff der „Selbsterfüllenden Prophezeiung“ in die Debatte ein. Die englischen Begriffe „self-fulfilling-prophecy“ und „self-destroying-prophecy“ haben sich weitgehend durchgesetzt.

⁷ Lutz (1955) schließt sich der Argumentation von Morgenstern an und differenziert diese weiter aus. Bosse (1957) und Bombach (1962) halten die self-destroying-prophecy für das gravierendste Problem.

Rechenkapazitäten in den letzten Jahrzehnten vervielfachten. Zuletzt hat Betz (2004) sich intensiv mit der These von Morgenstern beschäftigt. Während er einige Argumente für nicht mehr zeitgemäß hält, kann er im Grundsatz die Morgensternsche Argumentationslinie nicht widerlegen.

Erste Theorien zur Erwartungsbildung

Von zentraler Bedeutung für die moderne Evaluation von Wirtschaftsprognosen sind die Theorien zur Erwartungsbildung der Wirtschaftsakteure. Diese war zunächst bei den vorherrschenden ökonomischen Gleichgewichtsmodellen nicht explizit berücksichtigt worden, wenngleich wirtschaftliches Handeln per se zukunftsgerichtet ist. Auf Fisher (1930) gehen Überlegungen zurück, nach denen die Erwartungsbildung auf Basis vergangener Werte des Prognosegegenstands beruht. Dabei wird neben den vergangenen Werten auch der Prognosefehler berücksichtigt, weshalb von der adaptiven Erwartungsbildung gesprochen wird.

Keynes (1936) setzt sich im Rahmen seiner „General Theory“ intensiv mit der Erwartungsbildung der Wirtschaftsakteure auseinander und präzisiert einzelne Passagen ein Jahr später im *Quarterly Journal of Economics*. Er sieht im Grundsatz eine erhebliche Unsicherheit über die zukünftige Entwicklung,⁸ weshalb die Annahme des Fortbestehens der jeweils aktuellen Situation der Ausgangspunkt jeder Überlegung sein sollte.⁹ Eine Änderung dieser Annahme sollte nur auf Basis guter Gründe erfolgen. Keynes sah die Erwartungsbildung mehrheitlich als einen qualitativen Prozess an, bei dem auch Daumenregeln und Gefühle eine große Rolle spielen.¹⁰

Theorie rationaler Erwartungen

In den Jahrzehnten nach der Veröffentlichung der „General Theory“ kam es zunächst zu keinen maßgeblichen Forschungsbemühungen im Bereich der Erwartungsbildung. Dann jedoch erschienen von Muth (1961) und später Lucas (1972) maßgebliche Arbeiten, die in der Theorie der rationalen Erwartungen mündeten. Im Rahmen dieser Theorie lassen die Wirtschaftssubjekte sämtliche Informationen in ihre Erwartungsbildung einfließen. Hierzu gehören auch absehbare Strukturbrüche – beispielsweise durch politische Eingriffe. Diese wären seitens der

⁸ Keynes (1936), S. 97.

⁹ Keynes (1937), S. 214.

¹⁰ Keynes (1937), S. 222.

Wirtschaftssubjekte nach den bis dahin vorherrschenden Erwartungsbildungstheorien nicht berücksichtigt worden.

Eine zentrale Folge der Theorie der rationalen Erwartungsbildung ist die Schlussfolgerung, dass Prognosefehler nicht systematisch sein dürfen. Mincer und Zarnowitz (1969) stellen hierzu einen Test vor, mit dem Prognosezeitreihen auf das Vorliegen rationaler Erwartungsbildung untersucht werden können. Der sogenannte Test auf Unverzerrtheit findet in zahlreichen empirischen Studien zur Erforschung von Konjunktur- und Kapitalmarktprognosen Anwendung, so auch in allen Beiträgen im Rahmen dieser Dissertation. Mincer und Zarnowitz haben mit ihrer Arbeit erheblich dazu beigetragen, dass in der empirischen Prognoseforschung nicht allein auf die Höhe der Prognosefehler sondern auf die Art und Weise der Fehlerverteilung ein besonderes Augenmerk gelegt wird.

Die Theorie der rationalen Erwartungsbildung führte auch in der Kapitalmarktforschung zu neuen theoretischen Ansätzen. Fama (1970) verknüpft sie in Form der Effizienzmarkthypothese mit der Preisbildung auf den Kapitalmärkten. Er unterscheidet verschiedene Stufen der Informationseffizienz. Bei der sogenannten schwachen Form der Informationseffizienz fließen sämtliche vergangenen Kursinformationen in die Erwartungsbildung der Marktteilnehmer und somit in die Preisbildung ein. Bei der halbstrengen Informationseffizienz spiegeln sich alle öffentlich verfügbaren Informationen in der Preisbildung wider. Zu guter Letzt würden bei der strengen Informationseffizienz auch Insiderinformationen in den Preisen auf den Kapitalmärkten enthalten sein. Für Kapitalmarktprognosen ist die Theorie von Fama von großer Bedeutung: Läge strenge Informationseffizienz vor, so wären Kapitalmarktprognosen völlig unmöglich, da sich kein Informationsvorsprung generieren ließe.

Die von Fama initiierte mehrstufige Einordnung von Informationen, die der Erwartungsbildung und den Prognosen zugrunde liegen, führt auch abseits der Kapitalmärkte zu neuen Forschungsansätzen. Nordhaus (1987) überträgt die Theorie von Fama auch auf volkswirtschaftliche Prognosen. Starke Informationseffizienz sei jedoch nicht zu erwarten, da Ökonomen keine Insiderinformationen vorlägen. Allerdings müssten sie ihre eigenen Prognosefehler zur Kenntnis nehmen und korrigieren, so dass schwache Informationseffizienz vorliegen sollte. Auch Holden und Peel (1990) wenden sich der Informationseffizienz von Prognosen zu. Sie stellen einen Test vor, mit dessen Hilfe geprüft wird, ob eine bestimmte Information in den Prognosen berücksichtigt worden ist, oder ob sich die Prognose mit Hilfe der Information hätte verbessern

lassen. Dieser sogenannte Effizienztest ist eine Erweiterung des Tests auf Unverzerrtheit und findet ebenfalls häufig Verwendung.

Kritik an der Annahme rationaler Erwartungen

Das Konzept der rationalen Erwartungsbildung ist leicht nachvollziehbar und passt hervorragend zum modelltheoretischen Ausgangspunkt zahlreicher ökonomischer Theorien: Dem rational agierenden Agenten als Homo Öconomicus. Tversky und Kahneman (1974) weisen jedoch experimentell nach, dass Menschen keinesfalls in jedem Fall rationale Entscheidungen treffen. Stattdessen kommt es regelmäßig zu Heuristiken und kognitiven Verzerrungen, die nicht zwangsläufig zum höchsten Nutzen für den Anwender führen. Sie sehen besonders Entscheidungen unter Unsicherheit von diesem Phänomen betroffen.¹¹ Dies ist für Wirtschaftsprognosen von Bedeutung, da die Abgabe unter Unsicherheit erfolgt. Zudem sehen sie auch die Informationsverarbeitung der Prognostiker als zu langsam und verzerrt an.¹² Mit dem Status quo bias weisen Samuelson und Zeckhauser (1988) auf eine Form der kognitiven Verzerrung hin, die zur Erklärung „mutloser“ Prognosen beitragen könnte.¹³

Weiterhin besteht auch die Möglichkeit, dass sich die Nutzenfunktion der Prognostiker aus anderen Bestandteilen als einfach nur dem Erreichen eines möglichst geringen Prognosefehlers zusammensetzt. Eine solche Nutzenfunktion könnte zu Verhaltensweisen führen, welche nach außen hin den Kriterien der rationalen Erwartungsbildung nicht genügen, jedoch aus Sicht der einzelnen Prognostiker in ihrem institutionellen Umfeld durchaus rational sind. Es könnte beispielsweise sein, dass Abweichungen in die eine oder andere Richtung unterschiedlich gewichtet werden.¹⁴ Bei Prognosen, die der Unterstützung der Wirtschaftspolitik dienen, wird politische Einflussnahme zur Durchsetzung von Maßnahmen im Sinne der Prognostiker diskutiert.¹⁵ Besteht die Nutzenfunktion der Prognostiker aus mehreren Bestandteilen, kann es nach Laster u. a. (1999) bei der Evaluation ebenfalls zu einer fehlerhaften Feststellung des Ausbleibens rationaler Erwartungsbildung kommen.

¹¹ Kahneman und Tversky (1979).

¹² Tversky u. a. (1981): „We tend to break the good or bad news to ourselves slowly, taking too long to allow surprises to be incorporated into our forecasts“.

¹³ Gubaydullina u. a. (2011) vermuten den Status quo bias als Ursache für Zinsprognosen, die sich wenig vom Gegenwartswert zum Zeitpunkt der Prognoseerstellung unterscheiden.

¹⁴ In diesem Fall spricht man von einer asymmetrischen Verlustfunktion. Vgl. Theil (1971), S. 15-19.

¹⁵ Döpke (2000) untersucht diese Frage für Deutschland, kommt jedoch nicht zu eindeutigen Ergebnissen.

Empirische Studien

Das Vorliegen rationaler Erwartungsbildung bei Wirtschaftsprognosen ist immer wieder kontrovers diskutiert worden. McNees (1978) sieht bei einer Analyse von drei Anbietern von Konjunkturprognosen ein gemischtes Bild hinsichtlich der rationalen Erwartungsbildung – wobei die Ergebnisse besonders vom Prognosehorizont abhängen. Figlewski und Wachtel (1981) sehen bei Inflationsprognosen eher adaptive als rationale Erwartungen vorliegen. Keane und Runkle (1990) sehen vorhergehende Ergebnisse, die rationale Erwartungsbildung ablehnen, in einer ungeeigneten Untersuchungsmethodik begründet und schlagen ein erweitertes Verfahren vor, in dessen Licht die Prognosen rational seien. Bonham und Cohen (1995) kritisieren wiederum die neue Methodik als unzureichend. Eine ähnliche Debatte findet in Deutschland in den 80er Jahren statt.¹⁶ In den 2000ern finden Heilemann und Stekler (2003) Hinweise, dass die Prognosefehler von BIP- und Inflationsprognosen mit den Prognosen zusammenhängen und autokorreliert sind. Beides widerspricht der Theorie rationaler Erwartungsbildung. Döpke und Fritsche (2006) lehnen das Vorliegen von Rationalität in Abhängigkeit vom Maßstab mehrheitlich ab. Elliott u. a. (2008) sehen die bei vielen Studien vorangestellte Annahme als problematisch an, dass die Verlustfunktion der Prognostiker symmetrisch sei.

Neben der Rationalität steht die Frage im Raum, wie es um die Qualität von Wirtschaftsprognosen aus Sicht der Anwender bestellt ist. Bei volkswirtschaftlichen Prognosen, die in der zweiten Jahreshälfte für das Folgejahr abgegeben werden, kann je nach Zeitpunkt der Abgabe ein 10-30% niedrigeres Fehlerniveau erreicht werden, als es das einfache Fortschreiben der jeweiligen gemessenen Vorjahreswerte erwarten lässt.¹⁷ Nichtsdestotrotz liegen die gemessenen Prognosefehler in einer aktuellen Studie von Heilemann und Stekler (2013) mit 1,3 Prozentpunkten bei BIP-Prognosen und 0,7 Prozentpunkten bei Inflationsprognosen auf einem hohen Niveau. Dies entspricht dem allgemeinen Stand der Forschung der letzten Jahre.

Wegen des Fehlerniveaus keimt immer wieder fundamentale Kritik an der Abgabe und Nutzung volkswirtschaftlicher Prognosen auf. Zarnowitz (1992) fragt „Has macro-forecasting failed?“ und kann die Frage nicht eindeutig beantworten. Heilemann und Stekler (2003) werfen die

¹⁶ Kirchgässner (1982) sieht Umfrageergebnisse des Allensbach-Instituts zur Wirtschaftsentwicklung in Deutschland als rational an, während Neumann und Buscher (1985) die Vorgehensweise von Kirchgässner als völlig ungeeignet ansehen. Nachdem er zunächst in zahlreichen Publikationen Rationalität festgestellt hat, hat Kirchgässner inzwischen Zweifel an den Fähigkeiten der Prognostiker, aus ihren eigenen Fehlern zu lernen. Vgl. Kirchgässner und Müller (2006).

¹⁷ Siehe beispielsweise Heilemann und Stekler (2013) oder auch Scheier (2012).

Frage auf, ob sich makroökonomische Prognosen für Deutschland in der zweiten Hälfte des 20. Jahrhunderts verbessert haben. Sie sehen zwar verringerte Prognosefehler, welche jedoch auf geringere Niveaus der prognostizierten Variablen zurückzuführen sein könnten. 10 Jahre später sehen sie sich in dieser Auffassung bestätigt, da mit der höheren Volatilität der Prognosegegenstände auch die absoluten Prognosefehler gestiegen seien.¹⁸ Die Ergebnisse ließen sich auch auf die USA und Großbritannien übertragen.

Bei Kapitalmarktprognosen ist die Frage nach der Qualität aus Sicht der Anwender einfacher zu beantworten: Der Prognoseerfolg bleibt weitgehend aus. Dies ist gleichermaßen für Prognosen von Zinsen¹⁹, Aktien-²⁰ und Wechselkursen²¹ festzustellen. Schwarzbach u. a. (2012) beantworten die Frage aus Sicht der Kapitalanlage von Versicherungsunternehmen und raten von der Prognosenutzung ab. Ein wiederkehrendes Muster zeigt sich bei der Art und Weise wie die Prognostiker bei Kapitalmarktprognosen scheitern. Sie scheinen bei ihren Einschätzungen sehr dicht am jeweiligen Wert zum Zeitpunkt der Prognoseerstellung zu liegen und darüber jeden Zukunftsbezug zu verlieren. Andres und Spiwoks (1999) nennen dieses Phänomen gegenwartsorientierte Verlaufsanpassung und stellen ein Verfahren für dessen Messung vor.

Erster Beitrag – Zur Beurteilung von Konjunkturprognosen

Der erste Beitrag dieser Dissertation mit dem Titel „Zur Beurteilung von Konjunkturprognosen – Eine Auswertung von Prognosen zur Entwicklung des BIP, der Industrieproduktion und der privaten Konsumausgaben in zwölf Industrienationen“ setzt sich kritisch mit den aus der Theorie der rationalen Erwartungsbildung nach Muth (1961) heraus entwickelten Maßstäben für Wirtschaftsprognosen im Allgemeinen und dem Test auf Unverzerrtheit von Mincer und Zarnowitz (1969) im Speziellen auseinander. Dieser führt bei volkswirtschaftlichen Prognosen einer umfangreichen Datenbasis zu einem eindeutigen Urteil: Keine der betrachteten Prognosezeitreihen kann als rational angesehen werden.

¹⁸ Heilemann und Stekler (2013).

¹⁹ So stellt beispielsweise Simon (1989) für Zinsprognosen von 30 professionellen Analysten fest, dass diese kaum besser als ein random-walk sind. Die 3- und 12-Monats-Prognosen des Renditeniveaus 10-jähriger Staatsanleihen gelten als guter Indikator für die Möglichkeit, das Zinsniveau vorherzusehen. Scheier und Spiwoks (2006) für Großbritannien, Spiwoks u. a. (2008) für die USA, Spiwoks u. a. (2009) für die Schweiz sowie Spiwoks (2002) für Deutschland kommen zu dem Schluss, dass die überwältigende Mehrheit der Prognosen schlechter als die sogenannten naiven Prognosen abschneiden.

²⁰ Spiwoks (2004), Benke (2006) sowie Söderlind (2010) sprechen den Aktienkurs- bzw. Aktienindexprognosen jeden Nutzen ab.

²¹ Bofinger und Schmidt (2003) stellen für Wechselkurse ein Fehlerniveau fest, welches wie bei den anderen Kapitalmarktprognosen höher als bei der sogenannten naiven Prognose liegt.

Hier greift der Rationalitätsbegriff aus Sicht der Autoren zu kurz, da dieser auf eine Kenntnis der volkswirtschaftlichen Zusammenhänge seitens der Prognostiker abstellt. In der Studie werden zwei Testansätze präsentiert, die besser zur Beurteilung der Prognosequalität und der Erwartungsbildung der Prognostiker geeignet sind. Mit dem erstmals vorgestellten Test auf Informationszuwachs wird überprüft, ob die Prognostiker bei sich verkürzendem Prognosehorizont von der zunehmenden Informationsbasis profitieren können und sich die Prognosen systematisch verbessern. Dieses durchaus als rational zu wertende Verhalten ist bei der Mehrheit der betrachteten Zeitreihen festzustellen.

Weiterhin wird der Test auf gegenwartsorientierte Verlaufsanpassung nach Andres und Spiwoks (1999) erstmals auf volkswirtschaftliche Prognosezeitreihen angewandt. Dabei stellt sich heraus, dass anders als bei Kapitalmarktprognosen der Zukunftsbezug der Prognosen deutlich ausgeprägt ist und den Bezug zur Gegenwart zum Zeitpunkt der Prognoseerstellung mehrheitlich übertrifft.

Zweiter Beitrag - Verbesserung von Konjunkturprognosen

Der zweite Beitrag mit dem Titel „Verbesserung von Konjunkturprognosen mittels des OECD Composite Leading Indicators und den Ergebnissen der ifo World Economic Survey – eine empirische Untersuchung für die G7-Staaten“ greift ebenfalls Aspekte der Evaluationsmethodik nach der Theorie der rationalen Erwartungsbildung auf. Eine zentrale Forderung an die Prognostiker ist es, alle öffentlich verfügbaren Informationen in die Prognosen mit einfließen zu lassen.

In der Studie werden mit dem OECD Composite Leading Indicator und den Ergebnissen der ifo World Economic Survey zwei bekannte und öffentlich verfügbare Konjunkturindikatoren vorgestellt. Obwohl sie nicht zu diesem Zweck erhoben werden, weisen sie zeitweise einen stärkeren Zusammenhang zum Wirtschaftswachstum des folgenden Jahres auf als Punktprognosen professioneller Prognostiker von Banken und Wirtschaftsforschungsinstituten. Mittels des Effizienztests von Holden und Peel (1990) stellt sich heraus, dass die Mehrzahl der Prognosezeitreihen zweifelsohne von der Berücksichtigung der Konjunkturindikatoren profitiert hätte – ein klarer Verstoß gegen die Rationalitätsannahme.

Dritter Beitrag – Prognoserevisionen bei Konjunkturprognosen

Der dritte Beitrag mit dem Titel „Prognoserevisionen bei Konjunkturprognosen – Empirische Ursachenforschung anhand von Konsensprognosen für das Bruttoinlandsprodukt der G7-Staaten“ setzt sich kritisch mit der ebenfalls aus der Theorie der rationalen Erwartungsbildung abgeleiteten These auseinander, dass rationale Prognoserevisionen einem random-walk unterliegen müssten. Diese Forderung ist gemeinsam mit einer entsprechenden Untersuchungsmethodik von Nordhaus (1987) vorgestellt worden.

Die Studie bestätigt den Stand der Forschung zunächst insofern, als dass mehrheitlich kein random-walk vorliegt und die Annahme rationaler Erwartungsbildung somit abzulehnen ist. Allerdings wird die Frage aufgeworfen, wie hilfreich diese Erkenntnis im Hinblick auf die Beurteilung der Prognosequalität ist. Bei den längeren Prognosehorizonten liegt zwar ein random-walk der Revisionen vor – die Qualität gemessen an der Höhe der Prognosefehler ist aber sehr schlecht. Sobald sich diese verbessert, bleibt der random-walk aus. Im Fazit sind die betrachteten Konjunkturprognosen „rational schlecht“, bevor sie sich „irrational verbessern“. Möglicherweise liegt dies im Prognosegegenstand selbst begründet: Wenn sich die Konjunkturentwicklung schrittweise in eine Richtung vollzieht, könnte sich auch der Informationsstand der Prognostiker in dieser Form entwickeln.

Die Studie richtet den Blick darüber hinaus auf mögliche Ursachen der Prognoserevisionen. Es fällt auf, dass die frühen Prognosen im Durchschnitt deutlich zu hoch abgegeben werden. Ein weiteres Merkmal der Prognosen mit längerem Prognosehorizont ist die gleichmäßige Abgabe. Die jahresübergreifende Standardabweichung der abgegebenen Prognosen ist erheblich niedriger als die der prognostizierten Wachstumsraten.

Vierter Beitrag – Planspiel Kapitalmarktprognose

Der vierte Beitrag mit dem Titel „Planspiel Kapitalmarktprognose – Ein empirischer Vergleich der Prognosekompetenz von Amateuren und Experten“ setzt sich mit der Frage auseinander, ob sich bestimmte wiederkehrende Muster bei der Beurteilung von Kapitalmarktprognosen von Finanzanalysten in einem Planspiel replizieren lassen. Hierzu geben Studierende in kleinen Gruppen über ein ganzes Semester hinweg Aktienkursprognosen ab. Methodisch lässt sich die Vorgehensweise zur Datenerhebung als eine Mischung zwischen Labor- und Feldexperiment einordnen.

Ein zentrales Ergebnis ist es, dass sich die Prognosequalität zwischen Analysten und Planspielteilnehmern nicht maßgeblich unterscheidet. Weiterhin sind beide nicht rational im Sinne der Theorie der rationalen Erwartungsbildung und weisen darüber hinaus überwiegend eine starke gegenwartsorientierte Verlaufsanpassung auf. Die Studienergebnisse stellen bestimmte Erklärungen für die beobachteten Prognosemuster in Frage, die auf das institutionelle Umfeld von Finanzanalysten abstellen. So sollten Reputationsverluste durch Fehlprognosen, die sich von der Mehrheit der Prognosen abheben, kein Bestandteil des Anreizsystems der Planspielteilnehmer gewesen sein.

Literatur

- Andres, P. / Spiwox, M. (1999): „Prognosequalitätsmatrix - Ein methodologischer Beitrag zur Beurteilung der Güte von Kapitalmarktprognosen“, *Jahrbücher für Nationalökonomie und Statistik*, Bd. 219, Nr. 5, S. 513–542.
- Benke, H. (2006): „Was leisten Kapitalmarktprognosen? Die Sicht eines Stiftungsmanagers, in: Zeitschrift für das gesamte Kreditwesen“, *Zeitschrift für das gesamte Kreditwesen*, Nr. 17, S. 40–49.
- Betz, G. (2004): „Apriorische und empirische Grenzen von Wirtschaftsprognosen: Oskar Morgenstern nach 70 Jahren Prognoseerfahrung“. In Frank, U. (Herausgeber) *Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik: Theoriebildung und -bewertung, Ontologien, Wissensmanagement*, Wiesbaden, S. 171–190.
- Bofinger, P. / Schmidt, R. (2003): „On the Reliability of Professional Exchange Rate Forecasts: An Empirical Analysis for the €US-\$ Rate“, *Financial Markets and Portfolio Management*, Bd. 17, Nr. 4, S. 437–449.
- Bombach, G. (1962): „Über die Möglichkeit wirtschaftlicher Voraussagen“, *Kyklos*, Bd. 15, Nr. 1, S. 29–67.
- Bonham, C. / Cohen, R. (1995): „Testing the rationality of price forecasts: Comment.“, *American Economic Review*, Bd. 85, Nr. 1, S. 284.
- Bosse, L. (1957): „Über die Möglichkeit und den Nutzen von kurzfristigen Wirtschaftsprognosen“, *Weltwirtschaftliches Archiv: Zeitschrift des Instituts für Weltwirtschaft an der Universität Kiel*, Bd. 79, S. 65–83.
- Crum, W. L. (1923): „Cycles of rates on commercial paper“, *The Review of Economic Statistics*, S. 17–29.
- Dominguez, K. M. / Fair, R. C. / Shapiro, M. D. (1988): „Forecasting the Depression: Harvard versus Yale“, *The American Economic Review*, S. 595–612.
- Döpke, J. (2000): „Haben Konjunkturprognosen in Deutschland einen politischen Bias?“, *Schmollers Jahrbuch*, Bd. 120, Nr. 4, S. 587–620.

- Döpke, J. / Fritsche, U. (2006): „Growth and inflation forecasts for Germany: a panel-based assessment of accuracy and efficiency“, *Empirical economics*, Bd. 31, Nr. 3, S. 777–798.
- Elliott, G. / Komunjer, I. / Timmermann, A. (2008): „Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?“, *Journal of the European Economic Association*, Bd. 6, Nr. 1, S. 122–157.
- Fama, E. (1970): „Efficient Capital Markets, A Review of Theory and Empirical Work“, *Journal of Finance*, Bd. 25, S. 383–417.
- Figlewski, S. / Wachtel, P. (1981): „The Formation of Inflationary Expectations“, *The review of economics and statistics*, Bd. 63, Nr. 1, S. 1–10.
- Fisher, I. (1930): *The Theory of Interest*, New York.
- De Groot, A. D. / Spiekerman, J. A. A. (1969): *Methodology (Methodologie, engl.) Foundations of inference and research in the behavioral sciences*, MTH.
- Gubaydullina, Z. / Hein, O. / Spiwoks, M. (2011): „The Status Quo Bias of Bond Market Analysts“, *Journal of Applied Finance & Banking*, Bd. 1, Nr. 1, S. 31–51.
- Heilemann, U. / Stekler, H. O. (2003): „Has the accuracy of German macroeconomic forecasts improved?“, *Technical Report Universität Dortmund, SFB 475 Komplexitätsreduktion in Multivariaten Datenstrukturen*, Bd. 7, Nr. 31.
- Heilemann, U. / Stekler, H. O. (2013): „Has The Accuracy of Macroeconomic Forecasts for Germany Improved?“, *German Economic Review*, Bd. 14, Nr. 2, S. 235–253.
- Holden, K. / Peel, D. A. (1990): „On testing for unbiasedness and efficiency of forecasts“, *The Manchester School of Economic and Social Studies*, Bd. 58, Nr. 2, S. 120–127.
- Juglar, C. (1862): *Des crises commerciales et de leur retour périodique en France, en Angleterre et aux Etats Unis*, Paris.
- Kahneman, D. / Tversky, A. (1979): „Prospect Theory: An Analysis of Decision under Risk“, *Econometrica*, Bd. 47, Nr. 2, S. 263–292.
- Keane, M. P. / Runkle, D. E. (1990): „Testing the Rationality of Price Forecasts: New Evidence from Panel Data“, *The American Economic Review*, Bd. 80, Nr. 4, S. 714–735.
- Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money*, London.
- Keynes, J. M. (1937): „The General Theory of Employment“, *The Quarterly Journal of Economics*, Bd. 51, Nr. 2, S. 209–223.
- Kirchgässner, G. (1982): „Sind die Erwartungen der Wirtschaftssubjekte ‚rational‘? Eine empirische Untersuchung für die Bundesrepublik Deutschland“, *Review of World Economics*, Bd. 118, Nr. 2, S. 215–240.

- Kirchgässner, G. / Müller, U. K. (2006): „Are forecasters reluctant to revise their predictions? Some German evidence“, *Journal of Forecasting*, Bd. 25, Nr. 6, S. 401–413.
- Kitchin, J. (1923): „Cycles and trends in economic factors“, *The Review of Economic Statistics*, S. 10–16.
- Kondratjev, N. N. (1926): Die langen Wellen der Konjunktur, Mohr.
- Krengel, R. (1984): „Die wirtschaftliche Entwicklung im Deutschen Reich in den 20er und 30er Jahren aus der Sicht des Instituts für Konjunkturforschung“, *Vierteljahrshefte zur Wirtschaftsforschung*, Nr. 1, S. 30–37.
- Laster, D. / Bennett, P. / Geom, I. S. (1999): „Rational Bias in Macroeconomic Forecasts“, *Quarterly Journal of Economics*, Bd. 114, Nr. 1, S. 293–318.
- Lucas, R. E. (1972): „Expectations and the neutrality of money“, *Journal of Economic Theory*, Bd. 4, Nr. 2, S. 103–124.
- Lutz, F. A. (1955): „Das Problem der Wirtschaftsprognosen“. In *Recht und Staat in Geschichte und Gegenwart - Eine Sammlung von Vorträgen und Schriften aus dem Gebiet der gesamten Staatswissenschaften*, Tübingen, Mohr, S. 1–20.
- McNees, S. K. (1978): „The ‚Rationality‘ Of Economic Forecasts.“, *American Economic Review*, Bd. 68, Nr. 2, S. 301.
- Merton, R. K. (1995): Soziologische Theorie und soziale Struktur, Bod Third Party Titles.
- Mincer, J. / Zarnowitz, V. (1969): „The Evaluation of Economic Forecasts“. In Mincer, J. (herausgeber) *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, New York, S. 14–20.
- Morgenstern, O. (1928): Wirtschaftsprognose. Eine Untersuchung ihrer Möglichkeiten und Voraussetzungen, Wien.
- Muth, J. F. (1961): „Rational Expectations and the Theory of Price Movements“, *Econometrica*, Bd. 29, Nr. 3, S. 315–335.
- Neumann, M. J. M. / Buscher, H. S. (1985): „Wie gut sind die Prognosen der Arbeitsgemeinschaft wirtschaftswissenschaftlicher Forschungsinstitute in der BRD? - Ein Kommentar“, *Weltwirtschaftliches Archiv*, Bd. 121, Nr. 2, S. 321–330.
- Nordhaus, W. D. (1987): „Forecasting Efficiency: Concepts and Applications“, *The review of economics and statistics*, Bd. 69, Nr. 4, S. 667–674.
- Popper, K. R. (1957): The poverty of historicism, London.
- Samuelson, W. / Zeckhauser, R. (1988): „Status Quo Bias in Decision Making“, *Journal of Risk and Uncertainty*, Bd. 1, Nr. 1, S. 7–59.

- Scheier, J. (2012): „The improvement of annual economic forecasts by using non-annual indicators - An empirical investigation for the G7 states“, *Wolfsburg Working Papers*, Bd. 12-01.
- Scheier, J. / Spiwoks, M. (2006): Aktives Portfoliomanagement am britischen Anleihenmarkt, in: Meyer-Bullerdiek, F. / Spiwoks, M. (Herausgeber), Reihe Bank- und Finanzwirtschaft, Bd. 4, Frankfurt am Main.
- Schöpf, A. (1966): Das Prognoseproblem in der Nationalökonomie - Versuch einer Gesamtbeurteilung, Reihe Beiträge zur ganzheitlichen Wirtschafts- und Gesellschaftslehre, Berlin, München.
- Schwarzbach, C. / Kunze, F. / Rudschuck, N. / Windels, T. (2012): „Asset management in the German insurance industry: the quality of interest rate forecasts“, *Zeitschrift für die gesamte Versicherungswirtschaft*, Bd. 101, Nr. 5, S. 693–703.
- Simon, D. P. (1989): „The rationality of federal funds rate expectations: evidence from a survey“, *Journal of Money, Credit and Banking*, S. 388–393.
- Söderlind, P. (2010): „Predicting stock price movements: regressions versus economists“, *Applied Economics Letters*, Bd. 17, Nr. 9, S. 869–874.
- Spiwoks, M. (2002): Vermögensverwaltung und Kapitalmarktprognose, in: Meyer-Bullerdiek, F. / Spiwoks, M. (Herausgeber), Reihe Bank- und Finanzwirtschaft, Bd. 1.
- Spiwoks, M. (2004): „Die Verwendbarkeit der ZEW-Aktienindex-Prognosen für aktive Portfoliomanagement-Strategien“, *Jahrbücher für Nationalökonomie und Statistik*, Bd. 224, Nr. 5, S. 557–578.
- Spiwoks, M. / Bedke, N. / Hein, O. (2008): „Forecasting the past - The case of U.S. interest rate forecasts“, *Financial Markets and Portfolio Management*, Bd. 22, S. 357–379.
- Spiwoks, M. / Bedke, N. / Hein, O. (2009): „The Pessimism of Swiss Bond Market Analysts and the Limits of the Sign Accuracy Test – An Empirical Investigation of Their Forecasting Success Between 1998 and 2007“, *International Bulletin of Business Administration*, Bd. 4, S. 6–19.
- Stavenhagen, G. (1951): Geschichte der Wirtschaftstheorie, Vandenhoeck und Ruprecht.
- Theil, H. (1971): Applied Economic Forecasting, in: Theil, H., Reihe Studies in Mathematical and Managerial Economics, Bd. 4, Amsterdam, North-Holland Publishing Company.
- Tversky, A. / Kahneman, D. (1974): „Judgement under Uncertainty: Heuristics and Biases“, *Science*, Bd. 185, S. 1124–1131.
- Tversky, A. / Kahneman, D. / Choice, R. (1981): „The framing of decisions“, *Science*, Bd. 211, S. 453–458.
- Zarnowitz, V. (1992): „Has macro-forecasting failed?“, *Cato Journal*, Bd. 12, Nr. 1, S. 129–160.

Kapitel 2

Zur Beurteilung von Konjunkturprognosen

Eine Auswertung von Prognosen zur Entwicklung des BIP, der Industrieproduktion und der privaten Konsumausgaben in zwölf Industrienationen

mit Markus Spiwoks und Oliver Hein

Erschienen als:

Diskussionsbeitrag der Sonderforschungsgruppe Institutionenanalyse (sofia), Nr. 11-1, Darmstadt 2011.

Auch erschienen als:

On Assessing Economic Forecasts: An Evaluation of Predictions on GDP, Industrial Production and Private Consumer Spending Trends in Twelve Industrial Nations, in: Advances in Management & Applied Economics, Bd. 4, Nr. 2, S. 49-68, London 2014.

Abstract

Diese Studie befasst sich mit Konjunkturprognosen und ihrer Beurteilung. Wir werten Prognosen zum Wachstum des BIP, der Industrieproduktion und des privaten Konsums in zwölf bedeutenden Industrienationen aus. Neben dem Test auf Unverzerrtheit und dem Vergleich zur naiven Prognose werden auch der Test auf Informationszuwachs und der GOVA-Koeffizient angewendet. Es zeigt sich, dass die Konjunkturprognosen im Großen und Ganzen recht erfolgreich sind. Es zeigt sich ferner, dass der Test auf Unverzerrtheit ungeeignet ist, um Erfolge und Misserfolge der Prognostiker angemessen zu differenzieren.

1 Einleitung

Konjunkturprognosen sind unverzichtbar. Sowohl das Regierungshandeln als auch viele Unternehmensentscheidungen hängen wesentlich von der erwarteten konjunkturellen Entwicklung ab.

Die Abschätzung der künftigen Steuereinnahmen, die Prognose der notwendigen Sozialausgaben und damit natürlich die gesamte Finanzpolitik eines Staates hängen wesentlich von der erwarteten konjunkturellen Entwicklung ab. Sofern sich eine Regierung dafür verantwortlich fühlt, steuernd in die konjunkturelle Entwicklung einzugreifen, ist sie ebenfalls auf Konjunkturprognosen angewiesen. Auch die Notenbanken erstellen in aller Regel aufwendige Konjunkturprognosen, beispielsweise um eine drohende Überhitzung des Wirtschaftswachstums und somit eine Inflationsgefahr frühzeitig abwenden zu können.

Unternehmen des produzierenden Gewerbes orientieren sich an Konjunkturprognosen, um ihren künftigen Investitions- und Personalbedarf planen zu können oder um sich rechtzeitig den Zugang zu Rohstoffen zu sichern. Für den Bankensektor spielen Prognosen des Wirtschaftswachstums eine besonders wichtige Rolle. Konjunkturprognosen sind notwendig, damit die Banken die Wahrscheinlichkeit für den Ausfall von aktuellen und künftigen Kreditnehmern abschätzen können. Nur so können die Banken einen angemessenen Risikoaufschlag in der Zinsmarge berücksichtigen. Konjunkturprognosen sind auch für das Investmentgeschäft von erheblicher Bedeutung. Die konjunkturelle Entwicklung entfaltet wesentlichen Einfluss auf die Aktien- und Anleihenkurse. Finanzmarktprognosen basieren somit in aller Regel auch auf Konjunkturprognosen.

So wichtig wie Konjunkturprognosen sind, so umstritten sind sie zugleich auch. Das hängt insbesondere damit zusammen, dass der Prognoseerfolg häufig hinter den hochgesteckten Erwartungen zurückbleibt. Die Prognose der konjunkturellen Entwicklung ist eine schwierige Aufgabe. Eine Volkswirtschaft ist ein komplexes Gebilde, das sich darüber hinaus permanent verändert. Evolutionäre Wissenschaftsgegenstände sind besonders schwierig zu ergründen. Das musste die Ökonomen-Gemeinde 2008 im Zuge der Finanzmarktkrise und der darauf folgenden Wirtschaftskrise erneut erkennen.

Die eingeschränkten Möglichkeiten für die Erstellung zuverlässiger Konjunkturprognosen stehen in starkem Kontrast zu den häufig vollkommen überzogenen Ansprüchen der Nutzer dieser

Prognosen. Muth (1961) hat mit seiner Hypothese rationaler Erwartungen einen vielleicht entscheidenden Beitrag zu der unrealistischen Einschätzung der Möglichkeiten von Konjunkturprognosen geleistet. Die Hypothese rationaler Erwartungen geht davon aus, dass es bei Berücksichtigung aller relevanten Informationen möglich ist, einen ökonomischen Sachverhalt weitgehend zutreffend zu prognostizieren. Lediglich unvorhersehbare Zufallseinflüsse (das so genannte weiße Rauschen) tragen dazu bei, dass die Prognostiker nicht jede Entwicklung vollkommen zutreffend vorhersagen können.

Ob die Prognosefehler wirklich nur auf unsystematische Zufallseinflüsse zurückzuführen sind, wird in der Regel mit dem Test auf Unverzerrtheit überprüft. Sind die Prognosen nicht unverzerrt, so gelten sie auch nicht als rational.

Die Nutzer von Prognosen fordern von den Experten vernünftige und sinnvolle Prognosen. Es ist nicht immer leicht zu vermitteln, dass vernünftige und sinnvolle Prognosen nicht zwangsläufig auch rational im Sinne von Muth (1961) sein müssen. Denn Muth setzt voraus, dass das relevante Modell der Ökonomie bekannt ist, dass also das Funktionieren der Volkswirtschaften weitgehend ergründet ist. Diese Annahme ist aber offensichtlich unzutreffend. Die Bedingungen für wirtschaftliches Wachstum sind nur in Grundzügen bekannt. Viele Einflussfaktoren werden bisher nicht, nicht ausreichend oder nicht korrekt berücksichtigt. In einer Welt unvollkommener Kenntnis über die Zusammenhänge und Wirkmechanismen in der Volkswirtschaft muss man bei der Beurteilung von Prognostikern andere Maßstäbe als den Test auf Unverzerrtheit anlegen.

Die vorliegende Studie verfolgt vier Ziele:

1. Die Studie soll einen kurzen Überblick über die wichtigsten Verfahren der Prognosegütemessung bei makroökonomischen Prognosen geben.
2. Es sollen zwei neue Verfahren zur Beurteilung makroökonomischer Prognosen vorgestellt werden: Der Test auf Informationszuwachs und der GOVA-Koeffizient. Unseres Erachtens handelt es sich um besonders geeignete Ansätze, weil sie realistische Anforderungen an die Prognostiker stellen.
3. Es soll gezeigt werden, dass Prognoseerfolge und Prognosemisserfolge mit dem Test auf Unverzerrtheit nicht angemessen differenziert werden können.

4. Die Studie soll einen Überblick über den Erfolg von Konjunkturprognosen in zwölf Industrienationen geben. Im Einzelnen handelt es sich um Konsensprognosen zur Entwicklung des BIP, der Industrieproduktion und der privaten Konsumausgaben in den USA, Japan, Deutschland, Frankreich, Großbritannien, Italien, Spanien, Kanada, Niederlande, Schweiz, Schweden und Norwegen.

Das nächste Kapitel stellt die verwendeten Prognosedaten näher vor. Die Methodologie steht im Mittelpunkt des übernächsten Kapitels. Im vorletzten Kapitel präsentieren wir die empirischen Ergebnisse. Das letzte Kapitel fasst die wichtigsten Ergebnisse der Studie zusammen.

2 Datenbasis

Im Mittelpunkt der Untersuchung steht die Beurteilung von Konjunkturprognosen. Im Einzelnen handelt es sich um Prognosen zum Wachstum des BIP, der Industrieproduktion und der privaten Konsumausgaben. Die Prognosen beziehen sich jeweils auf die prozentuale Veränderung gegenüber dem letzten Kalenderjahr. Die Prognosen stammen aus der Zeitschrift *Consensus Forecasts*. Dieses Magazin veröffentlicht seit Oktober 1989 in monatlichem Rhythmus Prognosen zu verschiedenen makroökonomischen Größen. *Consensus Forecasts* befragt Analysten und Wissenschaftler aus Banken, Versicherungen, Investmentgesellschaften, Forschungsinstituten und Verbänden. Das Magazin veröffentlicht alle Prognosen der Experten und fasst diese schließlich zu Konsensprognosen zusammen.

Die Prognosen beziehen sich auf die prozentuale Veränderung des laufenden Jahres im Vergleich zum Vorjahr und auf die prozentuale Veränderung des folgenden Jahres im Vergleich zum laufenden Jahr. Die Experten geben ihre Einschätzungen jeweils zu Monatsbeginn ab. Die Prognosen beziehen sich jeweils auf das Ende des Jahres, weil erst dann die prozentuale Veränderung gegenüber dem vorherigen Kalenderjahr ermittelt werden kann. Da die Experten jeden Monat diese Prognosen abgeben, resultieren für jeden Prognosegegenstand 24 unterschiedliche Prognosehorizonte. Gibt man beispielsweise Anfang Januar 2011 eine Prognose für das folgende Jahr ab (Stichtag 31.12.2012), so weist diese Prognose einen Prognosehorizont von 24 Monaten auf. Gleichzeitig gibt man eine Prognose für das laufende Jahr ab (Stichtag 31.12.2011). Diese Prognose weist somit einen Prognosehorizont von zwölf Monaten auf. Gibt man Anfang Februar wieder je eine Prognose zum folgenden Jahr und zum laufenden Jahr ab, verkürzen sich die Prognosehorizonte auf 23 Monate beziehungsweise auf elf Monate. Gibt man die zwei Prognosen Anfang Dezember ab, so betragen die Prognosehorizonte noch 13

Monate beziehungsweise einen Monat. Für jeden Prognosegegenstand und jedes prognostizierte Jahr existieren somit 24 Konsensprognosen mit den Prognosehorizonten 24 Monate bis einen Monat. Die Experten beginnen also jeweils zwei Jahre vor Ende eines Kalenderjahres, ihre Prognosen abzugeben. Sie geben jeden Monat neue Prognosen ab, bis nach 24 Monaten das prognostizierte Ereignis eintritt. Diese besondere Struktur der Prognosedaten eröffnet interessante Analysemöglichkeiten.

Consensus Forecasts veröffentlicht bereits seit Oktober 1989 Prognosen für die USA, Japan, Deutschland, Frankreich, Großbritannien, Italien und Kanada. Ab Dezember 1994 finden auch Spanien, Schweden und die Niederlande Berücksichtigung. Im Juni 1998 kommen schließlich noch Norwegen und die Schweiz hinzu. Wir berücksichtigen in unserer Studie nur die Jahre, für die alle 24 Prognosehorizonte vorliegen. Die Analyse der ersten Ländergruppe umfasst somit die Jahre 1991 bis 2009, die der zweiten Ländergruppe von 1996 bis 2009 und die der dritten Ländergruppe 2000 bis 2009 (siehe Tabelle 1).

Tabelle 1: Verwendeter Datenbestand aus der Zeitschrift Consensus Forecasts

Land	Beginn der Veröffentlichung	Erstes vollständig prognostiziertes Jahr	Letztes vollständig prognostiziertes Jahr	Anzahl der analysierten Prognosen
USA	Okt. 1989	1991	2009	1.368
Japan	Okt. 1989	1991	2009	1.368
Deutschland	Okt. 1989	1991	2009	1.368
Frankreich	Okt. 1989	1991	2009	1.368
Großbrit.	Okt. 1989	1991	2009	1.368
Italien	Okt. 1989	1991	2009	1.368
Spanien	Dez. 1994	1996	2009	1.008
Kanada	Okt. 1989	1991	2009	1.368
Niederlande	Dez. 1994	1996	2009	1.008
Schweiz	Juni 1998	2000	2009	720
Schweden	Dez. 1994	1996	2009	1.008
Norwegen	Juni 1998	2000	2009	720
Gesamt				14.040

Für die erste Ländergruppe liegen also je 24 Prognosen für jedes der 19 betrachteten Jahre (von 1991 bis 2009) vor. Das sind insgesamt 456 Prognosen. Da wir bei jedem Land jeweils drei Prognosegegenstände betrachten (BIP, Industrieproduktion, privater Konsum), führt dies zu

1.368 Prognosen pro Land der ersten Ländergruppe. Insgesamt werten wir 14.040 Konsensprognosen aus.

3 Methodik

Prognosen sollten im günstigsten Fall unverzerrt sein (Diebold, 2007). Das bedeutet, dass der Erwartungswert der Abweichungen zwischen den Prognosen und den tatsächlichen Ereignissen Null sein sollte. Außerdem sollte die Zeitreihe der Residuale keine systematischen Komponenten enthalten, sondern zufällig verteilt sein.

Am häufigsten findet man den Test auf Unverzerrtheit in Form der Mincer-Zarnowitz-Regression (Ball, 1962; Mincer und Zarnowitz, 1969). Seien A_t der tatsächliche Wert, P_t die dazugehörige Prognose und u_t das Residual zum Zeitpunkt t . Die Mincer-Zarnowitz-Regression hat dann folgende Gestalt:

$$(1) \quad A_t = \alpha + \beta P_t + u_t$$

Basierend auf dieser Gleichung gelten die Prognosen als unverzerrt, wenn sich α nicht signifikant von 0 und β sich nicht signifikant von 1 unterscheiden. Zudem darf der Fehlerterm u nicht autokorreliert sein.

Wir stehen dem Test auf Unverzerrtheit allerdings sehr kritisch gegenüber. Er wurzelt in der Theorie rationaler Erwartungen. Damit Prognosen sich als unverzerrt erweisen können, müssen mehrere Voraussetzungen erfüllt sein: 1. Das Geschehen in der realen Volkswirtschaft muss stabile Strukturen aufweisen. 2. Das Model von dieser Volkswirtschaft muss zutreffend und eindeutig sein. 3. Die Prognostiker müssen stets alle relevanten Informationen angemessen berücksichtigen.

In der euphorischen Anfangszeit der neoklassischen Theorie mag man noch bereit gewesen sein, solche Annahmen zumindest als näherungsweise zutreffend anzuerkennen. Nicht zuletzt die Banken- und Finanzmarktkrise 2008, die darauf folgende Wirtschaftskrise 2009, die darauf folgende Staatsschuldenkrise 2010 und die bereits drohende Währungskrise haben viele Wirtschaftswissenschaftler dazu gebracht, an den bisherigen Modellen zu zweifeln.¹ Die Behavioral Economics haben vielfältige Belege dafür gefunden, dass selbst hochbegabte Subjekte nicht

¹ Nur wenige Ökonomen, wie beispielsweise Thomas Sargent von der New York University oder Robert Barro von der Harvard University, erklärten auf der Jahrestagung der American Economic Association (AEA) 2010 in

in der Lage sind, immer alle relevanten Informationen angemessen auszuwerten. Die Transaktionskostentheorie zeigt, dass dies – selbst wenn es denn möglich wäre – nicht immer sinnvoll ist. Die Evolutionsökonomik hat tiefgreifende Zweifel geweckt, ob die Volkswirtschaften überhaupt dauerhafte, stabile Strukturen aufweisen. Vor diesem Hintergrund erscheint es uns vollkommen unangemessen, von den Prognostikern unverzerrte Prognosen zu verlangen.

Wir haben den Test auf Unverzerrtheit in der Mincer-Zarnowitz-Variante trotzdem in unserer Studie berücksichtigt. Wir wollen aufzuzeigen, dass dieser sehr verbreitete Ansatz zu ganz anderen Ergebnissen führt, als die von uns vorgeschlagenen Verfahren der Prognosebeurteilung.

Der Effizienztest von Holden und Peel (1990) baut unmittelbar auf der Mincer-Zarnowitz-Regression auf.

$$(2) \quad A_t = \alpha + \beta P_t + \gamma X_t + u_t$$

X_t ist eine beliebige zur Verfügung stehende Information. Basierend auf Gleichung (2) gelten die Prognosen als effizient, wenn sich α und γ nicht signifikant von 0 und β nicht signifikant von 1 unterscheiden. Auch hier darf der Fehlerterm nicht autokorreliert sein.

Wenn γ signifikant von Null unterschiedlich ist, dann wurde die Information X nicht angemessen in den Prognosen berücksichtigt. Bei diesem Ansatz greifen im Wesentlichen dieselben Kritikpunkte wie beim Test auf Unverzerrtheit.

Eine andere übliche Variante des Effizienztests betrachtet, ob die letzten tatsächlichen Werte vor Abgabe einer Prognose systematischen Einfluss auf den Prognosefehler haben (Simon, 1989). Ist dies der Fall, so stecken in den tatsächlichen Ereignissen Informationsbestandteile, die in den Prognosen unberücksichtigt bleiben. Seien A_t wieder das aktuelle Ereignis, P_t die entsprechende Prognose und u_t das Residual zum Zeitpunkt t . Des Weiteren erfolgt eine Berücksichtigung des Prognosehorizonts h .

$$(3) \quad A_t - P_t = \alpha + \sum_{i=1}^I \beta_i A_{t-h-i} + u_t$$

Atlanta ihr unbeirrtes Festhalten an den Grundlagen der Neoklassik wie zum Beispiel der Theorie rationaler Erwartungen (Storbeck, 2010).

Wenn die zur Verfügung stehenden Informationen effizient ausgewertet wurden, dürfen die Prognosefehler nicht signifikant mit den LAG-Variablen der tatsächlichen Ereignisse korreliert sein. Ob dies der Fall ist, wird mit Hilfe eines F-Tests ermittelt.

Diesen Ansatz beurteilen wir positiver. Die tatsächlichen Ereignisse der jüngsten Vergangenheit vor einer Prognoseabgabe müssen von einem Prognostiker zur Kenntnis genommen werden. Wenn sich ein systematischer Zusammenhang zwischen den zurückliegenden und den künftigen Ereignissen herstellen lässt, dann sollte man von einem Prognostiker erwarten, dass er diese Möglichkeit nutzt.

Der Vorzeichentest (Merton, 1981; Henriksson und Merton, 1981) ist ein weiteres weit verbreitetes Instrument zur Beurteilung von Prognosen. Bei diesem bekannten Verfahren wird nicht das Ausmaß einer prognostizierten Veränderung thematisiert. Es wird lediglich überprüft, ob die Richtung der Prognosen stimmt. Die Prognosen werden dann in eine 2×2 - Matrix eingeteilt. Dabei unterscheidet man einerseits, ob eine Wachstumsbeschleunigung oder eine Wachstumsverlangsamung vorhergesagt wurde. Andererseits unterscheidet man, ob eine Wachstumsbeschleunigung oder eine Wachstumsverlangsamung eingetreten ist. Die Hauptdiagonale in der 2×2 - Matrix kennzeichnet die in der Entwicklungstendenz zutreffenden Prognosen. Die Nebendiagonale kennzeichnet die in der Entwicklungstendenz unzutreffenden Prognosen. Mittels eines Chi-Quadrat-Tests wird nun überprüft, ob sich die Häufigkeitsverteilung in den 4 Feldern signifikant von einer Zufallsprognose unterscheidet (vgl. Diebold und Lopez, 1996; Joutz und Stekler, 2000). Sollte dies der Fall sein, so muss weiterhin geprüft werden, ob die Prognosen besser oder schlechter als die Zufallsprognose abschneiden.

Wir halten diesen Ansatz für adäquat. Er überprüft, ob die Prognostiker unwissend sind und lediglich raten oder ob sie zumindest einen groben Einblick in den prognostizierten Sachverhalt gewonnen haben. Hier werden also keine überzogenen Ansprüche an die Experten formuliert.

Der „Goldstandard“ der Prognosegütemessung ist nach wie vor der Vergleich zur naiven Prognose. Angenommen eine Blackbox produziert in regelmäßigen Abständen ein quantitatives Ereignis. Soll nun, in völliger Unkenntnis der in der Blackbox stattfindenden Prozesse, eine Prognose über den weiteren Verlauf der Zeitreihe abgegeben werden, so spricht eben so viel für einen steigenden wie für einen fallenden Verlauf in der Zukunft. Es erscheint also sinnvoll, von einer unveränderten Situation in der Zukunft auszugehen (naive Prognose). Der französische Mathematiker Pierre Simon Laplace (1814) hat diese Überlegung als „Prinzip des unzureichenden

Grundes“ in die Literatur eingeführt. Seitdem gilt die naive Prognose als absoluter Nullpunkt der Prognosequalität, da diese mühelos und ohne jede Kenntnis der prognostizierten Zusammenhänge erreichbar ist. Schon ein rudimentäres Verständnis der zu Grunde liegenden Prozesse müsste zu einer deutlich besseren Prognosegüte als bei der naiven Prognose führen.²

Es gibt vielfältige Möglichkeiten, um einen Vergleich zur naiven Prognose vorzunehmen. Zunächst lassen sich einfache Prognosegütemaße jeweils für die Prognosen der Experten und für die naiven Prognosen berechnen. Die gebräuchlichsten einfachen Prognosegütemaße sind der mittlere Prognosefehler (*MP*), Root Mean Squared Error (*RMSE*), der mittlere absolute Prognosefehler (*MAP*), Median Absolute Error (*MdAE*), Mean Absolute Percentage Error (*MAPE*), Median Absolute Percentage Error (*MdAPE*), Symmetric Mean Absolute Percentage Error (*sMAPE*), Symmetric Median Absolute Percentage Error (*sMdAPE*), Mean Relative Absolute Error (*MRAE*), Median Relative Absolute Error (*MdRAE*) und Geometric Mean Relative Absolute Error (*GMRAE*) (De Gooijer und Hyndman, 2006; Mathews und Diamantopoulos, 1994).

Etwas bequemer ist die Handhabung des Vergleichs zur naiven Prognose, wenn er bereits implizit in dem Prognosegütemaß vorgesehen ist. Theil's U ist ein bekanntes Beispiel dafür (Theil, 1971). Will man auf eine überproportionale Gewichtung von großen Prognosefehlern verzichten, dann bietet sich der mittlere absolute Prognosefehler bezogen auf die naive Prognose (*MRAPnP*) an.

$$(4) \quad MRAPnP = \frac{\sum_{t=1}^T |P_t - A_t|}{\sum_{t=1}^T |A_{t-h} - A_t|}$$

Dieser besteht im Zähler aus der Summe der absoluten Prognosefehler (P_t = Prognose zum Zeitpunkt t , A_t = tatsächliches Ereignis zum Zeitpunkt t). Diese wird durch die Summe der Prognosefehler der naiven Prognose geteilt, wobei A_{t-h} den tatsächlichen Wert zum Zeitpunkt $t-h$ darstellt (h =Prognosehorizont).

Fair und Shiller (1990) haben gezeigt, dass diese Ansätze in bestimmten Konstellationen zu fragwürdigen Ergebnissen führen können. Deshalb findet der Diebold Mariano Test for Forecast Encompassing (Diebold und Mariano, 1995) immer häufiger Verwendung. Ausgangspunkt

² Vgl. Scheier, J. / Spiwoks, M. (2006), S. 63.

ist hierbei die Vorhersage einer Situation P_m durch zwei konkurrierende Prognosemodelle i und j :

$$(5) \quad P_m = (1 - \lambda) P_{i,m} + \lambda P_{j,m}$$

wobei $0 \leq \lambda \leq 1$. Für $\lambda = 0$ spricht man von einer Umfassung der vom Modell j generierten Prognosen durch die Vorhersagen des Modells i . Das Modell j kann in diesem Fall nichts zu einer optimalen Kombinationsprognose aus beiden Modellen beitragen, da es über die in Modell i bereits enthaltenen Informationen hinaus keine weiteren nützlichen Zusatzinformationen beisteuert. Harvey, Leybourne und Newbold (1998) stellen eine Möglichkeit vor, die Nullhypothese $H_0: \lambda = 0$ gegen die Alternative $H_1: \lambda > 0$ zu testen. Bei einer Ablehnung der Nullhypothese enthalten beide Modelle unterschiedliche Informationen mit Zukunftsgehalt, so dass sie zur Bildung der optimalen Kombinationsprognose P_m herangezogen werden können. Betrachtet man nun die naiven Prognosen als Modell i und die Expertenprognosen als Modell j , so kann ermittelt werden, ob der Informationsgehalt der Expertenprognosen signifikant über den Informationsgehalt der naiven Prognosen hinausgeht.

Wir halten den Vergleich zur naiven Prognose für sinnvoll. Prognoseverfahren, die nicht zu besseren Ergebnissen führen als die entsprechenden naiven Prognosen, sind zweifelsfrei untauglich. Der Diebold Mariano Test for Forecast Encompassing bietet eine besonders elegante Möglichkeit, um einen Vergleich zur naiven Prognose vorzunehmen. Allerdings haben wir Zweifel, ob es vernünftig ist, immer und überall nur statistisch signifikant bessere Ergebnisse als Erfolg anzuerkennen. Beispielsweise fragt bei der Olympiade niemand danach, ob der Goldmedaillengewinner beim 100-Meter-Lauf statistisch signifikant schneller laufen kann als der Zweitplatzierte.³ Wenn es darum geht, keine übertriebenen Anforderungen an die Prognostiker zu stellen, so kommen Vergleiche auf der Basis einfacher Prognosegütemaße nach wie vor in Betracht. Wir verwenden den Mittleren absoluten Prognosefehler bezogen auf die naive Prognose (*MRAPnP*), weil es sich dabei um ein einfaches, übersichtliches sowie ein leicht und eindeutig zu interpretierendes Prognosegütemaß handelt.

³ Es wäre ja ohne weiteres möglich, auch bei der Beurteilung sportlicher Leistungen statistische Signifikanztests zu etablieren: Die Sprinter müssten 30 Mal gegeneinander antreten. Aus den durchschnittlichen Zeiten und der Streuung der Einzelergebnisse ließe sich dann ermitteln, ob der Erste signifikant schneller laufen kann als der Zweitplatzierte. Eine solche Vorgehensweise dürfte den meisten Sportbegeisterten jedoch als weltfremd und unsinnig erscheinen.

Allerdings bietet es sich nicht an, dieses Prognosegütemaß für alle 24 Prognosehorizonte zu ermitteln. Die tatsächlichen Ereignisse treten nicht monatlich, sondern nur jahresweise auf. Dementsprechend gibt es pro Jahr auch nur eine naive Prognose, die als Vergleichsmaßstab beim *MRAPnP* verwendet werden kann. Die Prognosen, die im Januar abgegeben werden, können sinnvoll mit den naiven Prognosen (tatsächliches Ereignis des Vorjahres) verglichen werden. Deshalb werten wir die Prognosen mit 24 und mit 12 Monaten Prognosehorizont aus. Hier könnte allerdings eingewendet werden, dass Anfang Januar die Daten zum tatsächlichen Wirtschaftswachstum des Vorjahres noch nicht vorliegen. Üblicherweise dauert es bis März, bis die ersten weitgehend verlässlichen Zahlen vorliegen. Deshalb werten wir außerdem die Prognosen, die Anfang April abgegeben werden, mit dem *MRAPnP* aus. Das betrifft somit die Prognosen mit 21 und mit 9 Monaten Prognosehorizont. Da wir überzogene Ansprüche an die Prognostiker vermeiden wollen, erscheint die Betrachtung der Prognosen von Anfang April sogar sinnvoller. Denn die kostengünstige Alternative, sich stets auf die naiven Prognosen zu stützen, steht schließlich erst ab Anfang April zur Verfügung.

Je weiter (näher) nicht-deterministische Ereignisse in der Zukunft liegen, desto schwieriger (leichter) sind sie in aller Regel zu prognostizieren. Das ist darauf zurückzuführen, dass im Zeitverlauf Ereignisse zu beobachten sind, die zu einer besseren Beurteilung des Sachverhaltes beitragen können (Poulizac, Weale und Young, 1996; Ackert und Hunter, 1994, S. 390-391; Döpke und Fritsche, 2006, S. 789). Den Ausgang eines Fußballspiels kann man in der 80. Spielminute weitaus besser prognostizieren als bei Anpfiff des Spiels. Hat eine Fußballmannschaft bis zur 80. Spielminute zwei Spieler durch rote Karten verloren und liegt zugleich mit 0:6 Toren zurück, dann ist es nicht mehr schwierig, diese Mannschaft als Verlierer zu prognostizieren. Müsste man den Ausgang von 100 Fußballspielen jeweils zu Spielbeginn prognostizieren, so würde man gewiss viel häufiger falsch liegen, als wenn man den Ausgang von 100 Fußballspielen jeweils in der 80. Minute prognostizieren würde. Voraussetzung dafür ist nur, dass man über einige rudimentäre Grundkenntnisse des Fußballs verfügt und die relevanten Ereignisse (Tore, Platzverweise, etc.) bis zur 80. Spielminute zur Kenntnis nimmt.

Von Wirtschaftsexperten sollte man das gleiche verlangen können. Sie sollten zumindest über rudimentäre Grundkenntnisse wirtschaftlicher Prozesse verfügen⁴ und zumindest die wichtigsten der relevanten Ereignisse zur Kenntnis nehmen. Sind diese beiden Voraussetzungen erfüllt,

⁴ Mehr ist bei dem aktuellen Forschungsstand der Wirtschaftswissenschaften vielleicht auch gar nicht möglich.

dann sollten die Prognosen im langjährigen Durchschnitt umso genauer sein, je kürzer der Prognosehorizont ist.

Man könnte vermuten, dass Wirtschaftsprognostiker diese Anforderung üblicherweise ohne jede Anstrengung und ohne jede Schwierigkeit bewältigen. Tatsächlich zeigt sich aber, dass beispielsweise Finanzmarktanalysten diesem Anforderungsprofil keineswegs immer entsprechen. So zeigt Spiwoks (2009), dass der Prognosefehler bei der Vorhersage der Gewinnentwicklung US-amerikanischer Unternehmen mit abnehmendem Prognosehorizont nicht zurückgeht. Die Prognosen mit einem Prognosehorizont von nur einem Monat sind im Durchschnitt schlechter als die Prognosen mit einem Prognosehorizont von 24 Monaten. Dies ist besonders überraschend, weil die Unternehmen quartalsweise ihre Gewinne veröffentlichen, so dass bei kurzen Prognosehorizonten bereits wesentliche Bestandteile des zu prognostizierenden Ereignisses bekannt sind. Offensichtlich sind die Analysten aber nicht bereit, diese Zwischenergebnisse in ihren Prognosen zu berücksichtigen.

Wir schlagen daher einen Test auf Informationszuwachs vor. Dieser neue und zugleich sehr einfache Ansatz überprüft, ob die Prognostiker bereit und in der Lage sind, relevante Ereignisse während eines sich allmählich verkürzenden Prognosehorizontes zur Kenntnis zu nehmen. Es wird also überprüft, ob der Prognosefehler mit abnehmendem Prognosehorizont statistisch signifikant zurückgeht.

Die Prognosefehler werden als absoluter Prognosefehler (*AP*) oder als absoluter prozentualer Prognosefehler (*APP*) für jeden Prognosehorizont berechnet. Anschließend wird der Korrelationskoeffizient zwischen Prognosehorizont und Prognosefehler geschätzt. Weist der Korrelationskoeffizient statistisch signifikant ein positives Vorzeichen auf, so verringert sich der Prognosefehler erkennbar mit abnehmendem Prognosehorizont. Eine wichtige Mindestanforderung an die Prognostiker ist in diesem Fall als erfüllt anzusehen.

Der Test auf Informationszuwachs kann der Gruppe der Effizienztests zugeordnet werden. Denn es wird überprüft, ob relevante Informationen angemessene Berücksichtigung in den Prognosen finden.

Wenn Prognosen stärker mit den tatsächlichen Werten zum Zeitpunkt ihrer Erstellung als mit den Entwicklungen zu ihren Gültigkeitszeitpunkten korreliert sind, spricht man von gegenwartsorientierter Verlaufsanpassung (*GOVA*). Hierbei erfolgt eine fortlaufende Anpassung der Prognosen an das aktuelle Marktgeschehen, so dass ihr zukunftsorientierter Charakter unter

Umständen vollständig verloren geht. Mit Hilfe des GOVA-Koeffizienten lässt sich eine etwaige gegenwartsorientierte Verlaufsanpassung feststellen.

Hierzu wird zunächst das Bestimmtheitsmaß (R^2_A) zwischen den Prognosen und den tatsächlich eingetretenen Werten berechnet (siehe Andres und Spiwoks, 1999; sowie Bofinger und Schmidt, 2003). Anschließend erfolgt die Ermittlung des Bestimmtheitsmaßes zwischen den Prognosen und den Werten des Prognosegegenstands zum Zeitpunkt der Prognoseerstellung (R^2_B). Die Division der beiden Bestimmtheitsmaße ergibt nun den GOVA-Koeffizienten.

$$(6) \quad \text{GOVA-Koeffizient} = \frac{R^2_{\text{Prognosen (Geltungszeitpunkte); tatsächliche Ereignisse}}}{R^2_{\text{Prognosen (Entstehungszeitpunkte); tatsächliche Ereignisse}}} = \frac{R^2_A}{R^2_B}$$

Bei einem GOVA-Koeffizienten < 1 liegt gegenwartsorientierte Verlaufsanpassung vor. In diesem Fall haben die Prognosen den Prognosegegenstand zum Zeitpunkt ihrer Erstellung besser erfasst als zum Zeitpunkt ihrer eigentlichen Gültigkeit. Sie spiegeln die Vergangenheit stärker als die Zukunft wider.

Bei Finanzmarktprognosen treten gegenwartsorientierte Verlaufsanpassungen regelmäßig auf. Spiwoks, Gubaydullina und Hein (2011) untersuchen beispielsweise rund 1.200 Zinsprognosezeitreihen mit rund 160.000 einzelnen Prognosen. Es zeigt sich, dass 98,5% aller Prognosezeitreihen von einer gegenwartsorientierten Verlaufsanpassung geprägt sind. Finanzmarktanalysten können oder wollen sich offenbar nicht von der jeweils aktuell vorherrschenden Marktsituation lösen. Sie „prognostizieren“ also vor allem die Gegenwart, aber kaum die Zukunft. Schuh (2001, S. 42), Mehra (2002, S. 21-22), Mankiw, Reis und Wolfers (2003, S. 212), Bowles et al. (2007, S. 18), Andolfatto, Hendry und Moran (2008, S. 407) und Dovern, Fritsche und Slacalek (2009, S. 38-43) geben erste Hinweise darauf, dass gegenwartsorientierte Verlaufsanpassungen durchaus auch bei makroökonomischen Prognosen auftreten können.

Der GOVA-Koeffizient wurde jedoch bisher noch nie zur Analyse makroökonomischer Prognosen angewendet. Das ist sicherlich darauf zurückzuführen, dass der GOVA-Koeffizient nur für die Analyse von Prognosen mit konstantem Prognosehorizont geeignet ist. Wählt man einzelne Prognosehorizonte aus den vorhandenen 24 Prognosehorizonten aus, so lässt sich der GOVA-Koeffizient jedoch auch auf makroökonomische Prognosezeitreihen anwenden. Wir wenden ihn auf die Prognosen mit Prognosehorizonten von zwölf und neun Monaten an. Diese beiden Prognosehorizonte erscheinen besonders geeignet. Im Januar dürfte das abgelaufene Jahr bei den Prognostikern besonders präsent sein, so dass sich möglicherweise eine Neigung

zu gegenwartsorientierten Verlaufsanpassungen einstellt. Anfang April sind in aller Regel bereits recht verlässliche vorläufige Daten zum Wirtschaftswachstum des Vorjahres bekannt. Vielleicht verführen gerade diese Daten zu gegenwartsorientierten Verlaufsanpassungen.

Wir analysieren die Konjunkturprognosen also mit vier der vorgestellten Verfahren. Den mittleren absoluten Prognosefehler bezogen auf die naive Prognose (*MRAPnP*), den Test auf Informationszuwachs und den GOVA-Koeffizienten wenden wir an, weil wir diese Instrumente als besonders geeignet betrachten. Den Test auf Unverzerrtheit wenden wir an, um zu zeigen, zu welcher gravierenden Fehleinschätzungen ein ungeeigneter Ansatz zur Beurteilung von Prognosen führen kann.

4 Ergebnisse

Der Vergleich zur naiven Prognose auf der Basis des mittleren absoluten Prognosefehlers bezogen auf die naive Prognose (*MRAPnP*) zeigt, dass die Konjunkturprognostiker bis auf wenige Ausnahmen erfolgreich sind (Tabelle 2). Bei den Prognosehorizonten 24 und 12 Monate sind die Prognosen jeweils zu 33 der 36 Prognosegegenstände (91,7%) genauer als die Alternative der naiven Prognosen.⁵ Beim Prognosehorizont von 21 Monaten weisen die Prognosen der Experten in 34 von 36 Fällen (94,4%) geringere Prognosefehler auf als die entsprechenden naiven Prognosen. Beim kürzesten der betrachteten Prognosehorizonte (neun Monate) sind die Prognosen für alle 36 Prognosegegenstände der naiven Prognose überlegen.

Die naive Prognose stellt den absoluten Nullpunkt der Prognosegüte dar (Fildes und Stekler, 2002, S. 439). Ein Prognostiker muss Prognosen abgeben, die erkennbar besser als die entsprechenden naiven Prognosen sind, um seine Tätigkeit überhaupt rechtfertigen zu können. Denn die naive Prognose steht als kostenlose Alternative zur Verfügung. Insofern scheint es keinen Grund zur Euphorie zu geben, wenn die Konjunkturprognostiker in 136 von 144 Fällen (94,4%) besser als die entsprechenden naiven Prognosen abschneiden.

Um den Erfolg der Konjunkturprognostiker angemessen würdigen zu können, sollte man sich die Erfolgsquoten von Finanzmarktprognostikern in Erinnerung rufen. In vielen Studien zeigt sich,

⁵ Osterloh (2008) verwendet Theil's U als Beurteilungsmaßstab und kommt für die deutschen BIP-Prognosen mit 24 Monaten Prognosehorizont zu einem ungünstigeren Befund. Die unterschiedlichen Ergebnisse können auf die Auswahl des Beurteilungsmaßstabes zurückgehen. Sie können aber auch in dem unterschiedlichen Betrachtungszeitraum begründet sein. Osterloh wertet Prognosedaten von 1995 bis 2005 aus.

dass zwischen 80% und 100% der Zeitreihen von Finanzmarktprognosen schlechter als die entsprechenden naiven Prognosen abschneiden (vgl. zu Aktienprognosen: Lakonishok, 1980; Fraser und MacDonald, 1993; Spiwoks, 2004; vgl. zu Zinsprognosen: Brooks und Gray, 2004; Mose, 2005; Spiwoks, Bedke und Hein, 2008; Spiwoks, Bedke und Hein, 2009; Spiwoks, Bedke und Hein, 2010; vgl. zu Wechselkursprognosen: Manzur, 1988; Chinn und Frankel, 1994; Bofinger und Schmidt, 2003). Insofern erscheinen die betrachteten Konjunkturprognosen durchaus als sehr erfolgreich.

Tabelle 2: Ergebnisse des Vergleichs mit der naiven Prognose auf der Basis des mittleren absoluten Prognosefehlers bezogen auf die naive Prognose (MRAPnP) für ausgewählte Prognosehorizonte

Land	Prognose-gegenstand	MRAPnP 24 Mon.	MRAPnP 21 Mon.	MRAPnP 12 Mon.	MRAPnP 9 Mon.
USA	BIP	0,972	0,941	0,793	0,639
	Industrieprod.	0,989	0,966	0,808	0,645
	privater Konsum	0,914	0,827	1,012	0,741
Japan	BIP	0,825	0,780	0,736	0,551
	Industrieprod.	0,741	0,726	0,643	0,432
	privater Konsum	0,932	0,836	0,878	0,633
Deutschland	BIP	0,813	0,772	0,670	0,519
	Industrieprod.	0,750	0,739	0,620	0,415
	privater Konsum	0,810	0,871	0,952	0,885
Frankreich	BIP	0,785	0,779	0,751	0,502
	Industrieprod.	0,879	0,859	0,825	0,623
	privater Konsum	0,836	0,860	1,074	0,914
Großbritannien	BIP	0,798	0,769	0,809	0,640
	Industrieprod.	0,953	0,927	0,708	0,485
	privater Konsum	0,879	0,820	0,800	0,595
Italien	BIP	0,891	0,879	0,728	0,561
	Industrieprod.	0,951	0,968	0,867	0,612
	privater Konsum	0,943	0,892	0,888	0,746
Spanien	BIP	0,914	0,838	0,931	0,822
	Industrieprod.	0,846	0,831	0,852	0,555
	privater Konsum	0,744	0,726	0,983	0,813
Kanada	BIP	0,789	0,794	0,774	0,535
	Industrieprod.	1,087	1,084	1,217	0,877
	privater Konsum	0,737	0,749	0,800	0,669
Niederlande	BIP	0,875	0,853	0,883	0,730
	Industrieprod.	0,788	0,797	0,767	0,568
	privater Konsum	1,010	0,978	0,970	0,932
Schweiz	BIP	0,773	0,773	0,713	0,653
	Industrieprod.	0,647	0,653	0,625	0,561
	privater Konsum	0,702	0,685	0,815	0,884
Schweden	BIP	0,823	0,827	0,855	0,673
	Industrieprod.	0,754	0,781	0,690	0,511
	privater Konsum	0,854	0,871	0,859	0,735
Norwegen	BIP	0,836	0,842	0,756	0,714
	Industrieprod.	1,218	1,182	0,497	0,525
	privater Konsum	0,809	0,767	0,742	0,689

MRAPnP = mittlerer absoluter Prognosefehler bezogen auf die naive Prognose.

Der Test auf Informationszuwachs vermittelt ebenfalls ein recht positives Bild (Tabelle 3). Betrachtet man die Korrelationskoeffizienten zwischen den Prognosehorizonten und den absoluten Prognosefehlern (*AP*), so zeigt sich bei 35 von 36 Prognosegegenständen (97,2%) ein positives Vorzeichen. Bei 33 dieser 35 Fälle (94,3%) erweist sich dieses Ergebnis bei einer Irrtumswahrscheinlichkeit von 1% als statistisch signifikant. In den übrigen beiden Fällen zeigt sich eine statistische Signifikanz bei einer Irrtumswahrscheinlichkeit von 5%. Lediglich bei den

Prognosen zur Industrieproduktion in Norwegen zeigt sich ein negatives Vorzeichen. Allerdings ist dieses Ergebnis statistisch nicht signifikant.

Tabelle 3: Ergebnisse des Tests auf Informationszuwachs für ausgewählte Prognosegütemaße und GOVA-Koeffizienten für ausgewählte Prognosehorizonte.

Land	Prognosegegenstand	Korr. PH - AP	Korr. PH - APP	GOVA 12 Mon.	GOVA 9 Mon.
USA	BIP	0,408***	0,215***	1,562	2,102
	Industrieproduktion	0,342***	0,233***	1,606	2,072
	privater Konsum	0,375***	0,177***	1,029	1,432
Japan	BIP	0,312***	0,182***	0,627	1,585
	Industrieproduktion	0,368***	0,272***	1,168	4,904
	privater Konsum	0,233***	0,201***	0,507	1,011
Deutschland	BIP	0,386***	0,127***	3,032	6,248
	Industrieproduktion	0,371***	0,083***	3,099	6,329
	privater Konsum	0,187***	0,227***	1,439	1,957
Frankreich	BIP	0,415***	0,122***	1,178	2,225
	Industrieproduktion	0,335***	0,276***	1,643	2,617
	privater Konsum	0,283***	0,164***	0,760	1,592
Großbritannien	BIP	0,212***	0,089**	1,704	1,860
	Industrieproduktion	0,390***	0,234***	2,680	2,765
	privater Konsum	0,287***	0,205***	1,624	1,900
Italien	BIP	0,417***	0,155***	1,413	1,692
	Industrieproduktion	0,370***	0,251***	1,598	2,969
	privater Konsum	0,280***	0,223***	0,836	1,285
Spanien	BIP	0,281***	0,208***	1,309	1,405
	Industrieproduktion	0,352***	0,257***	0,775	1,104
	privater Konsum	0,250***	0,144**	1,166	1,273
Kanada	BIP	0,425***	0,314***	2,394	2,888
	Industrieproduktion	0,334***	0,156***	3,004	3,403
	privater Konsum	0,283***	0,147***	1,347	1,882
Niederlande	BIP	0,330***	0,171***	1,352	2,033
	Industrieproduktion	0,285***	0,154***	2,749	3,323
	privater Konsum	0,420***	0,218***	1,409	1,295
Schweiz	BIP	0,500***	0,267***	4,501	9,678
	Industrieproduktion	0,366***	0,163**	2,476	2,858
	privater Konsum	0,160**	0,054	0,816	0,643
Schweden	BIP	0,310***	0,133**	0,840	1,229
	Industrieproduktion	0,242***	0,163***	1,573	2,634
	privater Konsum	0,325***	0,077	0,268	0,909
Norwegen	BIP	0,141**	0,102*	1,004	2,524
	Industrieproduktion	-0,043	0,111	1,739	3,372
	privater Konsum	0,391***	0,192**	1,138	1,438

Korr. = Korrelationskoeffizient; PH = Prognosehorizont; AP = absoluter Prognosefehler; APP = absoluter prozentualer Prognosefehler; GOVA = GOVA-Koeffizient; Signifikanzniveau: 1%***, 5%** und 10%*.

Tabelle 4: Ergebnisse des Tests auf Unverzerrtheit in der Mincer-Zarnowitz-Variante

Land	Prognosegegenstand	α	β	F-Test p-Wert	DW p-Wert
USA	BIP	0,264	0,890	0,001 ¹	0,000
	Industrieproduktion	-0,402	0,937	0,000	0,000
	privater Konsum	0,272	1,029	0,000 ¹	0,000
Japan	BIP	-0,059	0,650	0,000	0,000
	Industrieproduktion	-1,627	0,803	0,000	0,000
	privater Konsum	0,347	0,484	0,000 ¹	0,000
Deutschland	BIP	-0,064	0,879	0,000	0,000
	Industrieproduktion	-1,101	0,973	0,000	0,000
	privater Konsum	0,453	0,717	0,000 ¹	0,000
Frankreich	BIP	-0,230	0,915	0,000	0,000
	Industrieproduktion	-1,759	0,977	0,000	0,000
	privater Konsum	0,010	0,914	0,001 ¹	0,000
Großbritannien	BIP	-0,259	1,143	0,000 ¹	0,000
	Industrieproduktion	-1,151	0,907	0,000	0,000
	privater Konsum	-0,382	1,249	0,000 ¹	0,000
Italien	BIP	-0,671	1,012	0,000	0,000
	Industrieproduktion	-1,765	1,047	0,000	0,000
	privater Konsum	-0,197	0,832	0,000	0,000
Spanien	BIP	-0,475	1,300	0,000 ¹	0,000
	Industrieproduktion	-2,044	1,028	0,000 ¹	0,000
	privater Konsum	-0,984	1,410	0,000 ¹	0,000
Kanada	BIP	-0,699	1,172	0,000	0,000
	Industrieproduktion	-0,648	0,654	0,000 ¹	0,000
	privater Konsum	0,143	1,010	0,002 ¹	0,000
Niederlande	BIP	0,089	1,045	0,081	0,000
	Industrieproduktion	-0,443	0,940	0,000	0,000
	privater Konsum	0,178	0,984	0,179 ¹	0,000
Schweiz	BIP	0,133	1,014	0,191	0,000
	Industrieproduktion	-0,816	1,065	0,062	0,000
	privater Konsum	0,693	0,581	0,000	0,000
Schweden	BIP	-0,387	1,153	0,125	0,000
	Industrieproduktion	-3,188	1,228	0,000 ¹	0,000
	privater Konsum	0,801	0,688	0,000 ¹	0,000
Norwegen	BIP	0,702	0,419	0,000 ¹	0,000
	Industrieproduktion	-1,535	0,396	0,000	0,000
	privater Konsum	0,465	1,005	0,000	0,000

¹ = Heteroskedastizität kann nicht ausgeschlossen werden, so dass der p-Wert mit robusten Standardfehlern ermittelt wurde; die p-Werte, die sich durch Schätzung mit robusten Standardfehlern geändert haben, sind fett hervorgehoben; DW = Durbin-Watson-Test.

Tabelle 5: Gegenüberstellung von Prognoseerfolgen (+) und Prognosemisserfolgen (-) nach ausgewählten Beurteilungsmaßstäben

Land	Prognosegegenstand	MRAPnP 9 Mon.	Korrel. PH-AP	GOVA 9 Mon.	Test auf Un- verzerrtheit
USA	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Japan	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Deutschland	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Frankreich	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Großbritannien	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Italien	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Spanien	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Kanada	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Niederlande	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	+	-
Schweiz	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	-	-
Schweden	BIP	+	+	+	-
	Industrieproduktion	+	+	+	-
	privater Konsum	+	+	-	-
Norwegen	BIP	+	+	+	-
	Industrieproduktion	+	-	+	-
	privater Konsum	+	+	+	-

+ = Prognoseerfolg; - = Prognosemisserfolg; PH = Prognosehorizont; AP = absoluter Prognosefehler; MRAPnP = mittlerer absoluter Prognosefehler bezogen auf die naive Prognose; GOVA = GOVA-Koeffizient.

Etwas weniger zwingend stellt sich dieser Zusammenhang dar, wenn man die Korrelationskoeffizienten zwischen den Prognosehorizonten und den absoluten prozentualen Prognosefehlern (APP) betrachtet (Tabelle 3). Zwar weisen alle 36 Korrelationskoeffizienten ein positives Vorzeichen auf, aber nur 27 davon (75,0%) sind bei einer Irrtumswahrscheinlichkeit von 1% als statistisch signifikant zu betrachten. Bei einer Irrtumswahrscheinlichkeit von 5% erweisen sich

immerhin 32 der 36 positiven Korrelationskoeffizienten (88,9%) als statistisch signifikant. Bei einer Irrtumswahrscheinlichkeit von 10% sind es 33 der 36 Korrelationskoeffizienten (91,7%).

Der Test auf Informationszuwachs zeigt somit nur in 4 von 72 Fällen (5,6%) keinen statistisch signifikanten Zusammenhang. In 68 von 72 Fällen (94,4%) zeigt sich hingegen, dass der Prognosefehler umso geringer ausfällt, je kürzer der Prognosehorizont ist. Die überwiegende Mehrheit der Konjunkturprognostiker ist also offenbar bereit und in der Lage, im Zeitverlauf relevante Informationen aufzunehmen und sinnvoll in ihre Prognosen einzubeziehen.

Es mag auf den ersten Blick wenig beeindruckend wirken, dass Konjunkturprognostiker ihre Prognosen nicht abgeben, ohne im Laufe der Zeit relevante Fakten zur Kenntnis zu nehmen. Doch zeigt sich beispielsweise bei Finanzmarktprognostikern, dass dies keineswegs als selbstverständlich zu betrachten ist.⁶

Der GOVA-Koeffizient vermittelt ebenso wie der Vergleich zur naiven Prognose und der Test auf Informationszuwachs ein verhältnismäßig positives Bild von den Leistungen der Konjunkturprognostiker (Tabelle 3). Bei 28 der 36 Prognosezeitreihen mit 12 Monaten Prognosehorizont (77,8%) und bei 34 von 36 Prognosezeitreihen mit 9 Monaten Prognosehorizont (94,4%) zeigen sich keine gegenwartsorientierte Verlaufsanpassungen. Das bedeutet, dass die weit überwiegende Mehrzahl der Prognosezeitreihen stärker die Zukunft als die Gegenwart reflektiert. Lediglich 10 von 72 Prognosezeitreihen (13,9%) weisen einen GOVA-Koeffizienten von > 1 auf. Diese Zeitreihen korrespondieren stärker mit den tatsächlichen Ereignissen zu den jeweiligen Zeitpunkten der Prognoseabgabe als mit den tatsächlichen Ereignissen zu den jeweiligen Zeitpunkten der Prognosegeltung. Sie sind also von gegenwartsorientierten Verlaufsanpassungen geprägt.

Prognosen sollten auf die Zukunft gerichtet sein. Prognosen, die eher die Gegenwart als die Zukunft widerspiegeln verfehlen im Prinzip ihren Zweck. Insofern sollte man erwarten, dass gegenwartsorientierte Verlaufsanpassungen im ökonomischen Prognosewesen krasse Ausnahmen darstellen. Jedoch zeigt sich, dass bei Finanzmarktprognosen gegenwartsorientierte Verlaufsanpassungen die Regel und nicht die Ausnahme sind. Eine ganze Reihe von Studien gibt zu erkennen, dass mehr als 98% aller Finanzmarktprognosezeitreihen von gegenwartsorientierten Verlaufsanpassungen geprägt sind (Bofinger und Schmidt, 2003; Spiwoks, 2004; Spiwoks,

⁶ Vgl. beispielsweise Spiwoks (2009).

Bedke und Hein, 2008; Spiwox, Bedke und Hein, 2009; Spiwox, Bedke und Hein, 2010; Spiwox, Gubaydullina und Hein, 2011). Insofern muss es als Erfolg der Konjunkturprognostiker gewertet werden, dass nur ein kleiner Teil der Prognosezeitreihen von gegenwartsorientierten Verlaufsanpassungen geprägt ist.

Der Test auf Unverzerrtheit führt zu einer vollkommen anderen Bewertung des Prognoseerfolgs (Tabelle 4). Er vermittelt den Eindruck, dass die Anstrengungen der Prognostiker vollkommen gescheitert sind. Bei 31 der 36 betrachteten Prognosegegenstände (86,1%) zeigt der F-Test bei einer Irrtumswahrscheinlichkeit von 1%, dass $\alpha \neq 0$ und / oder $\beta \neq 1$ ist. Bei einer Irrtumswahrscheinlichkeit von 10% trifft dies sogar auf 33 von 36 Prognosegegenständen (91,7%) zu.⁷

Noch ungünstiger stellt sich die Situation dar, wenn man mit Hilfe des Durbin-Watson-Tests die Residuale auf systematische Komponenten hin untersucht. In allen 36 Fällen liegen bei einer Irrtumswahrscheinlichkeit von deutlich unter 1% Autokorrelationen vor.

Der Test auf Unverzerrtheit kennzeichnet somit die Prognosen zu allen 36 Prognosegegenständen als verzerrt.⁸ Die Prognoseanstrengungen sind, sofern man diesen Beurteilungsmaßstab als adäquat betrachtet, vollständig gescheitert. Keane und Runkle (1990) sowie Bonham und Cohen (2001) weisen darauf hin, dass Konsensprognosen als unverzerrt erscheinen können, obwohl die darin enthaltenen individuellen Prognosen durchaus verzerrt sind. Selbst dieser Effekt hat bei der vorliegenden Datenanalyse zu keinem günstigeren Ergebnis geführt.

Viele der betrachteten Prognosegegenstände genügen – abgesehen vom Test auf Unverzerrtheit – allen acht Beurteilungsmaßstäben (Tabellen 2 und 3):

Die Prognosen zum BIP-Wachstum für die USA, Deutschland, Frankreich, Italien, Spanien, Kanada, die Niederlande und die Schweiz (a) erweisen sich für alle vier Prognosehorizonte (24,

⁷ Der Breusch-Pagan-Test (Breusch and Pagan, 1979) zeigt, dass bei 16 der 36 Datenbestände Heteroskedastizität angenommen werden muss. Bei diesen 16 haben wir den F-Test mit robusten Standardfehlern durchgeführt. Dies führte allerdings – bei Berücksichtigung von drei Nachkommastellen – nur in vier Fällen (fett hervorgehoben) zu einer Änderung des p-Wertes.

⁸ Ager, Kappler and Osterloh (2007) untersuchen unter anderem BIP-Prognosen für die meisten der hier betrachteten Länder im Zeitraum von 1996 bis 2006. Allerdings stellen sie in der Mehrzahl der Fälle fest, dass die Prognosen unverzerrt sind (S. 13). Dovern und Weisser (2009) untersuchen unter anderem BIP-Prognosen für die G7-Länder im Zeitraum von 1991 bis 2005. Auch sie kommen zu dem Ergebnis, dass die meisten Prognosen – insbesondere in Phasen ohne gravierende Strukturbrüche – unverzerrt sind (S. 20). Diese abweichenden Ergebnisse sind möglicherweise auf die kürzeren Untersuchungszeiträume, auf die abweichende Anzahl von Prognosen oder auf andere Zuschnitte des Tests auf Unverzerrtheit zurückzuführen. Batchelor (2007) untersucht unter anderem BIP-Prognosen für den Zeitraum von 1990 bis 2005. Für Japan, Italien, Deutschland und Frankreich erweisen sich die BIP-Prognosen als verzerrt (S. 20).

21, 12 und 9 Monate) als den naiven Prognosen überlegen, (b) lassen sowohl anhand des absoluten Prognosefehlers wie auch des absoluten prozentualen Prognosefehlers bei einer Irrtumswahrscheinlichkeit von 1% einen Informationszuwachs im Zeitverlauf erkennen und (c) weisen bei keinem der beiden Prognosehorizonte (12 und 9 Monate) gegenwartsorientierte Verlaufsanpassungen auf.

Die Prognosen zum Wachstum der Industrieproduktion für die USA, Japan, Deutschland, Frankreich, Großbritannien, Italien und die Niederlande (a) erweisen sich für alle vier Prognosehorizonte (24, 21, 12 und 9 Monate) als den naiven Prognosen überlegen, (b) lassen sowohl anhand des absoluten Prognosefehlers wie auch des absoluten prozentualen Prognosefehlers bei einer Irrtumswahrscheinlichkeit von 1% einen Informationszuwachs im Zeitverlauf erkennen und (c) weisen bei keinem der beiden Prognosehorizonte (12 und 9 Monate) gegenwartsorientierte Verlaufsanpassungen auf.

Die Prognosen zum Wachstum des privaten Konsums für Deutschland, Großbritannien und Kanada (a) erweisen sich für alle vier Prognosehorizonte (24, 21, 12 und 9 Monate) als den naiven Prognosen überlegen, (b) lassen sowohl anhand des absoluten Prognosefehlers wie auch des absoluten prozentualen Prognosefehlers bei einer Irrtumswahrscheinlichkeit von 1% einen Informationszuwachs im Zeitverlauf erkennen und (c) weisen bei keinem der beiden Prognosehorizonte (12 und 9 Monate) gegenwartsorientierte Verlaufsanpassungen auf.

Bei 33 der 36 Prognosegegenstände (91,7%) werden folgende Anforderungen gleichzeitig erfüllt: (a) bei einem Prognosehorizont von neun Monaten werden die naiven Prognosen übertroffen, (b) bei Betrachtung des absoluten Prognosefehlers kann ein Informationszuwachs statistisch signifikant festgestellt werden und (c) bei einem Prognosehorizont von neun Monaten treten keine gegenwartsorientierte Verlaufsanpassungen auf (Tabelle 5). Lediglich die Prognosen zum Wachstum des privaten Konsums in der Schweiz und in Schweden sowie die Prognosen zum Wachstum der Industrieproduktion in Norwegen können diese drei Kriterien nicht vollständig erfüllen.

In Tabelle 5 zeigt sich auch die starke Diskrepanz zwischen den drei von uns bevorzugten Beurteilungsmaßstäben einerseits und dem Test auf Unverzerrtheit andererseits.

Wir stellen abschließend fest, dass Konjunkturprognostiker wesentlich erfolgreicher sind als Finanzmarktanalysten. Dieser gravierende Unterschied wird vom Test auf Unverzerrtheit allerdings nivelliert und verschleiert. Wir halten den Test auf Unverzerrtheit deshalb für weitgehend ungeeignet, um ökonomische Prognosen zu beurteilen.

5 Fazit

Die vorliegende Studie setzt sich mit der Beurteilung von Konjunkturprognosen auseinander. Neben der Erörterung herkömmlicher Beurteilungsmaßstäbe werden auch zwei neue Verfahren vorgestellt: Der Test auf Informationszuwachs und der GOVA-Koeffizient.

Der Test auf Informationszuwachs betrachtet, ob die Prognosen bei allmählich rückläufigem Prognosehorizont genauer werden. Es wird also untersucht, ob die Prognostiker bereit und in der Lage sind, während des Prognosezeitraums relevante neue Informationen aufzunehmen und in ihren Prognosen zu berücksichtigen. Der GOVA-Koeffizient zeigt, ob die Prognosen stärker die Zukunft oder die Gegenwart reflektieren.

Wir untersuchen Konsensprognosen zur Entwicklung des BIP, der Industrieproduktion und der privaten Konsumausgaben in den USA, Japan, Deutschland, Frankreich, Großbritannien, Italien, Spanien, Kanada, Niederlande, Schweiz, Schweden und Norwegen.

Dabei stützen wir uns auf vier ausgewählte Beurteilungsverfahren: 1. Vergleich zur naiven Prognose mit Hilfe des mittleren absoluten Prognosefehlers bezogen auf die naive Prognose (*MRAPnP*), 2. Test auf Informationszuwachs, 3. GOVA-Koeffizient und 4. Test auf Unverzerrtheit.

Der mittlere absolute Prognosefehler bezogen auf die naive Prognose (*MRAPnP*) zeigt, dass die Konjunkturprognosen in 136 von 144 betrachteten Fällen (94,4%) exakter sind als die entsprechenden naiven Prognosen (Tabelle 2).

Auch der Test auf Informationszuwachs führt zu einer recht positiven Beurteilung der Konjunkturprognostiker. In 68 der 72 betrachteten Fälle (94,4%) zeigt sich eine statistisch signifikante Verringerung des Prognosefehlers bei allmählicher Verkürzung des Prognosehorizontes (Tabelle 3). In der weit überwiegenden Mehrzahl der Fälle sind die Konjunkturrexperten also bereit und in der Lage, relevante Ereignisse während des sich allmählich verkürzenden Prognosehorizontes in ihren Prognosen zu berücksichtigen.

Die Analyse der Konjunkturprognosen mit Hilfe des GOVA-Koeffizienten führt ebenfalls zu einem günstigen Urteil. Bei 62 von 72 betrachteten Prognosezeitreihen (86,1%) liegen keine gegenwartsorientierten Verlaufsanpassungen vor (Tabelle 3). Die weit überwiegende Mehrzahl der Prognosen spiegelt also die Entwicklungen der Zukunft deutlich besser wider als die der Gegenwart.

Bei 33 der 36 Prognosegegenstände (91,7%) werden folgende Anforderungen gleichzeitig erfüllt: (a) bei einem Prognosehorizont von neun Monaten werden die naiven Prognosen übertroffen, (b) bei Betrachtung des absoluten Prognosefehlers kann ein Informationszuwachs statistisch signifikant festgestellt werden und (c) bei einem Prognosehorizont von neun Monaten treten keine gegenwartsorientierte Verlaufsanpassungen auf (Tabelle 5). Dieses Ergebnis weist durchaus gut entwickelte Prognosefähigkeiten bei den Konjunkturexperten aus.

Der Test auf Unverzerrtheit betrachtet hingegen die untersuchten Prognosen ausnahmslos als gescheitert (Tabellen 4 und 5). Wir sind der Auffassung, dass die Diskrepanz zwischen dem Test auf Unverzerrtheit und den anderen drei Beurteilungsmaßstäben darauf zurückzuführen ist, dass der Test auf Unverzerrtheit unrealistische Ansprüche an die Prognostiker stellt.

Konjunkturprognosen sind wesentlich exakter als Finanzmarktprognosen. Dieser deutliche Vorsprung wird bei Anwendung des Tests auf Unverzerrtheit nivelliert und verschleiert. Wir plädieren deshalb dafür, dem Test auf Unverzerrtheit in Zukunft bedeutend weniger Aufmerksamkeit zu schenken als dies bisher der Fall ist.

Literatur

- Ackert, L. F. / Hunter, W. C. (1994), Rational Expectations and the Dynamic Adjustment of Security Analysts' Forecasts to New Information, in: *The Journal of Financial Research*, Bd. 17, S. 387-401.
- Ager, P / Kappler, M. / Osterloh, S. (2007), The Accuracy and Efficiency of the Consensus Forecasts: A Further Application and Extension of the Pooled Approach, *Centre for European Economic Research Discussion Paper*, Nr. 07-058, Mannheim.
- Andolfatto, D / Hendry S. / Moran, K. (2008), Are inflation expectations rational? in: *Journal of Monetary Economics*, Bd. 55, S. 406-422.
- Andres, P. / Spiwoks, M. (1999), Forecast Quality Matrix, A Methodological Survey of Judging Forecast Quality of Capital Market Forecasts, in: *Journal of Economics and Statistics*, Bd. 219, S. 513-542.

- Ball, R. J. (1962), The Prediction of Wage-Rate Changes in the United Kingdom Economy 1957-60, in: *The Economic Journal*, Bd. 72, S. 27-44.
- Batchelor, R. (2007), Forecaster Behaviour and Bias in Macroeconomic Forecasts, *ifo Working Papers*, Nr. 39, ifo Institute of Economic Research at the University of Munich, Munich.
- Bofinger, P. / Schmidt, R. (2003), On the reliability of professional exchange rate forecasts: An empirical analysis for the EUR/US-\$ rate, in: *Financial Markets and Portfolio Management*, Bd. 17, S. 437-449.
- Bonham, C. S. / Cohen, R. H. (2001), To Aggregate, Pool, or Neither: Testing the Rational-Expectations Hypothesis Using Survey Data, in: *Journal of Business & Economic Statistics*, Bd. 19, S. 278-291.
- Bowles, C / Friz, R / Genre, V / Kenny, G / Meyler, A. / Rautanen, T. (2007), The ECB Survey of Professional Forecasters (SPF) – A Review After Eight Years' Experience, *European Central Bank Occasional Paper Series*, Nr. 59, April 2007, Frankfurt am Main.
- Breusch, T. S. / Pagan, A. R. (1979), Simple test for heteroscedasticity and random coefficient variation, in: *Econometrica*, Bd. 47, S.1287–1294.
- Brooks, R. / Gray, B. (2004), History of forecasters: Naive forecasts are better than the consensus WSJ forecast, in: *Journal of Portfolio Management*, Bd. 31, S. 113-117.
- Chinn, M. / Frankel, J. (1994), Patterns in Exchange Rate Forecasts for Twenty-five Currencies, in: *Journal of Money, Credit, and Banking*, Bd. 26, S. 759-770.
- De Gooijer, J. G. / Hyndman, R. J. (2006), 25 years of time series forecasting, in: *International Journal of Forecasting*, Bd. 22, S. 443-473.
- Diebold, F. X. (2007), *Elements of Forecasting*, 4th International Edition, Thomson Learning.
- Diebold, F. X. / Lopez, J. A. (1996), Forecast Evaluation and Combination, in: Maddala, G. S. / Rao, C. R. (Hrsg.), *Handbook of Statistics*. 14: S. 241-268, Amsterdam.
- Diebold, F. X. / Mariano, R. S. (1995), Comparing Predictive Accuracy, in: *Journal of Business and Economic Statistics*, Bd. 13, S. 253-263.
- Döpke, J. / Fritsche, U. (2006), Growth and inflation forecasts for Germany – A panel-based assessment of accuracy and efficiency, in: *Empirical Economics*, Bd. 31, S. 777-798.
- Dovern, J / Fritsche, U. / Slacalek, J. (2009), Disagreement Among Forecasters in G7 Countries, *European Central Bank Working Paper Series*, Nr. 1082, Frankfurt am Main.
- Dovern, J. / Weisser, J. (2009), Accuracy, Unbiasedness and Efficiency of Professional Macroeconomic Forecasts: An Empirical Comparison for the G7, *Jena Economic Research Papers*, Nr. 2009-091, Friedrich Schiller University, Jena.
- Fair, R. C. / Shiller, R. J. (1990), Comparing Information in Forecasts from Econometric Models, in: *The American Economic Review*, Bd. 80, S. 375-389.

- Fildes, R. / Stekler, H. (2002), The state of macroeconomic forecasting, in: *Journal of Macroeconomics*, Bd. 24, S. 435-468.
- Fraser, P. / MacDonald, R. (1993), The Efficiency of CAC Stock Price Forecasts, in: *Revue économique*, Bd. 44, S. 991-1000.
- Harvey, D. I / Leybourne, S. J. / Newbold, P. (1998), Tests for Forecast Encompassing, in: *Journal of Business & Economic Statistics*, Bd. 16, S. 254-259.
- Henriksson, R. D. / Merton, R. C. (1981), On Market Timing and Investment Performance, II. Statistical Procedures for Evaluating Forecasting Skills, in: *Journal of Business*, Bd. 54, S. 513-533.
- Joutz, F. / Stekler, H. O. (2000), An evaluation of the predictions of the Federal Reserve, in: *International Journal of Forecasting*, Bd. 16, S. 17-38.
- Keane, M. P. / Runkle, D. E. (1990), Testing the Rationality of Price Forecasts: New Evidence from Panel Data, in: *The American Economic Review*, Bd. 80, S. 714-735.
- Lakonishok, J. (1980), Stock Market Return Expectations: Some General Properties, in: *The Journal of Finance*, Bd. 35, S. 921-931.
- Laplace, P. S. (1814), *Essai philosophique sur les probabilités*, Paris.
- Mankiw, N. G / Reis, R. / Wolfers, J. (2003), Disagreement About Inflation Expectations, in: *NBER Macroeconomics Annual*, Bd. 18, S. 209-248.
- Manzur, M. (1988), How Much Are Exchange Rate Forecasts Worth? in: *Australian Journal of Management*, Bd. 13, S. 93-113.
- Mathews, B. P. / Diamentopoulos, A. (1994), Towards a Taxonomy of Forecast Error Measures, A Factor-comparative Investigation of Forecast Error Dimensions, in: *Journal of Forecasting*, Bd. 13, S. 409-416.
- Mehra, Y. P. (2002), Survey Measures of Expected Inflation: Revisiting the Issues of Predictive Content and Rationality, in: *Federal Reserve Bank of Richmond Economic Quarterly*, Bd. 88/3, S. 17-36.
- Merton, R. C. (1981), On Market Timing and Investment Performance, I. An Equilibrium Theory of Value for Market Forecasts, in: *Journal of Business*, Bd. 54, S. 363-406.
- Mincer, J. / Zarnowitz, V. (1969), The Evaluation of Economic Forecasts, in: Mincer, J. (Ed.), *Economic Forecasts and Expectation*, New York, Columbia University Press, S. 3-46.
- Mose, J. S. (2005), Expert Forecasts of Bond Yields and Exchange Rates, in: *Danmarks Nationalbank Monetary Review*, Bd. 4th Quarter 2005, S. 91-95.
- Muth, J. F. (1961), Rational Expectations and the Theory of Price Movements, in: *Econometrica*, Bd. 29, S. 315-334.

- Osterloh, S. (2008), Accuracy and Properties of German Business Cycle Forecasts, in: *Applied Economics Quarterly*, Bd. 54, S. 27-57.
- Poulizac, D / Weale, M. / Young, G. (1996), The Performance of National Institute Economic Forecasts, in: *National Institute Economic Review*, Nr. 156, S. 55-62.
- Schuh, S. (2001), An Evaluation of Recent Macroeconomic Forecast Errors, in: *New England Economic Review*, Bd. Jan./Feb. 2001, S. 35-56.
- Simon, D. P. (1989), The Rationality of Federal F/s Rate Expectations: Evidence from a Survey, in: *Journal of Money, Credit, and Banking*, Bd. 21, S. 388-393.
- Scheier, J./ Spiwoks, M. (2006), Aktives Portfoliomanagement am britischen Anleihenmarkt, in: Meyer-Bullerdiek, F. / Spiwoks, M. (Hrsg.), *Reihe Bank- und Finanzwirtschaft*, Bd. 4.
- Spiwoks, M. (2004), The Usefulness of ZEW Stock Market Forecasts for Active Portfolio Management Strategies, in: *Journal of Economics and Statistics*, Bd. 224, 2004, S. 557-578.
- Spiwoks, M. (2009), Deceptive Unity: Surprising Characteristics of Pre-Tax Corporate Profit Forecasts, in: *International Research Journal of Finance and Economics*, Bd. 26, S. 170-177.
- Spiwoks, M / Bedke, N. / Hein, O. (2008), Forecasting the past: The case of US interest rate forecasts, in: *Financial Markets and Portfolio Management*, Bd. 22, S. 357-379.
- Spiwoks, M / Bedke, N. / Hein, O. (2009), The Pessimism of Swiss Bond Market Analysts and the Limits of the Sign Accuracy Test – An Empirical Investigation of Their Forecasting Success Between 1998 and 2007, in: *International Bulletin of Business Administration*, Bd. 4, S. 6-19.
- Spiwoks, M / Bedke, N. / Hein, O. (2010), Topically Orientated Trend Adjustment and Autocorrelation of the Residuals - An Empirical Investigation of the Forecasting Behavior of Bond Market Analysts in Germany, in: *Journal of Money, Investment, and Banking*, Bd. 14, S. 16-35.
- Spiwoks, M / Gubaydullina, Z. / Hein, O. (2011), Trapped in the Here and Now – New Insights into Financial Market Analyst Behavior, *Wolfsburg Working Papers*, Nr. 11-01, Ostfalia University of Applied Sciences, Wolfsburg.
- Storbeck, O. (2010), Wie die Finanzkrise die VWL auf den Kopf stellt, in: *Handelsblatt*, 11. Januar 2010, S. 20-21.
- Theil, H. (1971), *Principles of Econometrics*, New York.

Kapitel 3

Verbesserung von Konjunkturprognosen mittels des OECD Composite Leading Indicators und den Ergebnissen der ifo World Economic Survey

Eine empirische Untersuchung für die G7-Staaten

Erschienen als:

Diskussionsbeitrag der Sonderforschungsgruppe Institutionenanalyse (sofia), Nr. 11-6, Darmstadt 2011.

Auch erschienen als:

The Improvement of Annual Economic Forecasts by Using Non-Annual Indicators: An Empirical Investigation for the G7 States, in: Wolfsburg Working Papers (WWP), Ostfalia Hochschule für angewandte Wissenschaften, Nr. 12-01.

Abstract

Eine zentrale Forderung an Konjunkturprognosen ist die effiziente Verarbeitung frei verfügbarer Informationen. Dieses Papier zeigt anhand zweier Frühindikatoren, dass diese Möglichkeiten nicht ausgeschöpft werden. Hierzu werden Konsensprognosen für die Konjunktur der G7-Staaten von 1991 – 2009 untersucht. Gemäß dem aktuellen Forschungsstand zeigt sich eine mäßige Prognosequalität, da die Prognosen für das kommende Jahr erst ab dem Herbst des laufenden Jahres einen messbaren Zusammenhang zum tatsächlichen Konjunkturverlauf aufweisen. Anders stellt sich die Situation bei den untersuchten Frühindikatoren dar. Sowohl die Ergebnisse einer weltweiten Expertenumfrage des Münchener ifo-Instituts hinsichtlich der Entwicklung der Weltwirtschaft (World Economic Survey) als auch der landesspezifische Verbundindikator der OECD (Composite Leading Indicator) weisen für alle Staaten einige Monate früher einen messbaren Zusammenhang zur Wirtschaftsentwicklung auf. Bei gleichzeitiger Betrachtung von Prognosen und Frühindikatoren zeigt sich, dass sich die Prognosequalität für die meisten G7-Staaten signifikant verbessern ließe. Dies gilt für diejenigen Prognosen, die im Sommer für das jeweilige Folgejahr abgegeben werden. Da die Prognosen in diesem Zeitraum eine schlechte Qualität aufweisen, können sie in besonderem Maße von der Berücksichtigung der Frühindikatoren profitieren.

1 Einführung

An Konjunkturprognosen werden die verschiedensten Forderungen gestellt. Die Hauptforderung ist zunächst einmal eine gute Qualität, damit Sie als Planungsgrundlage für Staaten, Unternehmen und Akteure an den Finanzmärkten dienen können.

Die Grundlagen für eine Aufteilung dieser Hauptforderung in einzelne Bestandteile hat Muth (1961) mit seiner Hypothese rationaler Erwartungen geliefert. Demnach weicht eine perfekte Prognose nur in der Höhe der unvorhersehbaren Zufallseinflüsse von der tatsächlichen Realisierung des Prognosegegenstands ab. Hierfür müssen die Prognosen mehrere Eigenschaften aufweisen:

- Sie dürfen keine systematische Verzerrung aufweisen. Diese liegt vor, wenn es zu immer gleichen Fehleinschätzungen kommt.
- Alle relevanten Informationen müssen für die Erstellung der Prognosen berücksichtigt werden. Dies schließt die Kenntnisnahme der eigenen Prognosefehler der Vergangenheit mit ein.
- Für Konjunkturprognosen im Speziellen muss gelten, dass die Prognostiker Kenntnis über die ökonomischen Zusammenhänge für das Zustandekommen gesamtwirtschaftlicher Dynamik besitzen.

Sind alle 3 Punkte erfüllt, dann spricht man von „effizienten“ Prognosen. Die Finanzkrise hat den Ökonomen zum wiederholten Mal vor Augen geführt, dass scheinbar vor allem beim 3. Punkt große Defizite auszumachen sind. Bei den Volkswirtschaften handelt es sich um außerordentlich komplexe Gebilde, deren Strukturierung sich darüber hinaus laufend ändert. So wird selbst die Debatte um das Zustandekommen der Konjunktur trotz einer bald 150-jährigen Forschungstradition¹ noch mit völlig offenem Ausgang² geführt.

Die lückenhaften Kenntnisse der ökonomischen Zusammenhänge äußern sich in der eingeschränkten Treffsicherheit der Konjunkturprognosen. Davon unbenommen sind allerdings die

¹ Stavenhagen (1969) bezeichnet Clément Juglar (1819-1905) als Entdecker des Konjunkturzyklus. In dessen 1862 erschienenem Werk „Des crises commerciales et de leur retour périodique en France, en Angleterre et aux Etats Unis“ benennt Juglar die drei Phasen Aufschwung, Krise und Stockung, stellt eine Wellenbewegung von 8-10 Jahren fest und belegt dies erstmals anhand einer empirischen Analyse.

² Diskutiert werden unter anderem folgende Ursachen für Konjunkturzyklen: Temporäre Ungleichgewichte auf den Güter- und Faktormärkten (Keynesianismus), monetäre Gründe (Österreichische Schule), politisch induzierte Zyklen (neue politische Ökonomie), externe Schocks, Vertrauensverlust als Grund für Rezessionen.

ersten beiden Forderungen. Wird das Wirtschaftswachstum bspw. kontinuierlich überschätzt, so müssten die zukünftigen Prognosen eben nach unten korrigiert werden. Dennoch werden solche systematische Fehler von Konjunkturprognosen immer wieder festgestellt, so auch in dieser Studie.

Bezüglich des 2. Punkts weist Nordhaus (1987) darauf hin, dass man schlecht „alle“ Informationen auf ihre Relevanz hin testen könnte.³ An diesem Punkt setzt diese Studie an. Wenn frei verfügbare Frühindikatoren einen höheren Zukunftsbezug als professionelle Konjunkturprognosen aufweisen, kann man die Verarbeitung dieser Informationen mit aller Deutlichkeit von den Prognostikern fordern. Unter Zukunftsbezug wird die Stärke des gemessenen Zusammenhangs zwischen Prognose bzw. Frühindikator und der tatsächlich eingetretenen Wirtschaftsentwicklung verstanden.

Der Fokus liegt bei dieser Untersuchung auf Frühindikatoren mit einem hohen qualitativen Anteil. Einen frühen Hinweis auf ihren Nutzen liefert Kirchgässner (1982), der die Ergebnisse der Neujahrsumfrage des Allensbach-Instituts für die Wirtschaftsentwicklung in Deutschland mit der Konjunkturprognose des Sachverständigenrats der Bundesregierung vergleicht. Die Umfrageergebnisse weisen dabei einen stärkeren Zusammenhang zur Wirtschaftsentwicklung des Folgejahres auf als die professionellen Konjunkturprognosen. Der Nutzen qualitativer Frühindikatoren wird vor allem für die rechtzeitige Erkennung konjunktureller Wendepunkte immer wieder festgestellt (Fritsche 1999, Hüfner / Schröder 2002, Kunkel 2003, Abberger / Nierhaus 2007, Seiler 2009, Ziegler 2009). Claveria, Pons und Ramos (2007) testen die Verwendung solcher Indikatoren in ökonometrischen Prognosemodellen. Ein Vergleich mit tatsächlich abgegebenen Konjunkturprognosen für die Wirtschaftsentwicklung auf Jahresbasis liegt jedoch nicht vor.

Ein solcher Vergleich erfolgt in dieser Studie mittels zweier Datenreihen, die auf ihre Relevanz für die Konjunkturprognosen hin untersucht werden. Zum einen erfolgt eine Betrachtung des monatlich von der OECD veröffentlichten Composite Leading Indicators (CLI). Dieser wird für jede Volkswirtschaft individuell konstruiert und beinhaltet sowohl quantitative als auch qualitative Frühindikatoren. Zum anderen erfolgt eine Überprüfung einer weltweiten Umfrage

³ Er schlägt daher vor, lediglich einen Test auf die „schwache“ Effizienz der Prognosen durchzuführen. Diese liegt vor, wenn die Prognosefehler nicht autokorreliert sind sondern einem sog. random-walk unterliegen.

des Münchener ifo-Instituts, der World Economic Survey (WES). Dabei werden weltweit Wirtschaftsexperten unter anderem hinsichtlich ihrer Einschätzung zur Lage der Weltkonjunktur in den nächsten 6 Monaten befragt.

Stellvertretend für die Konjunkturprognosen werden Konsensprognosen der britischen Fachzeitschrift *Consensus Forecasts* als Datenbasis verwendet. Ihre Veröffentlichung erfolgt monatlich, wobei sich je eine Prognose auf das laufende und eine auf das kommende Jahr bezieht. Für jedes Jahr liegen also 24 separate Prognosen vor. Die Untersuchung erstreckt sich auf den Zeitraum 1991 – 2009 und umfasst die G7-Staaten.

Der 2. Abschnitt dieser Untersuchung dient der Beurteilung der Qualität der untersuchten Konjunkturprognosen. Im 3. Abschnitt erfolgt ein Vergleich des Zukunftsbezugs der Prognosen mit den Zeitreihen der OECD und des ifo-Instituts. Der 4. Abschnitt zeigt schließlich, dass die Berücksichtigung dieser Zusatzinformationen mit hoher Wahrscheinlichkeit zu einer Verbesserung der Konjunkturprognosen beigetragen hätte.

2 Prognosequalität

Zunächst erfolgt eine Einschätzung der Qualität der in dieser Studie untersuchten Konjunkturprognosen. Als Datenbasis dienen die Veröffentlichungen der britischen Fachzeitschrift *Consensus Forecasts*. Die Zeitschrift wurde 1989 in der Folge einer wissenschaftlichen Debatte um den Nutzen von zusammengefassten Prognosen gegründet.⁴ Seitdem werden monatlich von ca. 40 Banken und Wirtschaftsforschungsinstituten Prognosen für verschiedene volkswirtschaftliche Indikatoren und Staaten publiziert – gemeinsam mit dem jeweiligen arithmetischen Mittel der Einzelprognosen, der sogenannten Konsensprognose.

Die veröffentlichten Prognosen lassen sich in zwei Kategorien einteilen – Prognosen mit konstantem und Prognosen mit variablem Prognosehorizont. Bei konstantem Horizont (fixed-horizon-forecasts) erfolgt zwischen zwei Ausgaben keine Veränderung der Zeitspanne zwischen Veröffentlichung und Eintritt des Prognosegegenstands. Zinsprognosen (Horizont 3 Monate)

⁴ Bates und Granger haben die Möglichkeit der Konsensprognosen 1969 als Erste erörtert. Die Debatte über den Nutzen und die Grenzen von Konsensprognosen gipfelte 1989 in Sonderausgaben des „*Journal of Forecasting*“ und des „*International Journal of Forecasting*“. Anschließend kam es zur Gründung von „*Consensus Economics*“ und zum Start der Publikationsreihe „*Consensus Forecasts*“.

der Januar-Ausgabe haben ihren Gültigkeitszeitpunkt im April, die Vorhersagen der Februar-Ausgabe im Mai.

Die in dieser Studie untersuchten Prognosen für das Bruttoinlandsprodukt gehören der zweiten Kategorie an, den sogenannten fixed-event oder fixed-target-forecasts. Monatlich wird je eine Prognose für das laufende und eine für das kommende Jahr veröffentlicht. Die Prognosen der Januar-Ausgabe erreichen ihr volle Gültigkeit somit zum Ende des laufenden und des kommenden Jahres – ebenso wie die Prognosen der Februar-Ausgabe. Der Prognosehorizont für das laufende Jahr verringert sich dabei zwischen den beiden Ausgaben von 12 auf 11 Monate, der für das kommende Jahr von 24 auf 23 Monate. Für jedes Jahr werden von den Instituten also 24 Prognosen bei kontinuierlich rückläufigem Prognosehorizont abgegeben. Dies ermöglicht eine Betrachtung der Prognosegüte in Abhängigkeit vom Prognosehorizont.

Consensus Forecasts erscheint seit Oktober 1989. Die laufende Vergrößerung dieser Datenbasis führt zunehmend zu Publikationen, die sich mit deren Auswertung befassen. Batchelor (2001) stellt fest, dass die Prognosen des Weltwährungsfonds und der OECD schlechter abschneiden als die in Consensus Forecasts veröffentlichten Konsensprognosen, in die vorwiegend Prognosen privater Banken eingehen. Er rät zur Vorsicht bei der Interpretation der Ergebnisse, da die Datenbasis zu diesem Zeitpunkt noch recht schmal ist. Auch Blix et al. (2001) kommen bei ihrer Untersuchung zu diesem Schluss. Zudem sind in den 1990ern die Prognosen für die Inflationsraten besser geglückt als für das jeweilige Bruttoinlandsprodukt. Auch sehen sie Anzeichen für ein Herdenverhalten der Prognostiker, da die Prognoserevisionen im Zeitverlauf große Ähnlichkeiten aufweisen. Loungani (2001) kann keine größeren Unterschiede zwischen der Qualität von Prognosen für Industrie- und Entwicklungsländer ausmachen. Allerdings seien die Prognosen vor allem für Entwicklungsländer in der Tendenz durch Überschätzung verzerrt und generell durch jeweils hohe Korrelationen der Revisionen zu glatt, was zur Einschätzung des Vorliegens mangelnder „Effizienz“ der Prognosezeitreihen führt. Isiklar und Lahiri (2007) stellen bei einer Überprüfung von BIP-Prognosen für 18 Staaten fest, dass die Qualität erst bei einem Horizont von 14 Monaten ein befriedigendes Niveau erreicht.

Ager, Kappler und Osterloh (2009) identifizieren bei der Mehrzahl von Prognosen für das Bruttoinlandsprodukt von 12 Industrieländern systematische Prognosefehler, sofern sich die Vorhersagen auf das kommende Jahr beziehen. Sie vermuten ein Herdenverhalten als mögliche Ursache.

Osterloh (2008) kommt bei seiner Analyse von BIP-Prognosen für Deutschland zu dem Schluss, dass die Vorhersage des jeweils kommenden Jahres kaum gelingt. Als mögliche Erklärung verweist er auf die zahlreichen exogenen Schocks im Untersuchungszeitraum, welche zudem nach ihrem Eintritt zu zögerlich in den Prognosen berücksichtigt wurden. Doern und Weisser (2011) untersuchen die Daten von Consensus Economics mittels eines panel-basierten Ansatzes, wobei sie auch die individuellen Prognosen für die G7-Staaten berücksichtigen. Sie stellen eine große Heterogenität der Prognosegenauigkeit, und Rationalität fest, wobei in der Tendenz die Inflationsprognosen am besten gelingen.

In dieser Studie werden die BIP-Prognosen (BSP für Japan bis 1993) für die G7 Staaten untersucht. Diese wurden von Anfang an in Consensus Forecasts berücksichtigt, während weitere Staaten erst später hinzukamen. Das erste Jahr, für welches die Prognosen ab dem Januar des Vorjahres zur Verfügung stehen, ist das Jahr 1991. Die Auswertung berücksichtigt Prognosen für die Jahre 1991 – 2009.

Messen müssen sich die Prognosen an der tatsächlich eingetretenen Wachstumsrate des Bruttoinlandsprodukts der G7-Staaten, wobei die Wahl der Vergleichsdaten weniger trivial ist als sie zunächst scheint. Zieht man Zeitreihen der jeweiligen statistischen Ämter für den gesamten Untersuchungszeitraum heran, so ist mit deutlichen Revisionen als Folge von Änderungen bei der Datenerfassung bzw. Aufbereitung der Werte zu rechnen. Diese können zu „künstlichen“ Prognosefehlern führen, welche das Untersuchungsergebnis verfälschen würden. Überwiegend werden in anderen Studien mit diesem Datensatz daher die tatsächlichen Werte als Vergleichsmaßstab genommen, die in der Mitte des jeweils folgenden Jahres in Consensus Forecasts veröffentlicht worden sind. Dies ist auch in dieser Studie der Fall. Als Vergleichswert wird das jeweilige Bruttoinlandsprodukt der Mai-Ausgabe des Folgejahres verwendet.

Um eine Einschätzung der Prognosequalität vornehmen zu können, erfolgt zunächst die Betrachtung der mittleren absoluten Prognosefehler (*MAP*) für die einzelnen Länder und Prognosehorizonte. Dabei handelt es sich um den Mittelwert aller absoluten Prognosefehler, die sich wiederum aus der Differenz zwischen Prognose P und tatsächlichem Wert A zum jeweiligen Zeitpunkt t bei der Gesamtzahl $T=19$ Jahre ergeben.

$$(1) \quad \text{MAP} = \frac{1}{T} \sum_{t=1}^T |P_t - A_t|$$

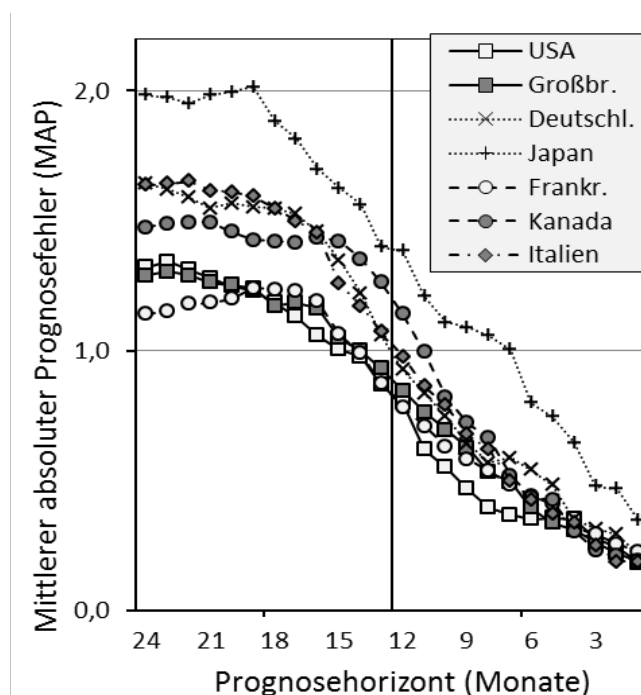


Abbildung 1: Mittlerer absoluter Prognosefehler der BIP-Prognosen

Für alle G7-Staaten zeigt sich ein sehr ähnliches Bild (Abbildung 1). Die Prognosen für das jeweils kommende Jahr weisen fast durchgängig einen $MAP > 1$ Prozentpunkt auf. Um den Jahreswechsel herum verbessern sich die Prognosen deutlich und nehmen bei einem Prognosehorizont von 11 Monaten bis auf die Prognosen für Japan durchgängig einen $MAP < 1$ Prozentpunkt an. Das Bruttoinlandsprodukt für das jeweils laufende Jahr ist also im Mittel mit einem Fehler von deutlich unter einem Prozentpunkt vorhergesagt worden. Der Informationsvorteil durch kürzere Prognosehorizonte und die daraus resultierenden niedrigeren Prognosefehler sind für alle Volkswirtschaften gleichermaßen erkennbar. Die einzelnen Werte sind Tabelle 1 zu entnehmen.

Tabelle 1: Mittlerer absoluter Prognosefehler der Konsensprognosen für das BIP der G7-Staaten von 1991-2009

PH Mon.	(1) USA	(2) UK	(3) DE	(4) JA	(5) FR	(6) KA	(7) IT
24	1,32	1,29	1,64	1,98	1,14	1,47	1,64
23	1,34	1,31	1,62	1,97	1,15	1,49	1,64
22	1,31	1,29	1,59	1,95	1,18	1,49	1,65
21	1,28	1,26	1,55	1,98	1,18	1,49	1,62
20	1,25	1,25	1,57	1,99	1,20	1,46	1,61
19	1,23	1,24	1,55	2,02	1,24	1,43	1,59
18	1,18	1,17	1,55	1,88	1,24	1,42	1,55
17	1,13	1,18	1,53	1,82	1,23	1,42	1,50
16	1,06	1,16	1,46	1,70	1,19	1,44	1,45
15	1,01	1,05	1,35	1,63	1,06	1,42	1,26
14	0,98	1,00	1,22	1,56	0,99	1,35	1,17
13	0,87	0,93	1,06	1,40	0,87	1,26	1,07
12	0,80	0,85	0,93	1,38	0,78	1,14	0,98
11	0,62	0,76	0,84	1,21	0,71	0,99	0,86
10	0,55	0,69	0,75	1,11	0,63	0,82	0,79
9	0,47	0,63	0,64	1,09	0,58	0,72	0,68
8	0,39	0,53	0,57	1,06	0,54	0,66	0,62
7	0,37	0,49	0,59	1,01	0,48	0,52	0,50
6	0,35	0,39	0,54	0,80	0,44	0,44	0,43
5	0,36	0,34	0,48	0,75	0,41	0,43	0,37
4	0,35	0,31	0,35	0,65	0,32	0,31	0,34
3	0,28	0,26	0,32	0,48	0,29	0,23	0,25
2	0,24	0,22	0,29	0,47	0,26	0,21	0,19
1	0,18	0,19	0,22	0,35	0,23	0,19	0,19

PH = Prognosehorizont in Monaten

Für die Beantwortung der Frage, ob diese Werte jetzt gut oder schlecht sind, ist ein Vergleichswert hilfreich. Üblich bei der Prognoseevaluation ist der Vergleich mit der naiven Prognose, sei es mit dem mittleren relativen absoluten Prognosefehler bezogen auf die naive Prognose (*MRAPnP*) oder Theils *U* (Theil 1971). Bei einer naiven Prognose erfolgt lediglich eine Übertragung der Gegenwartswerte in die Zukunft. Der französische Mathematiker Pierre Simon Laplace (1841) hat diese Überlegung in die Literatur als „Prinzip des unzureichenden Grundes“ eingeführt.

In dieser Studie wird mit Verweis auf das Untersuchungsdesign eine Alternative zu diesen gängigen Gütemaßen verwendet. So ließe sich zwar für eine Prognose zu Jahresbeginn möglicherweise der Wert des Vorjahres als naive Prognose nutzen – sofern dieser bereits von den jeweiligen Statistikämtern veröffentlicht worden ist. Ab der Jahresmitte kann jedoch der Gegenwartswert nicht in die Zukunft fortgeschrieben werden, da das Wirtschaftswachstum des laufenden Jahres noch nicht feststeht. Als Alternative bietet sich die gemessene durchschnittliche Schwankung des BIP im Untersuchungszeitraum an. Es ist legitim von den Prognostikern zu

fordern, dass die Prognosefehler niedriger als die durchschnittlichen Änderungsraten des Gegenstands der Prognosen sind. Sind die Fehler größer, so bieten die Prognosen keinen Nutzen.

Um leicht interpretierbare Ergebnisse zu erzielen, erfolgt ähnlich dem *MRAPnP* eine Division der absoluten Prognosefehler durch die jeweiligen absoluten Änderungsraten des Bruttoinlandsprodukts. Somit gelangt man zum mittleren absoluten Prognosefehler bezogen auf die mittlere Schwankung des Prognosegegenstands (*MRAPS*).

$$(2) \quad MRAPS = \frac{\frac{1}{T} \sum_{t=1}^T |P_t - A_t|}{\frac{1}{T} \sum_{t=1}^T |A_{t-1} - A_t|}$$

Ein *MRAPS* < 1 bedeutet, dass die Prognosefehler geringer als die Durchschnittsschwankung des BIP ausgefallen sind. Einen solchen Wert sollte man zumindest über einen längeren Untersuchungszeitraum von einer nutzbaren Prognose erwarten können.

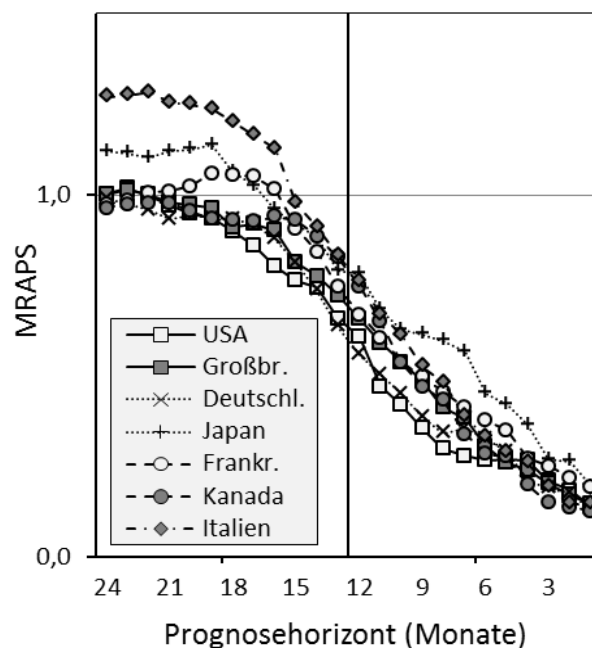


Abbildung 2: mittlerer absoluter Prognosefehler bezogen auf die mittlere Schwankung des Prognosegegenstands

In Abbildung 2 wird deutlich, dass die Prognostiker diesem Anspruch zunächst nicht gerecht werden. Die Prognosen für das kommende Jahr erreichen erst gegen Ende des laufenden Jahres einen Wert deutlich < 1. Die Prognosefehler der BIP-Prognosen, die in der ersten Jahreshälfte für das kommende Jahr abgegeben wurden, liegen in einer ähnlichen Größenordnung wie die

durchschnittliche Veränderungsrate des Wirtschaftswachstums. Die Prognosefehler der Prognosen für Italien und Japan liegen zunächst sogar deutlich darüber. Ab dem Herbst verbessern sich die Prognosen für alle G7-Staaten deutlich. Hier scheint sich der Informationsgewinn durch die spätere Prognoseabgabe für die Prognostiker auszuzahlen. Die Prognosen für das jeweils laufende Jahr (Prognosehorizont ≤ 12 Monate) weisen durchweg einen *MRAPS* von deutlich < 1 auf. Die Leistung relativiert sich allerdings dadurch ein wenig, da in diesem Fall zunehmend real gemessene Werte des Bruttoinlandsprodukts in die Prognosen eingehen können. Die einzelnen Werte sind Tabelle 2 zu entnehmen.

Tabelle 2: *MRAPS der Konsensprognosen für das BIP der G7-Staaten von 1991-2009*

PH Mon.	(1) USA	(2) UK	(3) DE	(4) JA	(5) FR	(6) KA	(7) IT
24	1,00	1,00	0,99	1,12	0,97	0,96	1,27
23	1,02	1,01	0,98	1,11	0,98	0,97	1,27
22	1,00	1,00	0,96	1,10	1,01	0,98	1,28
21	0,97	0,98	0,93	1,12	1,01	0,98	1,25
20	0,95	0,97	0,94	1,13	1,02	0,95	1,25
19	0,94	0,96	0,93	1,14	1,06	0,93	1,24
18	0,90	0,91	0,93	1,06	1,06	0,93	1,20
17	0,86	0,92	0,92	1,02	1,05	0,93	1,16
16	0,80	0,90	0,88	0,96	1,01	0,94	1,13
15	0,76	0,81	0,81	0,92	0,91	0,93	0,98
14	0,74	0,78	0,74	0,88	0,84	0,89	0,91
13	0,66	0,72	0,64	0,79	0,75	0,83	0,83
12	0,61	0,66	0,56	0,78	0,67	0,75	0,76
11	0,47	0,59	0,50	0,68	0,61	0,65	0,67
10	0,42	0,54	0,45	0,63	0,54	0,54	0,61
9	0,36	0,49	0,39	0,61	0,50	0,47	0,53
8	0,30	0,41	0,34	0,60	0,46	0,43	0,48
7	0,28	0,38	0,35	0,57	0,41	0,34	0,39
6	0,27	0,31	0,33	0,45	0,38	0,29	0,33
5	0,27	0,26	0,29	0,42	0,35	0,28	0,29
4	0,27	0,24	0,21	0,37	0,27	0,20	0,26
3	0,21	0,20	0,19	0,27	0,25	0,15	0,20
2	0,18	0,17	0,18	0,26	0,22	0,14	0,15
1	0,14	0,15	0,13	0,20	0,19	0,13	0,15

PH = Prognosehorizont in Monaten

Im Folgenden werden die Eigenschaften der Prognosefehler eingehender betrachtet. Für „effiziente“ Prognosen gilt die Forderung, dass deren Fehler keine Systematik aufweisen dürfen. Dies leitet sich aus der Hypothese der rationalen Erwartungsbildung von Muth (1961) ab. Die einfachste Form systematischer Prognosefehler ist eine kontinuierliche Über- oder Unterschätzung des Prognosegegenstands.

Einen ersten Überblick erhält man durch die Berechnung des mittleren Prognosefehlers (*MP*), der eine Aufhebung von Über- und Unterschätzungen seitens der Prognostiker zulässt. Eine

längere Prognosezeitreihe müsste einen durchschnittlichen MP nahe 0 aufweisen, um als unverzerrt gelten zu können.

Seien die Prognose wieder P und der tatsächliche Wert wieder A zum jeweiligen Zeitpunkt t bei der Gesamtzahl der Beobachtungen $T=19$ Jahre. Der mittlere Prognosefehler berechnet sich dann folgendermaßen:

$$(3) \quad MP = \frac{1}{T} \sum_{t=1}^T P_t - A_t$$

Die mittleren Prognosefehler der Konsensprognosen für die G7-Staaten sind der Abbildung 3 zu entnehmen.

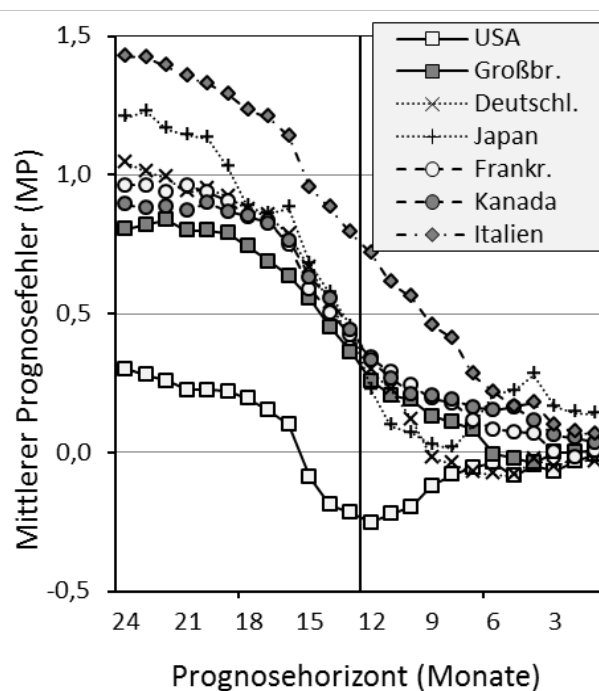


Abbildung 3: Mittlerer Prognosefehler der BIP-Prognosen

Bei den Prognosen für das kommende Jahr (Prognosehorizont > 12 Monate) erfolgt durch die Prognostiker fast ausschließlich eine Überschätzung des Wirtschaftswachstums. Lediglich ab einem Prognosehorizont von 15 Monaten sind durchschnittliche Unterschätzungen für die USA festzustellen. Bei den Prognosen für das laufende Jahr ergibt sich kein eindeutiges Bild mehr. Das Phänomen der durchschnittlichen Überschätzung des Wirtschaftswachstums verschwindet bei den geringeren Prognosehorizonten zusehends. Die einzelnen Werte sind der Tabelle 3 zu entnehmen.

Tabelle 3: Mittlerer Prognosefehler (MP) der Konsensprognosen für das BIP der G7-Staaten von 1991-2009

PH Mon.	(1) USA	(2) UK	(3) DE	(4) JA	(5) FR	(6) KA	(7) IT
24	0,30	0,81	1,04	1,21	0,96	0,89	1,43
23	0,28	0,82	1,01	1,23	0,96	0,88	1,42
22	0,26	0,84	0,99	1,16	0,94	0,88	1,39
21	0,23	0,80	0,94	1,14	0,96	0,87	1,35
20	0,23	0,80	0,95	1,13	0,94	0,90	1,33
19	0,22	0,79	0,92	1,03	0,91	0,87	1,29
18	0,19	0,74	0,87	0,88	0,85	0,85	1,23
17	0,15	0,69	0,85	0,86	0,83	0,83	1,21
16	0,10	0,64	0,78	0,88	0,75	0,76	1,14
15	-0,09	0,55	0,66	0,68	0,59	0,63	0,95
14	-0,19	0,45	0,54	0,57	0,51	0,55	0,88
13	-0,22	0,36	0,41	0,45	0,42	0,44	0,79
12	-0,25	0,26	0,29	0,23	0,34	0,33	0,72
11	-0,22	0,21	0,23	0,09	0,29	0,27	0,61
10	-0,19	0,19	0,12	0,07	0,24	0,21	0,56
9	-0,12	0,13	-0,02	0,03	0,19	0,21	0,46
8	-0,08	0,11	-0,04	0,02	0,18	0,19	0,41
7	-0,05	0,08	-0,07	0,10	0,12	0,16	0,28
6	-0,04	-0,01	-0,08	0,21	0,08	0,15	0,22
5	-0,08	-0,02	-0,08	0,22	0,07	0,16	0,16
4	-0,05	-0,04	-0,03	0,28	0,07	0,12	0,18
3	-0,07	0,00	-0,05	0,16	0,00	0,06	0,09
2	-0,03	0,00	-0,02	0,14	-0,02	0,05	0,07
1	-0,01	0,01	-0,04	0,14	0,01	0,04	0,06

PH = Prognosehorizont in Monaten

Für eine präzisere Überprüfung der Prognosefehler auf systematische Komponenten hin hat sich der Test auf Unverzerrtheit in Form der Mincer-Zarnowitz-Regression (Mincer / Zarnowitz 1969) als Standard etabliert. Seien A der jeweils reale Wert, P die dazugehörige Prognose und u der Fehlerterm jeweils zum Zeitpunkt t , dann hat die Mincer-Zarnowitz-Regression folgende Gestalt:

$$(4) \quad A_t = \alpha + \beta P_t + u_t$$

Für den Nachweis der Unverzerrtheit erfolgt ein Test der Gemeinschaftshypothese $\alpha=0$ und $\beta=1$. Zudem darf der Fehlerterm keine Autokorrelation aufweisen, wofür üblicherweise eine Überprüfung mittels des Durbin-Watson-Tests erfolgt.

An dieser Stelle ist es angebracht, auf die recht geringe Zahl der Beobachtungen pro durchgeführte Regression hinzuweisen. Für jeden Prognosehorizont und jedes Land stehen von 1991-2009 nur 19 Jahre und somit 19 Beobachtungen zur Verfügung. Dies führt zu großen Konfidenzintervallen, so dass die Ablehnung der Gemeinschaftshypothese mit einer geringen Irrtumswahrscheinlichkeit alleine durch die Beobachtungszahl unwahrscheinlich werden kann. Mit

entsprechender Vorsicht sind die Ergebnisse in Tabelle 4 zu interpretieren, in denen für jedes Land und jeden Prognosehorizont die Wahrscheinlichkeit der unberechtigten Ablehnung der Hypothese $\alpha=0$ und $\beta=1$ verzeichnet ist. Umgekehrt gilt jedoch: Werden systematische Prognosefehler festgestellt, so ist dies ob der geringen Beobachtungszahl als umso deutlicheres Ergebnis zu werten.

Tabelle 4: *p*-Werte des Tests auf Unverzerrtheit aus Gleichung (4)

PH Mon.	(1) USA	(2) UK	(3) DE	(4) JA	(5) FR	(6) KA	(7) IT
24	0,504	0,154	0,065*	0,027**	0,025**	0,118	0,008***
23	0,507	0,160	0,065*	0,028*	0,025**	0,145	0,008***
22	0,691	0,101	0,091*	0,041**	0,029**	0,131	0,009***
21	0,823	0,046**	0,133	0,039**	0,024**	0,130	0,009***
20	0,859	0,038**	0,136	0,036**	0,028**	0,076*	0,010**
19	0,815	0,020**	0,156	0,050*	0,028**	0,075*	0,012**
18	0,597	0,008*	0,183	0,096*	0,035**	0,072*	0,014**
17	0,501	0,029**	0,199	0,127	0,039**	0,081*	0,011**
16	0,589	0,038**	0,227	0,197	0,075*	0,150	0,014**
15	0,657	0,009*	0,206	0,353	0,145	0,252	0,007***
14	0,663	0,024**	0,163	0,467	0,189	0,318	0,005***
13	0,543	0,044**	0,140	0,584	0,243	0,405	0,002***
12	0,519	0,088*	0,143	0,853	0,332	0,512	0,001***
11	0,291	0,090*	0,167	0,963	0,381	0,440	0,002***
10	0,460	0,119	0,287	0,639	0,436	0,430	0,002***
9	0,682	0,158	0,827	0,532	0,438	0,516	0,010**
8	0,769	0,231	0,973	0,625	0,424	0,539	0,022**
7	0,885	0,170	0,636	0,419	0,346	0,199	0,127
6	0,860	0,289	0,555	0,262	0,297	0,123	0,237
5	0,642	0,489	0,608	0,252	0,289	0,049**	0,373
4	0,836	0,418	0,957	0,160	0,517	0,014**	0,219
3	0,751	0,589	0,845	0,146	0,746	0,015**	0,407
2	0,942	0,857	0,916	0,108	0,853	0,024**	0,130
1	0,958	0,970	0,695	0,217	0,994	0,093*	0,070

PH = Prognosehorizont in Monaten; * = Irrtumswahrscheinlichkeit < 10%, ** = Irrtumswahrscheinlichkeit < 5%, *** = Irrtumswahrscheinlichkeit < 1%;

Die Ergebnisse lassen eine deutlich Tendenz zu systematischen Prognosefehlern bei den Prognosen für das kommende Jahr erkennen (Prognosehorizont > 12 Monate). Die Hypothese einer Konstanten von 0 und eines Koeffizienten von 1 (siehe (4)) wird abgesehen von den Prognosen für die USA und Deutschland zumeist abgelehnt. Noch deutlicher wird das Ergebnis unter Einbeziehung des Durbin-Watson-Tests.

Bei den längeren Prognosehorizonten bis einschließlich 18 Monaten weist die Durbin-Watson-Statistik für fast alle Länder eine Autokorrelation 1. Ordnung auf (5%-Signifikanzniveau, Tabelle 5). Es sind also systematische Bestandteile bei den Residuen messbar.

Tabelle 5: Werte der Durbin-Watson-Statistik zur Überprüfung auf systematische Bestandteile in den Residuen

PH Mon.	(1) USA	(2) UK	(3) DE	(4) JA	(5) FR	(6) KA	(7) IT
24	0,90	0,93	1,31	1,01	1,30	1,06	1,25
23	0,90	0,96	1,30	1,02	1,30	1,07	1,28
22	0,91	0,98	1,34	1,02	1,32	1,15	1,29
21	0,93	1,10	1,39	1,03	1,40	1,13	1,30
20	0,97	1,10	1,42	1,03	1,39	1,23	1,33
19	1,00	1,10	1,49	1,02	1,35	1,31	1,32
18	0,96	1,24	1,51	1,05	1,37	1,36	1,33
17	1,12	1,16	1,54	1,07	1,36	1,48	1,42
16	1,30	1,25	1,61	1,14	1,66	1,49	1,42
15	1,78	1,64	1,71	1,23	2,00	1,85	1,78
14	1,87	1,67	1,83	1,34	2,22	1,96	2,09
13	1,97	1,88	2,10	1,59	2,22	2,13	2,24
12	1,88	2,02	2,23	1,69	2,28	2,37	2,46
11	1,96	2,10	2,09	1,99	2,49	2,48	2,48
10	1,92	2,20	1,97	2,21	2,51	2,59	2,51
9	1,98	2,33	1,87	2,19	2,48	2,55	2,66
8	2,35	2,36	1,57	2,25	2,39	2,57	2,67
7	2,29	2,34	1,57	2,39	2,08	2,71	2,31
6	2,33	2,32	1,58	2,23	2,09	2,34	2,22
5	2,06	2,35	1,45	2,27	1,92	2,23	2,28
4	2,04	2,22	1,14	2,21	1,86	1,63	2,43
3	2,29	1,90	0,99	2,34	1,78	1,59	2,49
2	2,11	1,99	0,94	2,29	1,67	1,50	2,14
1	2,17	1,68	0,77	2,30	1,55	2,02	2,13

PH = Prognosehorizont in Monaten; Durbin-Watson-Statistik: 0 = perfekte positive Autokorrelation, 2 = keine Autokorrelation, 4 = perfekte negative Autokorrelation; kritischer Wert (5% Signifikanzniveau) ist bei den hier vorliegenden 19 Beobachtungen und einer erklärenden Variable 1,4; Werte < 1,4 sind hervorgehoben.

Eine Möglichkeit, die Zahl der Beobachtungen zu erhöhen, ist der Verzicht auf eine länderspezifische Auswertung. Eine Paneldatenanalyse mit dem jeweiligen Land als Panelvariable und dem Jahr als Zeitvariable führt zu dem in Abbildung 5 dargestellten Ergebnis. Die Berechnung ist als Fixed-Effects-Modell in der Gestalt von (5) durchgeführt worden, wobei A wieder das tatsächlich eingetreten Wachstum des BIP in Prozentpunkten und P die entsprechende Konsensprognose darstellen. t sei wieder der Zeitpunkt und i ist jetzt der jeweils betrachtete G7-Staat. Analog zu (4) erfolgt der Test der Gemeinschaftshypothese einer Konstanten (α) von 0 und des geschätzten Parameters der Prognosen (β) von 1.

$$(5) \quad A_{i,t} = \alpha + \beta P_{i,t} + u_{i,t}$$

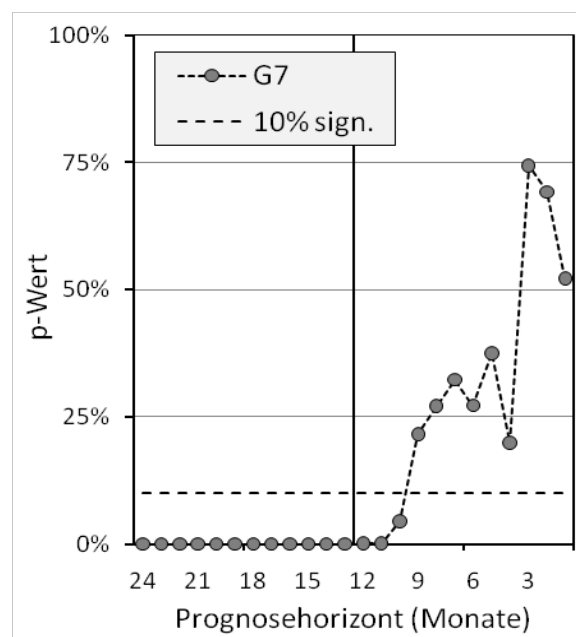


Abbildung 4: p-Werte der Panelregression des Tests auf Unverzerrtheit

Die Vergrößerung der Freiheitsgrade durch den Verzicht auf eine länderspezifische Auswertung führt zu einer Verdeutlichung der vorangegangenen Ergebnisse. Die systematischen Prognosefehler sind für die Gesamtheit der G7-Staaten bei einem Prognosehorizont von 11-24 Monaten mit einer Irrtumswahrscheinlichkeit von deutlich unter 1% nachweisbar (Abbildung 4). Dies schließt unverzerrte Konsensprognosen für einzelne Länder nicht aus, weist aber in aller Deutlichkeit auf die Existenz systematischer Prognosefehler in Abhängigkeit vom Prognosehorizont hin.

Die abschließende Beurteilung der Prognosequalität lässt sich in drei zeitlich Intervalle aufteilen. Den längerfristigen Prognosen mit einem Horizont > 15 Monate kann nur ein schlechtes Zeugnis ausgestellt werden. Der mittlere absolute Prognosefehler liegt hier für alle Staaten deutlich über einem Prozentpunkt und damit höher als die entsprechende durchschnittliche Fluktuation der jeweiligen Wachstumsraten des BIP. Je nach Land ist für den Untersuchungszeitraum eine mehr oder weniger ausgeprägte Überschätzung des Wachstums festzustellen. Systematische Prognosefehler lassen sich jedoch nur für die Minderheit der Konsensprognosen für die einzelnen G7-Staaten nachweisen. Betrachtet man jedoch die Prognosen für alle G7-Staaten zugleich, so ist eine systematische Verzerrung bis hin zu den Prognosen für das laufende Jahr eindeutig feststellbar.

Ab dem Herbst verbessern sich die Werte für die Prognosen für das kommende Jahr spürbar. Die Prognosefehler sinken unter die reale Fluktuation der Wachstumsraten und die Überschätzungen gehen dabei deutlich zurück. Dennoch sind systematische Prognosefehler in diesem Zeitraum nachweisbar, wenn auf eine länderspezifische Auswertung verzichtet wird.

Dies ändert sich bei den Prognosen für das laufende Jahr ab dem Frühjahr. Auch die Prognosefehler sind nun deutlich geringer. Allerdings können auch zunehmend real gemessene Werte in die Prognosen integriert werden.

Insgesamt scheinen die Prognosen erst wenige Monate vor Beginn des prognostizierten Jahres einen spürbaren Zukunftsbezug aufzuweisen. Dieses Ergebnis ist für die Konsensprognosen nicht besonders schmeichelhaft, entspricht jedoch dem aktuellen Forschungsstand.

3 Vergleich des Zukunftsbezugs

Im Folgenden steht die Frage im Vordergrund, ob sich die Konsensprognosen mit dem Zukunftsbezug anderer Informationen messen können. Hierfür erfolgt zum einen eine Analyse des OECD Composite Leading Indicators (CLI), der für die einzelnen G7-Staaten im Untersuchungszeitraum zur Verfügung steht. Zum anderen erfolgt die Hinzunahme der Ergebnisse der ifo World Economic Survey (WES), bei der weltweit Experten zu ihren Erwartungen bezüglich der zukünftigen Entwicklung der Weltwirtschaft befragt werden.

Zunächst werden die beiden Datensätze vorgestellt. Anschließend wird ihr Zukunftsbezug im Vergleich zu den Konsensprognosen untersucht. Da die Daten in unterschiedlicher Skalierung vorliegen, erfolgt eine Überprüfung ihres Zusammenhangs zur Wirtschaftsentwicklung mittels des Bestimmtheitsmaßes. Außerdem erfolgt für jeden Prognosehorizont und jedes Land ein Test der Hypothese, dass sich kein Zusammenhang zur Wirtschaftsentwicklung messen lässt.

Der CLI der OECD wurde unter dem Aspekt der Früherkennung konjunktureller Wendepunkte konstruiert und ist damit durchaus erfolgreich (Nilsson / Guidetti 2008). Für jede Volkswirtschaft erfolgt die Berechnung des CLI individuell, abhängig von der ökonomischen Signifikanz, dem zyklischen Verhalten und der Qualität der zur Verfügung stehenden Daten. Die für

die jeweilige Volkswirtschaft in die Aggregation eingehende Zahl der Zeitreihen schwankt zwischen 5 und 11, wobei keine Gewichtung erfolgt.⁵ Neben quantitativen Daten wie bspw. den Auftragseingängen in der Industrie, der Zinsstruktur, den Baugenehmigungen oder PKW-Neuzulassungen gehen auch landesspezifische Erhebungen zu den Erwartungen der Unternehmen und Verbraucher ein. Ziel des Indikators ist die rechtzeitige Erkennung konjunktureller Wendepunkte, wobei die frühzeitige Ermittlung des Ausmaßes der Änderung bei den Wachstumsraten als Zielsetzung keine Rolle spielt.

Die Zeitreihen sind bei der Statistiksparte auf der Webseite der OECD in verschiedenen Ausprägungen abrufbar. Für diese Untersuchung wird derjenige CLI verwendet, welcher sich auf die prozentuale Änderung der Referenzzeitreihe⁶ auf einer Basis von 12 Monaten bezieht. Zudem ist die Zeitreihe von der OECD zum jeweiligen Veröffentlichungsmonat um langfristige Trends bereinigt worden.⁷

Anders als bei den betrachteten Konsensprognosen bleibt der anvisierte Zukunftsbezug des Indikators immer konstant. Der CLI wird monatlich berechnet und mit einer Verzögerung von zwei Monaten endgültig veröffentlicht. In der Zwischenzeit erfolgen Revisionen, da nicht alle in die Aggregation zum CLI eingehenden Zeitreihen sofort verfügbar sind. Die Auswirkungen dieser Revisionen sind von Nilsson und Guidetti (2007) untersucht worden, wobei sich für den CLI für die in dieser Studie untersuchten G7-Staaten keine signifikanten Auswirkungen feststellen ließen.

Die zweite Datenreihe, die in die Untersuchung mit einbezogen wird, resultiert aus den Ergebnissen einer Erhebung durch das Münchener Institut für Wirtschaftsforschung (ifo) in Kooperation mit der Pariser International Chamber of Commerce (ICC). Dabei werden weltweit im jeweils ersten Monat eines Quartals über 1000 Experten aus 119 Ländern (Stand: Februar 2011) unter anderem zu ihrer Einschätzung der aktuellen konjunkturellen Situation und ihren Erwartungen für die nächsten sechs Monate befragt. Sie geben dabei lediglich eine Einschätzung zur Tendenz ab (besser / gleich / schlechter). Die Ergebnisse werden jeweils im 2. Monat eines Quartals unter dem Titel „World Economic Survey“ (WES) veröffentlicht.

⁵ Die genaue Zusammensetzung kann auf der Webseite der OECD nachvollzogen werden: www.oecd.org → Search: „CLI“

⁶ Für die meisten Länder ist das die Industrieproduktion, da dieser Wert im Gegensatz zum Bruttoinlandsprodukt auf monatlicher Basis vorliegt.

⁷ Vgl. OECD (2012), <http://www.oecd.org/dataoecd/26/39/41629509.pdf>

Das ifo-Institut transformiert die qualitativen Informationen folgendermaßen in Zeitreihen: Zunächst werden die beiden Aussagen mit Zahlen klassifiziert (besser = 9, gleich = 5, schlechter = 1). Der jeweils landesspezifische Durchschnitt wird dann mit dem Anteil der entsprechenden Volkswirtschaft am Welthandel gewichtet.⁸ Abschließend erfolgt eine Indizierung (Jahr 2005=100) und eine Zusammenfassung der Zeitreihen zur aktuellen Situation und den Erwartungen für die nächsten 6 Monate zu einer neuen Zeitreihe, die den ifo-Weltklimaindex darstellt.⁹

Für diese Untersuchung findet die Zeitreihe für die Erwartung der Experten für die nächsten 6 Monate Verwendung. Genau wie beim CLI bleibt auch hier der mögliche Zukunftsbezug konstant. Während die Umfrage zu Beginn des 1. Quartals die Erwartungen für das erste Halbjahr abfragt, beinhaltet die Umfrage zu Beginn des 2. Quartals die Erwartungen für das 2. und 3. Quartal eines Jahres.

Der Vergleich zwischen den Konsensprognosen einerseits und dem OECD-CLI und den Ergebnissen der ifo-WES andererseits ist ob der unterschiedlichen zeitlichen Zielsetzung der Vorhersageanstrengungen nicht ohne eine methodische Besonderheit möglich. Die Erläuterung der Vorgehensweise erfolgt anhand des folgenden Beispiels:

In der in der Mitte des Monats Januar erscheinenden Ausgabe 1/2000 von Consensus Forecasts werden – wie bereits im 2. Abschnitt erörtert – zwei Konsensprognosen für das Bruttoinlandsprodukt der G7-Staaten veröffentlicht. Eine bezieht sich auf das laufende, die andere auf das kommende Jahr (Tabelle 6, 2. + 3. Spalte). Sie weisen bis zum endgültigen Feststehen des prognostizierten Bruttoinlandsprodukts einen Horizont von 12 bzw. 24 Monaten auf. Nun werden der OECD-CLI des Januars 2000 und die Ergebnisse der ifo-WES des 1. Quartals für einen Vergleich mit den Konsensprognosen für beide Horizonte herangezogen. Die Daten der OECD und des ifo-Instituts werden also jeweils zweimal verwendet (Tabelle 6, 4. + 5. Spalte).

⁸ Plenk u. a. (2011), S. 2.

⁹ Plenk u. a. (2011), S. 2.

Tabelle 6: (Unterstellter) Abstand zwischen Erstellung und Gültigkeitszeitpunkt der Daten in Monaten (Prognosehorizont)

Ausgabe / Erhebungsmonat	Consensus Forecasts		OECD CLI	ifo WES
	lfd. Jahr	kom. Jahr		
Januar	12	24	12 / 24	12 / 24
Februar	11	23	11 / 23	-
März	10	22	10 / 22	-
April	9	21	9 / 21	9 / 21
Mai	8	20	8 / 20	-
Juni	7	19	7 / 19	-
Juli	6	18	6 / 18	6 / 18
August	5	17	5 / 17	-
September	4	16	4 / 16	-
Oktober	3	15	3 / 15	3 / 15
November	2	14	2 / 14	-
Dezember	1	13	1 / 13	-

Entgegen ihrem ursprünglichen – sich für jede Veröffentlichung konstant in die Zukunft verschiebendem – Geltungsbereich wird ihnen also ein fixer Bezug zum Wirtschaftswachstum des laufenden und kommenden Jahres unterstellt. Dies ist bei der Betrachtung der Untersuchungsergebnisse von besonderer Bedeutung.

Zunächst erfolgt ein Vergleich der Bestimmtheitsmaße der für jede der drei Vorhersageansätze, jedes Land und jeden Prognosehorizont einzeln durchgeführten Schätzung (6).

$$(6) \quad A_t = \alpha + \beta P_t + u_t$$

A ist hierbei das Wirtschaftswachstum des prognostizierten Jahres, P entweder die entsprechende Konsensprognose, der OECD CLI oder die Ergebnisse der ifo WES und t der jeweilige Zeitpunkt, welcher sich aus dem Untersuchungszeitraum 1991 – 2009 ergibt (19 Beobachtungen). Bei 7 Ländern, drei untersuchten Zeitreihen und 24 betrachteten Prognosehorizonten (8 für die ifo WES) ergeben sich also 392 einzeln durchgeführte Schätzungen.

Die Bestimmtheitsmaße geben Auskunft darüber, wie viel Prozent der Varianz des Wirtschaftswachstums durch die Vorhersageansätze erfasst wurde. Anschließend werden die Ergebnisse des Tests der Hypothese $\beta=0$ aus (5) präsentiert. Dies geschieht in Form einer grafischen Präsentation der p-Werte, also der Wahrscheinlichkeit, dass diese Hypothese zu Unrecht abgelehnt

wird. Bei Werten kleiner 0,05 (Signifikanzniveau 5%) wird davon ausgegangen, dass mit hoher Sicherheit ein Zukunftsbezug des Vorhersageansatzes zu diesem Zeitpunkt vorliegt. Die Tabellen mit den Werten sind nach Ländern sortiert dem Anhang zu entnehmen.

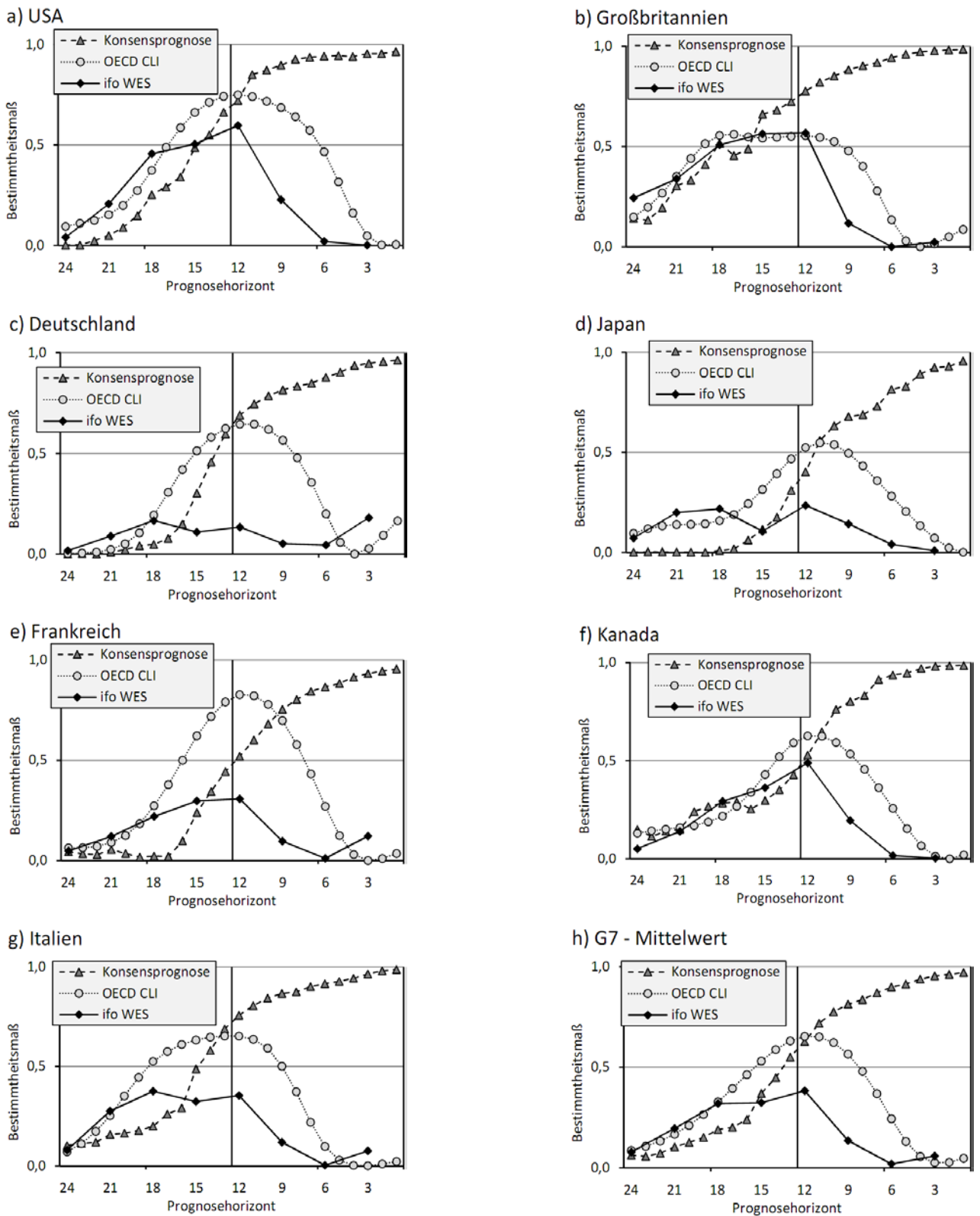


Abbildung 5: Bestimmtheitsmaße der verschiedenen Vorhersageansätze

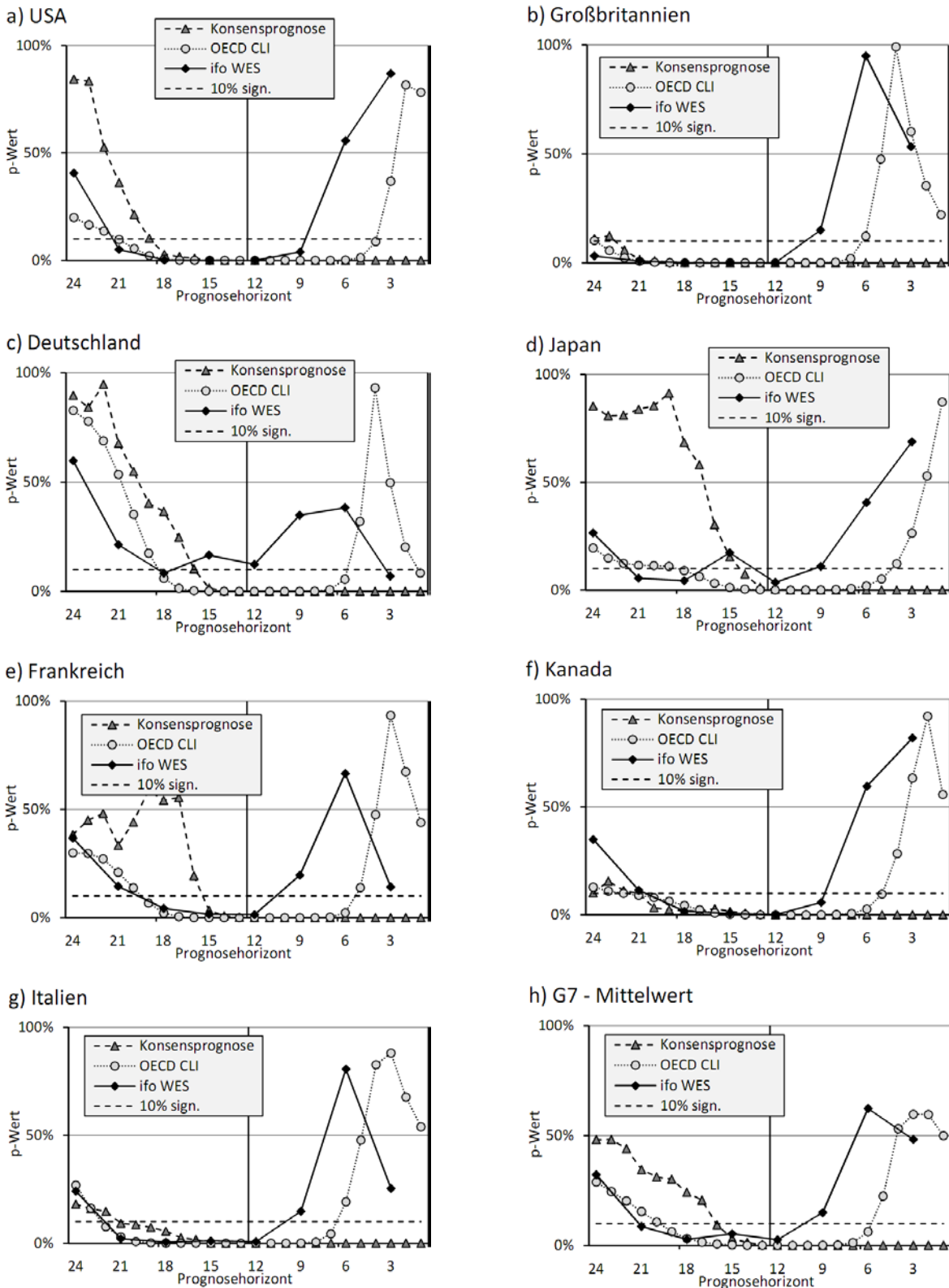


Abbildung 6: p-Werte für die Hypothese, dass kein signifikanter Bezug zum BIP messbar ist

Bei der Betrachtung der Bestimmtheitsmaße (R^2) in Abbildung 5 zeigen sich große Gemeinsamkeiten zwischen den einzelnen G7-Staaten. Nach der Analyse der Prognosequalität im 2. Abschnitt ist es wenig überraschend, dass die Konsensprognosen bei den hohen Prognosehorizonten ein geringes R^2 aufweisen, welches nach zunächst langsamem Anstieg ab einem Prognosehorizont von 15 Monaten deutlich zulegt. Bei den Prognosen, die für das Wirtschaftswachstum des laufenden Jahres erstellt wurden (Prognosehorizont < 13 Monate), lässt sich fast durchweg ein $R^2 > 0,5$ feststellen. Für die USA, Großbritannien, Kanada und Italien können schon bei den längeren Prognosehorizonten Bestimmtheitsmaße von deutlich über 0 festgestellt werden. Bei den Prognosen für die japanische Volkswirtschaft hingegen ist bei einem Prognosehorizont von 12 Monaten erst ein Bestimmtheitsmaß von 0,4 messbar. Korrespondierend hierzu fallen die Ergebnisse in Abbildung 7 aus. Mit Ausnahme der Prognosen für Japan ist der Zusammenhang zur Wirtschaftsentwicklung bei einem Horizont von 15 Monaten für alle Staaten auf 5%-Niveau signifikant – und bleibt dies erwartungsgemäß bis zu einem Horizont von einem Monat. Spiegelbildlich schlägt sich auch hier die vermeintlich¹⁰ bessere Prognoseperformance für die USA, Großbritannien, Kanada und Italien wieder. Die niedrige Irrtumswahrscheinlichkeit von 5% wird in diesen drei Staaten teilweise deutlich früher erreicht.

Die Bestimmtheitsmaße für den OECD CLI und die Ergebnisse der ifo WES Verhalten sich bei den größeren Prognosehorizonten zunächst ähnlich (Abbildung 5). Auch hier ist für fast alle Staaten ausgehend von einem niedrigen Niveau nahe 0 (Horizont = 24 Monate) ein kontinuierlicher Anstieg zu verzeichnen. Dieser erfolgt allerdings früher als bei den Konsensprognosen. So ist bei einem Prognosehorizont von 21 Monaten für alle Staaten außer Kanada gleichermaßen festzuhalten, dass die Bestimmtheitsmaße der untersuchten OECD- und ifo-Zeitreihen größer als die der zum gleichen Zeitpunkt abgegebenen Prognosen sind. Dies ist bemerkenswert, da die Zeitreihen der OECD und des ifo-Instituts nicht mit dem Ziel der Vorhersage des Wirtschaftswachstums im kommenden Jahr konstruiert worden sind. Für die ifo-Zeitreihe kommt erschwerend hinzu, dass die Ergebnisse der Expertenumfrage nicht länderspezifisch ausgewertet wurden, d. h. selbst die Betrachtung der Weltkonjunktur weist einen größeren Zusammenhang zur Wirtschaftsentwicklung in den einzelnen Ländern auf, als die entsprechenden landes-

¹⁰ „Vermeintlich“ deshalb, weil sich bei der Auswertung der Prognosequalität mittels einfacher Prognosefehlermaße im 2. Abschnitt keine Hinweise auf eine bessere Qualität ergeben.

spezifischen Prognosen der Konjunkturoxperten. Auch für einen Prognosehorizont von 18 Monaten lässt sich der überraschende Befund des stärkeren Zukunftsbezugs der beiden Zeitreihen für fast alle Staaten feststellen.

Der Zusammenhang der untersuchten Zeitreihen bei einem Prognosehorizont von 15 Monaten fällt unterschiedlich aus. Das R^2 des OECD CLI steigt in dieser Phase weiterhin für alle Staaten an und liegt außer für Großbritannien immer oberhalb des Bestimmtheitsmaßes der Konsensprognosen. Anders stellt sich die Situation bei der Zeitreihe der World Economic Survey des ifo-Instituts dar. Hier ist im Mittel kein weiterer Anstieg zu verzeichnen. Dennoch liegt der Wert in 3 Staaten immer noch oberhalb des Werts der entsprechenden Konsensprognosen.

Die Betrachtung der p-Werte in Abbildung 6 bestätigt den frühen Zukunftsbezug der OECD- und ifo-Zeitreihen. Im Mittel weisen diese bereits einige Monate vor den Konsensprognosen einen signifikanten Zusammenhang zur Wirtschaftsentwicklung des Folgejahres auf. Bei den einstelligen Prognosehorizonten verlieren die Zeitreihen der OECD und des ifo-Instituts deutlich an Kraft. Allerdings geschieht dies erwartungsgemäß, da sich der Geltungszeitpunkt der Zeitreihen mit jeder Veröffentlichung weiter in die Zukunft verschiebt. Sie beziehen sich also zunehmend auf das folgende und nicht mehr auf das laufende Jahr, was den abnehmenden Zusammenhang zur Wirtschaftsentwicklung des laufenden Jahres plausibel macht. Die schlechte Performance der ifo-Zeitreihe bei einem Horizont von 9 Monaten ist dennoch bemerkenswert, da die befragten Experten hier ihre Erwartung für das nächste halbe Jahr äußern. Die Werte für Prognosehorizonte von 12-18 Monaten sind deutlich besser, so dass die Erwartungen scheinbar erst mit einer deutlich größeren Verzögerung als 6 Monate einen Bezug zur tatsächlichen Entwicklung aufweisen.

4 Informationsgewinn

In diesem Abschnitt soll gezeigt werden, dass die Konjunkturprognosen durch den OECD-CLI und die Ergebnisse der ifo WES tatsächlich profitieren könnten. Wenn dies der Fall ist, sind die Prognosen nicht effizient. Seit Nordhaus (1987) wird eine Unterscheidung zwischen starker und schwacher Prognoseeffizienz vorgenommen. Starke Effizienz liegt vor, wenn bei der Erstellung der Prognosen sämtliche verfügbaren Informationen optimal verarbeitet wurden. Dies schließt die Kenntnisse um die ökonomischen Zusammenhänge und auch Insider-Info-

nen mit ein. Nordhaus verweist an dieser Stelle auf die Parallelität zur Effizienz an den Kapitalmärkten (Fama 1970). Die starke Effizienz ließe sich in der Praxis jedoch nicht erwarten, da die Ökonomen keinen Zugriff auf Insiderinformation hätten.

Als Alternative schlägt Nordhaus den Test auf schwache Effizienz vor. Hier wird von den Prognostikern erwartet, zumindest ihre eigenen Prognosefehler zu analysieren und daraus die richtigen Schlussfolgerungen zu ziehen. Dieser Test wird bei dieser Untersuchung nicht angewendet, da der Untersuchungszeitraum zu knapp bemessen ist, als dass sich vergangene Prognoserevisionen sinnvoll auswerten ließen.

An dieser Stelle sei wieder auf die Theorie der informationseffizienten Kapitalmärkte von Fama verwiesen. Famas Theorie beinhaltet eine Stufe der halbstrengen Effizienz, in der alle öffentlich verfügbaren Informationen zu einer auf dieser Stufe effizienten Kursbildung beitragen. Analog hierzu lässt sich die Nutzung der Zeitreihen der OECD und des ifo-Instituts für die Konjunkturprognosen betrachten. Beide Informationen sind öffentlich verfügbar. So gesehen ist hier eine weitergehende Differenzierung erforderlich. In dieser Studie wird gezeigt, dass die Prognosen zu den meisten Zeitpunkten nicht über halbstarke Effizienz verfügen.

Zur Überprüfung, ob die Informationen zu einer Verbesserung der Konjunkturprognosen beitragen könnten, wird die bereits im 2. Abschnitt vorgestellte Mincer-Zarnowitz-Regression erweitert (Holden / Peel 1990). Sei X eine beliebige Information zum Zeitpunkt t . A ist wieder der real gemessene Wert, P die entsprechende Prognose und u der Fehlerterm.

$$(7) \quad A_t = \alpha + \beta P_t + \gamma X_t + u_t$$

Um zu prüfen, ob die Zusatzinformation zu einer besseren Prognose beigetragen hätte, erfolgt der Test der Hypothese $\gamma=0$.

Bezogen auf diese Untersuchung ist A die Wachstumsrate des Bruttoinlandsprodukts auf Jahresbasis, P die entsprechende Konsensprognose, und für X werden einmal die Zeitreihe des OECD Composite Leading Indicators und einmal die Ergebnisse der ifo World Economic Survey eingesetzt. Der Untersuchungszeitraum umfasst wieder die Jahre 1991-2009 und es erfolgt eine individuelle Berechnung für jeden G7-Staat, wobei bei den Erwartungen für die Weltkonjunktur der ifo WES für jeden Staat die gleiche Datenreihe eingeht.

An dieser Stelle müssen die Zeitpunkte der Prognoseerstellung bzw. Datenerhebung thematisiert werden. Damit die Konjunkturprognosen von den Daten der OECD und des ifo-Instituts

profitieren können, müssen diese rechtzeitig vorliegen. Die Prognosen werden jeweils in der Monatsmitte veröffentlicht, so dass bis zum Monatsanfang Änderungen vorgenommen werden können. Die Aggregation des OECD CLI basiert auf den im Laufe eines Monats gewonnenen quantitativen und qualitativen Daten. Die Veröffentlichung erfolgt erst im Folgemonat. Die Prognostiker können den OECD CLI also erst einen Monat später berücksichtigt haben. Für die in Consensus Forecasts erschienenen Prognosen im Februar wird also überprüft, ob der CLI des Monats Januar ausreichend berücksichtigt wurde.

Die weltweite Expertenbefragung des ifo-Instituts erfolgt jeweils im ersten Monat eines Quartals. Die entsprechende Pressemitteilung mit den Ergebnissen wird zu Beginn des Folgemonats veröffentlicht. Um den Prognostikern für die Verarbeitung dieser Informationen ebenfalls ausreichend Zeit einzuräumen, wird auch hier ein LAG von einem Monat bei der Berechnung berücksichtigt.

Untersucht wird die Frage, ob sich die Prognosen für das kommende Jahr mit den beiden Zusatzinformationen verbessern lassen. Es werden die Prognosehorizonte 23 (Februar), 20 (Mai), 17 (August) und 14 (November) betrachtet. Tabelle 7 liefert eine Übersicht der untersuchten Prognosehorizonte der Konsensprognosen und die dazugehörigen Monate der jeweils eingehenden Zeitreihen der OECD und des ifo-Instituts.

Tabelle 7: Übersicht über den zeitlichen Eingang der Daten in den Test auf Nutzen der Zusatzinformationen

Monat der Veröffentlichung der Prognosen		Monat der Erhebung der Zusatzinformationen	
Consensus Forecasts	Prognosehorizont	OECD CLI	ifo WES
Februar	23 Monate	Januar	Januar
Mai	20 Monate	April	April
August	17 Monate	Juli	Juli
November	14 Monate	Oktober	Oktober

Wenn sich der geschätzte Parameter γ aus Gleichung (7) signifikant von 0 unterscheidet, so hätte die Zusatzinformation mit hoher Wahrscheinlichkeit zu einer Verbesserung der Konjunkturprognosen für das jeweilige Land beigetragen. Die Ergebnisse der Schätzungen sind den Tabellen 8 und 9 zu entnehmen.

Tabelle 8: Regressionsergebnisse inkl. der Zusatzinformation OECD CLI

	PH	(1) USA	(2) UK	(3) DE	(4) JA	(5) FR	(6) KA	(7) IT
Konsens- prognose		0,390 (0,642)	1,731 (0,327)	-0,182 (0,838)	0,068 (0,887)	0,515 (0,581)	0,825 (0,384)	0,942 (0,338)
OECD CLI		0,164 (0,189)	0,260 (0,264)	0,026 (0,825)	0,146 (0,214)	0,102 (0,377)	0,118 (0,307)	0,055 (0,652)
Konstante	23 Monate	1,035 (0,680)	-2,724 (0,525)	1,577 (0,443)	0,677 (0,556)	0,187 (0,936)	-0,554 (0,843)	-1,318 (0,549)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,107	0,200	0,006	0,098	0,081	0,172	0,125
Konsens- prognose		0,978 (0,199)	2,167* (0,093)	0,414 (0,614)	0,078 (0,869)	0,157 (0,862)	1,311 (0,119)	0,626 (0,429)
OECD CLI		0,195* (0,096)	0,396* (0,071)	0,063 (0,598)	0,179 (0,127)	0,114 (0,329)	0,098 (0,357)	0,180 (0,126)
Konstante	20 Monate	-0,666 (0,755)	-3,763 (0,220)	0,224 (0,902)	0,663 (0,545)	1,093 (0,626)	-2,024 (0,409)	-0,598 (0,726)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,239	0,459	0,039	0,141	0,093	0,281	0,282
Konsens- prognose		1,395** (0,022)	1,295* (0,096)	0,534 (0,428)	0,322 (0,515)	-0,532 (0,402)	1,262 (0,139)	0,423 (0,473)
OECD CLI		0,280*** (0,007)	0,438** (0,015)	0,187 (0,100)	0,175* (0,092)	0,264** (0,021)	0,092 (0,406)	0,263*** (0,007)
Konstante	17 Monate	-1,920 (0,218)	-1,441 (0,413)	-0,131 (0,927)	0,278 (0,780)	2,706* (0,085)	-1,766 (0,461)	-0,084 (0,945)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,554	0,629	0,226	0,182	0,306	0,321	0,541
Konsens- prognose		0,531* (0,085)	1,430** (0,012)	0,913** (0,042)	0,652 (0,116)	0,040 (0,921)	0,395 (0,508)	0,920* (0,052)
OECD CLI		0,248*** (0,006)	0,120 (0,436)	0,194** (0,016)	0,205** (0,021)	0,303*** (0,003)	0,187 (0,120)	0,182** (0,016)
Konstante	14 Monate	0,771 (0,237)	-1,349 (0,236)	-0,565 (0,462)	-0,065 (0,928)	1,404 (0,105)	0,812 (0,594)	-0,737 (0,361)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,721	0,694	0,627	0,416	0,622	0,445	0,712

Abhängige Variable: Entwicklung des Bruttoinlandsprodukt in den jeweiligen Staaten auf Jahresbasis im Vergleich zum Vorjahr in Prozent; PH = Prognosehorizont; CLI = Composite Leading Indicator (signifikante Werte hervorgehoben); p-Werte in Klammern (***) $p < 0,01$; ** $p < 0,05$, * $p < 0,1$)

Tabelle 9: Regressionsergebnisse der Konsensprognosen inkl. der Zusatzinformation ifo WES

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
	PH	USA	UK	DE	JA	FR	KA	IT
Konsens- prognose		-0,080 (0,930)	0,534 (0,792)	-0,077 (0,932)	0,195 (0,691)	0,862 (0,353)	1,079 (0,267)	1,197 (0,144)
ifo WES		0,023 (0,433)	0,058 (0,138)	0,015 (0,630)	0,038 (0,259)	0,022 (0,296)	0,013 (0,677)	0,032 (0,211)
Konstante	23 Monate	0,555 (0,859)	-5,221 (0,184)	-0,097 (0,981)	-3,200 (0,371)	-2,732 (0,414)	-2,345 (0,469)	-4,917 (0,119)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,041	0,248	0,017	0,082	0,099	0,124	0,198
Konsens- prognose		0,228 (0,802)	1,994 (0,183)	0,533 (0,496)	0,039 (0,933)	0,383 (0,626)	1,390 (0,143)	0,949 (0,157)
ifo WES		0,055 (0,138)	0,055 (0,165)	0,045 (0,208)	0,071* (0,064)	0,032 (0,195)	0,018 (0,633)	0,059** (0,040)
Konstante	20 Monate	-3,613 (0,251)	-8,840** (0,013)	-4,513 (0,272)	-6,425 (0,104)	-2,671 (0,347)	-3,927 (0,226)	-7,184** (0,019)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,210	0,411	0,117	0,200	0,134	0,251	0,363
Konsens- prognose		0,313 (0,723)	1,359 (0,113)	0,782 (0,234)	0,250 (0,601)	0,159 (0,777)	0,993 (0,261)	1,090** (0,045)
ifo WES		0,082** (0,040)	0,075** (0,043)	0,061* (0,084)	0,076* (0,051)	0,047* (0,057)	0,047 (0,247)	0,069** (0,010)
Konstante	17 Monate	6,565** (0,021)	9,104*** (0,003)	-6,532* (0,086)	7,282* (0,070)	-3,616 (0,165)	5,567* (0,078)	8,401*** (0,003)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,461	0,582	0,239	0,232	0,224	0,349	0,518
Konsens- prognose		0,763 (0,111)	1,370** (0,017)	1,435*** (0,003)	0,746 (0,116)	0,800** (0,027)	0,571 (0,426)	1,489*** (0,000)
ifo WES		0,029 (0,301)	0,025 (0,371)	0,022 (0,257)	0,029 (0,271)	0,029* (0,052)	0,036 (0,340)	0,034** (0,025)
Konstante	14 Monate	-1,961 (0,287)	-3,513* (0,060)	-3,374* (0,080)	-2,929 (0,240)	2,765** (0,043)	-2,702 (0,213)	4,829*** (0,001)
Beobach- tungen		19 (Zeitraum 1991 - 2009)						
R ²		0,580	0,698	0,501	0,238	0,487	0,388	0,697

Abhängige Variable: Entwicklung des Bruttoinlandsprodukt in den jeweiligen Staaten auf Jahresbasis im Vergleich zum Vorjahr in Prozent; PH = Prognosehorizont; ifo WES = World Economic Survey des ifo-Instituts (signifikante Werte hervorgehoben); p-Werte in Klammern (***) p<0,01; ** p<0,05, * p<0,1)

Bei einem Prognosehorizont von 23 Monaten sind die untersuchten Daten der OECD und des ifo-Instituts nicht hilfreich.

Dies ändert sich bei einem Prognosehorizont von 20 Monaten. Der Parameter für den OECD CLI wird für die USA und Großbritannien signifikant (10%-Niveau) ungleich 0 geschätzt. Die Konjunkturprognosen für zwei der bedeutendsten Volkswirtschaften der Welt könnten sich also

unter Zuhilfenahme des OECD CLI verbessern lassen. Die vom ifo-Institut erfasste Experteneinschätzung bezüglich der Entwicklung der Weltwirtschaft zeigt sich bei einem Prognosehorizont der Konsensprognosen von 20 Monaten für Japan (10%-Niveau) und Italien (5%-Niveau) signifikant.

Die Konjunkturprognosen, die im August für das folgende Jahr abgegeben werden (Horizont: 17 Monate), könnten in fast allen G7-Staaten von den hier getesteten Zusatzinformationen profitieren. Der OECD CLI zeigt sich in 5 von 7 Staaten zumindest auf 10%-Niveau signifikant, für die ifo-Zeitreihe gilt dies in 6 von 7 Staaten. Lediglich für Kanada ist bei beiden Zeitreihen keine Signifikanz feststellbar, was jedoch nicht mit erfolgreichen Konjunkturprognosen gleichzusetzen ist. Der geschätzte Parameter für die Konsensprognosen leistet bei einem Prognosehorizont von 17 Monaten in fast allen Fällen keinen signifikanten Beitrag zur Erklärung der Varianz der Wirtschaftsentwicklung in den betrachteten Ländern.

Die im November abgegebenen Prognosen für das kommende Jahr (Horizont 14 Monate) gelingen zwar insgesamt recht gut (siehe Abschnitt 2), aber auch hier könnten die Zusatzinformationen noch zu einer Verbesserung beitragen. Der OECD CLI zeigt sich für 5 der 7 Staaten zumindest auf einem 5%-Niveau signifikant, die vom ifo-Institut erhobenen Daten in 2 der 7 Staaten zumindest auf einem Niveau von 10%.

5 Schlussfolgerungen

In dieser Untersuchung wird gezeigt, dass bei der Erstellung von Konjunkturprognosen für die G7-Staaten nicht alle Mittel für die Erreichung einer guten Prognosequalität ausgeschöpft werden. Es werden relevante öffentlich verfügbare Informationen nicht oder nur unzureichend berücksichtigt. Mit dem Composite Leading Indicator der OECD und der Expertenbefragung World Economic Survey des Münchener ifo-Instituts sind zwei Zeitreihen hinsichtlich ihrer Eignung hierfür getestet worden. Beide Informationsquellen sind öffentlich ohne weiteres zugänglich. Zudem wurde den Prognostikern durch das Untersuchungsdesign ausreichend Zeit für eine Einarbeitung der Informationen in ihre Modelle zugestanden.

Als Ergebnis lässt sich folgendes festhalten:

- Die untersuchten Konjunkturprognosen weisen erst ab einem Prognosehorizont von 15 Monaten einen deutlichen Zukunftsbezug auf.

- Als eine mögliche Ursache hierfür lässt sich eine systematische Fehleinschätzung des Wirtschaftswachstums im Untersuchungszeitraum festhalten, die vor allem auf häufigen Überschätzungen beruht.
- Obwohl dies nicht der jeweiligen Zielsetzung der Erstellung bzw. Erhebungen entspricht, weisen die Zeitreihen der OECD und des ifo-Instituts frühzeitiger einen Zukunftsbezug zur Wirtschaftsentwicklung in den G7-Staaten auf Jahresbasis auf.
- Ab einem Prognosehorizont von 17 Monaten hätten die untersuchten Zeitreihen der OECD und des ifo-Instituts in den meisten G7-Staaten zu einer Verbesserung der Konjunkturprognosen während des Untersuchungszeitraums beitragen können.

Es stellt sich die Frage, warum die Prognostiker diese Informationen bisher nicht ausreichend zur Kenntnis genommen haben. Folgende Ursachen könnten hierzu geführt haben:

- Die Informationen beziehen sich auf die jeweils nächsten Monate und nicht auf die Wirtschaftsentwicklung eines ganzen Jahres. Vielleicht führt diese Tatsache zur Nichtbeachtung der Daten.
- Für die Zeitreihe des ifo-Instituts liegt kein landesspezifischer Bezug vor. Dies könnte zu einer Missachtung der Daten bei der Erstellung der landesspezifischen Konjunkturprognosen geführt haben.
- Bei den untersuchten Zusatzinformationen handelt es sich hauptsächlich um qualitativ gewonnene Daten – die vorwiegend Erwartungen widerspiegeln. Möglicherweise wird der Zukunftsbezug der Erwartungen der Wirtschaftssubjekte generell von den Prognostikern unterschätzt.

Mit den Befunden aus dieser Studie als Ausgangspunkt sollten weitere Forschungsanstrengungen unternommen werden. Die Möglichkeiten der Aufbereitung der Informationen müssten mit dem Ziel ergründet werden, die Prognosemodelle so weit wie möglich zu verbessern. Denkbar wären hier unterschiedliche Gewichtungen auf Basis des Prognosehorizonts oder eine Überarbeitung der Datenerhebungen mit der Zielsetzung des Vorlaufs für die Wirtschaftsentwicklung auf Jahresbasis.

Die Ursachenforschung für die ungenügende Berücksichtigung der Informationen sollte die qualitative Form der Datenerhebung und hier besonders Auswirkungen der Erwartungen der Wirtschaftssubjekte im Fokus haben. Es müsste getestet werden, ob quantitative Frühindikatoren im Vergleich besser von den Prognostikern berücksichtigt werden.

Weiterhin ist es von Interesse zu erfahren, ob individuelle Prognostiker die Informationen besser berücksichtigen. In dieser Studie erfolgte lediglich eine Betrachtung der Konsensprognosen ohne individuelle Zurechnung.

Literatur

- Abberger, K. & Nierhaus, W. (2007): Das Ifo Geschäftsklima und Wendepunkte der deutschen Konjunktur. Ifo-Schnelldienst, 60, 26-31.
- Ager, P., Kappler, M. & Osterloh, S. (2009): The accuracy and efficiency of the Consensus Forecasts: a further application and extension of the pooled approach. International Journal of Forecasting, 25, 167-181.
- Batchelor, R. (2001): How useful are the forecasts of intergovernmental agencies? The IMF and OECD versus the consensus. Applied Economics, 33, 225-236.
- Bates, J. M. & Granger, C. W. J. (1969): The Combination of Forecasts. Operational Research Quarterly, 20, 451–468.
- Blix, M., Wadefjord, J., Wienecke, U. & Ådahl, M. (2001): How good is the forecasting performance of major institutions? Sveriges Riksbank economic review, 12, 38-68.
- Claveria, O., Pons, E. & Ramos, R. (2007): Business and consumer expectations and macroeconomic forecasts. International journal of forecasting, 23, 47-69.
- Clements, M. P. & Galvão, A. B. (2008): Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. Journal of Business and Economic Statistics, 26, 546 – 554.
- Dovern, J. & Weisser, J. (2011): Accuracy, unbiasedness and efficiency of professional macroeconomic forecasts: an empirical comparison for the G7. International Journal of Forecasting, 27, 452 – 465.
- Fritsche, U. (1999): Vorlaufeigenschaften von Ifo-Indikatoren für Westdeutschland. DIW Discussion Papers, 179.
- Holden, K. & Peel, D. (1990): On testing for unbiasedness and efficiency of forecasts. The Manchester School of Economic and Social Studies, 58, 120-127.
- Hüfner, F. P. & Schröder, M. (2002): Prognosegehalt von Ifo-Geschäftserwartungen und ZEW-Konjunkturerwartungen: Ein ökonometrischer Vergleich. Journal of Economics and Statistics, 222, 316-336.
- Isiklar, G. & Lahiri, K. (2007): How far ahead can we forecast? Evidence from cross-country surveys. International Journal of Forecasting, 23, 167-187.
- Juglar, C. (1862): Des crises commerciales et de leur retour périodique en France, en Angleterre et aux états-Unis, Paris: Guillaumin et Cie.

- Kirchgässner, G. (1982): Sind die Erwartungen der Wirtschaftssubjekte "rational"? Eine empirische Untersuchung für die Bundesrepublik Deutschland. *Review of World Economics*, 118, 215-240.
- Kunkel, A. (2003): Zur Prognosefähigkeit des Ifo Geschäftsklimas und seiner Komponenten sowie der Überprüfung der "Dreimal-Regel". *Ifo-Diskussionsbeiträge*, 80.
- Laplace, P. (1814): *Essai philosophique sur les probabilités*, Paris: Courcier.
- Loungani, P. (2001): How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth. *International Journal of Forecasting*, 17, 419-432.
- Mincer, J. & Zarnowitz, V. (1969): The Evaluation of Economic Forecasts. In: C. Mincer (Ed.), *Economic Forecasts and Expectation*. New York: Columbia University Press, pp. 3-46.
- Muth, J. F. (1961): Rational expectations and the theory of price movements. *Econometrica*, 29, 315-334.
- Nilsson, R. & Guidetti, E. (2007): Current period performance of OECD Composite Leading Indicators (CLIs). *Journal of Business Cycle Measurement and Analysis*, 3, 235-266.
- Nilsson, R. & Guidetti, E. (2008): Predicting the Business Cycle - How good are early estimates of OECD Composite Leading Indicators? *OECD Statistics Brief*, 14.
- Nordhaus, W. D. (1987): Forecasting efficiency: concepts and applications. *The review of economics and statistics*, 69, 667-674.
- OECD (2012): OECD System of Composite Leading Indicators, <http://www.oecd.org/dataoecd/26/39/41629509.pdf>, assessed on 27 May 2012.
- Osterloh, S. (2008): Accuracy and Properties of German Business Cycle Forecasts. *Applied Economics Quarterly*, 54, 27-57.
- Plenk, J., Nerb, G., Abberger, K. & Wollmershäuser, T. (2011): CESIfo world economic survey. 10-1.
- Seiler, C. (2009): Prediction qualities of the Ifo indicators on a temporal disaggregated German GDP. *Ifo Working Paper*, 67.
- Stavenhagen, G. (1969): *Geschichte der Wirtschaftstheorie*, Göttingen: Vandenhoeck & Ruprecht, 4th edn.
- Theil, H. (1971): *Principles of Econometrics*, New York: John Wiley & Sons, 2nd edn.
- Ziegler, C. (2009): Testing Predictive Ability of Business Cycle Indicators for the Euro Area. *Ifo Working Paper*, 69.

Anhang

USA

Tabelle 10: Bestimmtheitsmaße und p-Werte der drei Vorhersageansätze bezogen auf die Wachstumsrate in den USA von 1991-2009

PH Mon.	Bestimmtheitsmaß			p-Werte		
	Prognosen	OECD CLI	ifo WES	Prognosen	OECD CLI	ifo WES
24	0,00	0,09	0,04	0,843	0,199	0,406
23	0,00	0,11	-	0,833	0,165	-
22	0,02	0,13	-	0,526	0,137	-
21	0,05	0,15	0,21	0,361	0,097	0,051
20	0,09	0,20	-	0,213	0,056	-
19	0,15	0,27	-	0,103	0,022	-
18	0,25	0,37	0,46	0,028	0,005	0,002
17	0,29	0,49	-	0,017	0,001	-
16	0,34	0,59	-	0,009	0,000	-
15	0,49	0,66	0,50	0,001	0,000	0,001
14	0,55	0,71	-	0,000	0,000	-
13	0,66	0,74	-	0,000	0,000	-
12	0,72	0,75	0,60	0,000	0,000	0,000
11	0,85	0,74	-	0,000	0,000	-
10	0,87	0,72	-	0,000	0,000	-
9	0,90	0,69	0,23	0,000	0,000	0,039
8	0,93	0,64	-	0,000	0,000	-
7	0,94	0,57	-	0,000	0,000	-
6	0,94	0,47	0,02	0,000	0,001	0,557
5	0,94	0,32	-	0,000	0,012	-
4	0,94	0,16	-	0,000	0,088	-
3	0,95	0,05	0,00	0,000	0,368	0,869
2	0,95	0,00	-	0,000	0,816	-
1	0,97	0,00	-	0,000	0,782	-

PH = Prognosehorizont; Abhängige Variable: Wachstum des Bruttoinlandsprodukt von 1991-2009 in Prozent; Unabhängige Variable: jeweils einzeln Konsensprognosen / OECD CLI / ifo WES; Anzahl der Beobachtungen: jeweils 19 (Zeitraum 1991-2009); p-Werte: Wahrscheinlichkeit, dass irrtümlich ein Zukunftsbezug des Vorhersageansatzes unterstellt wird, *** p<0,01 / ** p<0,05 / * p<0,1.

Großbritannien

Tabelle 11: Bestimmtheitsmaße und p-Werte der drei Vorhersageansätze bezogen auf die Wachstumsrate in Großbritannien von 1991-2009

PH Mon.	Bestimmtheitsmaß			p-Werte		
	Prognosen	OECD CLI	ifo WES	Prognosen	OECD CLI	ifo WES
24	0,14	0,15	0,24	0,111	0,102	0,031
23	0,13	0,20	-	0,124	0,056	-
22	0,20	0,27	-	0,058	0,023	-
21	0,30	0,35	0,34	0,014	0,007	0,009
20	0,33	0,44	-	0,010	0,002	-
19	0,41	0,51	-	0,003	0,001	-
18	0,52	0,56	0,51	0,001	0,000	0,001
17	0,46	0,56	-	0,002	0,000	-
16	0,49	0,55	-	0,001	0,000	-
15	0,66	0,54	0,56	0,000	0,000	0,000
14	0,68	0,55	-	0,000	0,000	-
13	0,72	0,55	-	0,000	0,000	-
12	0,78	0,55	0,57	0,000	0,000	0,000
11	0,82	0,55	-	0,000	0,000	-
10	0,85	0,52	-	0,000	0,000	-
9	0,88	0,48	0,12	0,000	0,001	0,150
8	0,90	0,40	-	0,000	0,004	-
7	0,92	0,28	-	0,000	0,020	-
6	0,94	0,13	0,00	0,000	0,122	0,950
5	0,96	0,03	-	0,000	0,476	-
4	0,97	0,00	-	0,000	0,991	-
3	0,98	0,02	0,02	0,000	0,602	0,533
2	0,98	0,05	-	0,000	0,354	-
1	0,98	0,09	-	0,000	0,221	-

PH = Prognosehorizont; Abhängige Variable: Wachstum des Bruttoinlandsprodukt von 1991-2009 in Prozent; Unabhängige Variable: jeweils einzeln Konsensprognosen / OECD CLI / ifo WES; Anzahl der Beobachtungen: jeweils 19 (Zeitraum 1991-2009); p-Werte: Wahrscheinlichkeit, dass irrtümlich ein Zukunftsbezug des Vorhersageansatzes unterstellt wird, *** $p < 0,01$ / ** $p < 0,05$ / * $p < 0,1$.

Deutschland

Tabelle 12: Bestimmtheitsmaße und p-Werte der drei Vorhersageansätze bezogen auf die Wachstumsrate in Deutschland von 1991-2009

PH Mon.	Bestimmtheitsmaß			p-Werte		
	Prognosen	OECD CLI	ifo WES	Prognosen	OECD CLI	ifo WES
24	0,00	0,00	0,02	0,897	0,827	0,597
23	0,00	0,00	-	0,842	0,778	-
22	0,00	0,01	-	0,947	0,688	-
21	0,01	0,02	0,09	0,677	0,535	0,213
20	0,02	0,05	-	0,548	0,352	-
19	0,04	0,11	-	0,403	0,174	-
18	0,05	0,19	0,17	0,365	0,059	0,083
17	0,08	0,31	-	0,247	0,014	-
16	0,15	0,42	-	0,102	0,003	-
15	0,30	0,51	0,11	0,015	0,001	0,166
14	0,46	0,58	-	0,001	0,000	-
13	0,60	0,62	-	0,000	0,000	-
12	0,69	0,64	0,13	0,000	0,000	0,123
11	0,75	0,65	-	0,000	0,000	-
10	0,79	0,62	-	0,000	0,000	-
9	0,82	0,57	0,05	0,000	0,000	0,348
8	0,83	0,48	-	0,000	0,001	-
7	0,85	0,36	-	0,000	0,007	-
6	0,88	0,20	0,05	0,000	0,055	0,383
5	0,90	0,06	-	0,000	0,319	-
4	0,94	0,00	-	0,000	0,931	-
3	0,95	0,03	0,18	0,000	0,497	0,069
2	0,96	0,09	-	0,000	0,203	-
1	0,96	0,17	-	0,000	0,084	-

PH = Prognosehorizont; Abhängige Variable: Wachstum des Bruttoinlandsprodukt von 1991-2009 in Prozent; Unabhängige Variable: jeweils einzeln Konsensprognosen / OECD CLI / ifo WES; Anzahl der Beobachtungen: jeweils 19 (Zeitraum 1991-2009); p-Werte: Wahrscheinlichkeit, dass irrtümlich ein Zukunftsbezug des Vorhersageansatzes unterstellt wird, *** $p < 0,01$ / ** $p < 0,05$ / * $p < 0,1$.

Japan

Tabelle 13: Bestimmtheitsmaße und p-Werte der drei Vorhersageansätze bezogen auf die Wachstumsrate in Japan von 1991-2009

PH Mon.	Bestimmtheitsmaß			p-Werte		
	Prognosen	OECD CLI	ifo WES	Prognosen	OECD CLI	ifo WES
24	0,00	0,10	0,07	0,853	0,195	0,264
23	0,00	0,12	-	0,808	0,148	-
22	0,00	0,13	-	0,810	0,124	-
21	0,00	0,14	0,20	0,838	0,115	0,055
20	0,00	0,14	-	0,854	0,114	-
19	0,00	0,14	-	0,913	0,110	-
18	0,01	0,16	0,22	0,685	0,090	0,044
17	0,02	0,19	-	0,581	0,062	-
16	0,06	0,24	-	0,303	0,031	-
15	0,12	0,32	0,11	0,154	0,012	0,173
14	0,18	0,39	-	0,074	0,004	-
13	0,31	0,47	-	0,013	0,001	-
12	0,40	0,52	0,24	0,004	0,000	0,035
11	0,56	0,55	-	0,000	0,000	-
10	0,63	0,54	-	0,000	0,000	-
9	0,68	0,50	0,14	0,000	0,001	0,110
8	0,69	0,43	-	0,000	0,002	-
7	0,73	0,36	-	0,000	0,007	-
6	0,82	0,28	0,04	0,000	0,019	0,406
5	0,83	0,21	-	0,000	0,051	-
4	0,89	0,13	-	0,000	0,123	-
3	0,92	0,07	0,01	0,000	0,264	0,688
2	0,93	0,02	-	0,000	0,529	-
1	0,96	0,00	-	0,000	0,872	-

PH = Prognosehorizont; Abhängige Variable: Wachstum des Bruttoinlandsprodukt von 1991-2009 in Prozent; Unabhängige Variable: jeweils einzeln Konsensprognosen / OECD CLI / ifo WES; Anzahl der Beobachtungen: jeweils 19 (Zeitraum 1991-2009); p-Werte: Wahrscheinlichkeit, dass irrtümlich ein Zukunftsbezug des Vorhersageansatzes unterstellt wird, *** p<0,01 / ** p<0,05 / * p<0,1.

Frankreich

Tabelle 14: Bestimmtheitsmaße und p-Werte der drei Vorhersageansätze bezogen auf die Wachstumsrate in Frankreich von 1991-2009

PH Mon.	Bestimmtheitsmaß			p-Werte		
	Prognosen	OECD CLI	ifo WES	Prognosen	OECD CLI	ifo WES
24	0,05	0,06	0,05	0,383	0,299	0,367
23	0,03	0,06	-	0,450	0,297	-
22	0,03	0,07	-	0,481	0,271	-
21	0,06	0,09	0,12	0,334	0,209	0,145
20	0,04	0,13	-	0,441	0,138	-
19	0,02	0,18	-	0,590	0,067	-
18	0,02	0,27	0,22	0,542	0,021	0,043
17	0,02	0,38	-	0,555	0,005	-
16	0,10	0,50	-	0,193	0,001	-
15	0,24	0,62	0,30	0,033	0,000	0,016
14	0,34	0,72	-	0,008	0,000	-
13	0,44	0,79	-	0,002	0,000	-
12	0,52	0,83	0,31	0,000	0,000	0,014
11	0,60	0,82	-	0,000	0,000	-
10	0,68	0,78	-	0,000	0,000	-
9	0,75	0,70	0,10	0,000	0,000	0,196
8	0,80	0,58	-	0,000	0,000	-
7	0,84	0,43	-	0,000	0,002	-
6	0,87	0,27	0,01	0,000	0,022	0,666
5	0,88	0,12	-	0,000	0,139	-
4	0,91	0,03	-	0,000	0,476	-
3	0,93	0,00	0,12	0,000	0,933	0,142
2	0,94	0,01	-	0,000	0,674	-
1	0,96	0,04	-	0,000	0,439	-

PH = Prognosehorizont; Abhängige Variable: Wachstum des Bruttoinlandsprodukt von 1991-2009 in Prozent; Unabhängige Variable: jeweils einzeln Konsensprognosen / OECD CLI / ifo WES; Anzahl der Beobachtungen: jeweils 19 (Zeitraum 1991-2009); p-Werte: Wahrscheinlichkeit, dass irrtümlich ein Zukunftsbezug des Vorhersageansatzes unterstellt wird, *** $p < 0,01$ / ** $p < 0,05$ / * $p < 0,1$.

Kanada

Tabelle 15: Bestimmtheitsmaße und p-Werte der drei Vorhersageansätze bezogen auf die Wachstumsrate in Kanada von 1991-2009

PH Mon.	Bestimmtheitsmaß			p-Werte		
	Prognosen	OECD CLI	ifo WES	Prognosen	OECD CLI	ifo WES
24	0,15	0,13	0,05	0,102	0,129	0,350
23	0,11	0,14	-	0,157	0,110	-
22	0,14	0,15	-	0,113	0,100	-
21	0,15	0,16	0,14	0,100	0,091	0,115
20	0,24	0,17	-	0,033	0,080	-
19	0,27	0,19	-	0,024	0,063	-
18	0,28	0,22	0,29	0,019	0,044	0,017
17	0,29	0,27	-	0,017	0,023	-
16	0,25	0,34	-	0,028	0,009	-
15	0,30	0,43	0,36	0,016	0,002	0,006
14	0,35	0,52	-	0,008	0,000	-
13	0,43	0,59	-	0,002	0,000	-
12	0,53	0,63	0,49	0,000	0,000	0,001
11	0,65	0,63	-	0,000	0,000	-
10	0,76	0,59	-	0,000	0,000	-
9	0,80	0,53	0,20	0,000	0,000	0,058
8	0,83	0,46	-	0,000	0,002	-
7	0,91	0,36	-	0,000	0,006	-
6	0,94	0,26	0,02	0,000	0,027	0,596
5	0,95	0,15	-	0,000	0,097	-
4	0,97	0,07	-	0,000	0,284	-
3	0,98	0,01	0,00	0,000	0,634	0,820
2	0,98	0,00	-	0,000	0,920	-
1	0,98	0,02	-	0,000	0,558	-

PH = Prognosehorizont; Abhängige Variable: Wachstum des Bruttoinlandsprodukt von 1991-2009 in Prozent; Unabhängige Variable: jeweils einzeln Konsensprognosen / OECD CLI / ifo WES; Anzahl der Beobachtungen: jeweils 19 (Zeitraum 1991-2009); p-Werte: Wahrscheinlichkeit, dass irrtümlich ein Zukunftsbezug des Vorhersageansatzes unterstellt wird, *** $p < 0,01$ / ** $p < 0,05$ / * $p < 0,1$.

Italien

Tabelle 16: Bestimmtheitsmaße und p-Werte der drei Vorhersageansätze bezogen auf die Wachstumsrate in Italien von 1991-2009

PH Mon.	Bestimmtheitsmaß			p-Werte		
	Prognosen	OECD CLI	ifo WES	Prognosen	OECD CLI	ifo WES
24	0,10	0,07	0,08	0,181	0,268	0,241
23	0,11	0,11	-	0,159	0,162	-
22	0,12	0,17	-	0,147	0,076	-
21	0,16	0,25	0,28	0,092	0,029	0,021
20	0,16	0,35	-	0,085	0,007	-
19	0,18	0,45	-	0,072	0,002	-
18	0,20	0,53	0,38	0,055	0,000	0,005
17	0,26	0,58	-	0,026	0,000	-
16	0,29	0,61	-	0,017	0,000	-
15	0,49	0,63	0,32	0,001	0,000	0,011
14	0,58	0,65	-	0,000	0,000	-
13	0,69	0,65	-	0,000	0,000	-
12	0,76	0,65	0,35	0,000	0,000	0,007
11	0,80	0,64	-	0,000	0,000	-
10	0,84	0,59	-	0,000	0,000	-
9	0,87	0,50	0,12	0,000	0,001	0,148
8	0,87	0,37	-	0,000	0,005	-
7	0,90	0,22	-	0,000	0,043	-
6	0,92	0,10	0,00	0,000	0,191	0,806
5	0,93	0,03	-	0,000	0,477	-
4	0,94	0,00	-	0,000	0,826	-
3	0,96	0,00	0,08	0,000	0,880	0,253
2	0,98	0,01	-	0,000	0,676	-
1	0,99	0,02	-	0,000	0,539	-

PH = Prognosehorizont; Abhängige Variable: Wachstum des Bruttoinlandsprodukt von 1991-2009 in Prozent; Unabhängige Variable: jeweils einzeln Konsensprognosen / OECD CLI / ifo WES; Anzahl der Beobachtungen: jeweils 19 (Zeitraum 1991-2009); p-Werte: Wahrscheinlichkeit, dass irrtümlich ein Zukunftsbezug des Vorhersageansatzes unterstellt wird, *** $p < 0,01$ / ** $p < 0,05$ / * $p < 0,1$.

Kapitel 4

Prognoserevisionen bei Konjunkturprognosen

*Empirische Ursachenforschung anhand von Konsensprognosen für
das Bruttoinlandsprodukt der G7-Staaten*

Erschienen in:

*Wolfsburg Working Papers (WWP), Ostfalia Hochschule für angewandte Wissenschaften,
Nr. 14-01.*

Abstract

Abfolgen von Prognosen für das Bruttoinlandsprodukt auf Jahresbasis sind häufigen Revisionen unterworfen. Der Theorie der rationalen Erwartungsbildung zufolge müssten Informationen effizient ausgewertet werden und die Revisionen einem random-walk unterliegen. In der Empirie zeigt sich jedoch eine gewisse Trägheit bzw. Ineffizienz der Informationsverarbeitung. Diese Studie zeigt anhand monatlich abgegebener Konsensprognosen für die Wachstumsrate in den G7-Staaten, dass die Effizienz der Informationsverarbeitung keinen Zusammenhang zur Prognosequalität aufweist. Der Revisionsbedarf ergibt sich im Wesentlichen aus zu hohen Prognosen, die zudem eine erheblich geringere Standardabweichung als die realisierten Wachstumsraten aufweisen.

1 Einführung

Wirtschaftsprognosen sind Bestandteil unternehmerischer Planungsprozesse und spielen im Finanzsektor (Kapitalanlage) sowie in der Politik (Steueraufkommen, wirtschaftspolitische Maßnahmen) eine Rolle. Im Fokus der Öffentlichkeit steht dabei besonders die Konjunkturprognose auf Jahresbasis. Dabei wird für ein bestimmtes Jahr eine Serie von Prognosen abgegeben, und zwar sowohl vor Beginn als auch während des entsprechenden Jahres. Man spricht von sogenannten fixed-event-Prognosen, da der Zeitpunkt der Prognosegültigkeit immer gleich bleibt.

Besondere Aufmerksamkeit der Forschung richtet sich auf den Verlauf der Prognoserevisionen bis zum Gültigkeitszeitpunkt. Ziel ist es, Einblick in die Art und Weise der Erwartungsbildung der Prognostiker zu bekommen. Nach der maßgeblich von Muth (1961) und Lucas (1972) entwickelten Theorie der rationalen Erwartungsbildung wären nur diejenigen Prognosen effizient, in die sämtliche verfügbaren Informationen vollständig eingearbeitet worden sind. Als Konsequenz müssten die Prognosefehler rein zufällig sein und dürften keinesfalls mit verfügbaren Informationen korrelieren.

Aus der Problematik heraus, dass sich dies wegen der Menge der denkbaren relevanten Informationen und der mangelnden Transparenz der Prognoseprozesse kaum hinreichend testen lässt, hat Nordhaus (1987) eine Methode zur Überprüfung von Prognosen auf das Vorliegen sogenannter schwacher Informationseffizienz entwickelt. Das Konzept der schwachen Informationseffizienz ist von Fama (1970) im Zuge der „Efficient Market Hypothesis“ entwickelt worden und besagt, dass sich alle öffentlich verfügbaren vergangenen Kursinformationen in den Kursen an den Kapitalmärkten widerspiegeln. Nordhaus kommt zu dem Schluss, dass wenn man schon nicht die effiziente Verarbeitung aller Informationen für die Generierung von Prognosen testen kann, so aber doch zumindest ob die Prognosen der jüngsten Vergangenheit effizient verarbeitet worden sind. Demnach dürften sich keine Zusammenhänge zwischen den Prognoserevisionen ergeben. Der Verlauf der Prognoserevisionen müsste also unsystematisch sein und der Form eines random-walks gleichen.

Dies ist empirisch in der Regel jedoch nicht der Fall, wie schon Nordhaus (1987) sowie bspw. Dovern u. a. (2013) in einer aktuellen Studie zu Konjunkturprognosen festgestellt haben. Statt eines random-walks ist eine „Glättung“ der Prognosen zu beobachten, das sogenannte forecast smoothing. Dabei erfolgen überproportional häufig kleine schrittweise Anpassungen der Prognosen in die jeweils gleiche Richtung.

Hinsichtlich der mangelnden Effizienz der Informationsverarbeitung von Wirtschaftssubjekten werden verschiedene Ursachen diskutiert: Mankiw u. a. (2003) vermuten die Kosten der Informationsbeschaffung und damit verbundene seltenere Beschaffungen als Ursache (sticky information theory). In den Momenten in denen Informationen beschafft werden, verfügen die Prognostiker allerdings über vollständige Information gemäß der Theorie der rationalen Erwartungsbildung. Bei einer weiteren Theorie nach Woodford (2001) und Sims (2003) wird davon ausgegangen, dass die Prognostiker Informationen zwar effizient verarbeiten, die Informationen aber von zweifelhafter Qualität sind (imperfect information theory).

Tversky u. a. (1981) bringen verhaltensökonomische Ursachen in die Diskussion ein und sehen hier besonders die Geschwindigkeit der Informationsverarbeitung als Problem an: „we tend to break the good or bad news to ourselves slowly, taking too long to allow surprises to be incorporated into our forecasts“. Nicht zuletzt könnte die Nutzenfunktion der Prognostiker weitere Aspekte beinhalten, die der bestmöglichen Prognose entgegenstehen. So weist Nordhaus (1987) darauf hin, dass starke Prognosekorrekturen Irritationen bei den Nutzern hervorrufen könnten und daher erklärungsbedürftig sind.

In dieser Studie werden die Ergebnisse einer empirischen Analyse der Qualität und Revisionen von Konsensprognosen für die Wachstumsraten in den G7-Staaten von 1991-2012 bei einem maximalen Prognosehorizont von 24 Monaten vorgestellt. Dabei werden die effiziente Informationsverarbeitung hinsichtlich der Prognoserevisionen und die Prognosequalität einer genauen Betrachtung unterzogen. Die zentrale Hypothese dieser Studie lautet:

H1: Je effizienter die Prognoserevisionen im Sinne von Nordhaus erfolgen, desto geringer fallen die Prognosefehler aus.

Zudem erfolgt eine Suche nach möglichen Ursachen für den fortwährenden Revisionsbedarf. Hierzu werden aufbauend auf einer grafischen Analyse der 24-Monats-Prognosen zwei weitere Hypothesen entwickelt und untersucht. Zum einen fällt auf, dass die 24-Monats-Prognosen des BIP jedes Jahr sehr ähnlich ausfallen. Als zweites richtet sich der Blick auf die Höhe der Prognosen – da die Prognostiker offensichtlich im Mittel deutlich zu optimistische Annahmen hinsichtlich des Wachstums in den G7-Staaten getroffen haben.

Die Studie ist folgendermaßen aufgebaut: Im zweiten Abschnitt wird die Datenbasis für die Analyse erläutert. In Abschnitt 3 erfolgen eine Betrachtung der Qualität der Prognosen und der Höhe und Effizienz der Prognoserevisionen. Die zentrale Untersuchungshypothese wird am Ende des dritten Abschnitts beantwortet. Der vierte Abschnitt dient der Suche nach den Ursachen für den Revisionsbedarf. Mit einer Diskussion der Ergebnisse und einem Fazit im fünften Abschnitt schließt diese Studie.

2 Datenbasis

Als Datenbasis für die Untersuchung dienen Konsensprognosen für das prozentuale Wachstum des Bruttoinlandsprodukts in den G7-Staaten, die durch die britische Fachzeitschrift *Consensus Forecasts* erhoben worden sind. Jeden Monat werden zahlreiche Banken und Wirtschaftsforschungsinstitute seitens des herausgebenden Unternehmens *Consensus Economics* hinsichtlich ihrer Einschätzung der zukünftigen Entwicklung volks- und finanzwirtschaftlicher Kennzahlen befragt. Für das Wachstum des Bruttoinlandsprodukts wird die Prognose der Entwicklung des laufenden sowie des kommenden Jahres erhoben. So wird bspw. in der März-Ausgabe 2011 je eine Prognose für das Gesamtjahr 2011 und eine Prognose für das Jahr 2012 publiziert. Der Prognosehorizont der Prognose für das Jahr 2011 beträgt dabei 10 Monate (bis zum Ende des laufenden Jahres) und für das Jahr 2012 22 Monate (bis zum Ende des Folgejahres). Aus dieser Form der Datenerhebung ergeben sich für jedes Jahr 24 Prognosen bei einem sich stetig verkürzenden Prognosehorizont, so dass eine Betrachtung der monatlichen Prognoserevisionen ohne weiteres möglich wird.

Die arithmetischen Mittelwerte der je nach Land etwa 15-35 Einzelprognosen pro Monat werden von *Consensus Forecasts* als sogenannte Konsensprognose veröffentlicht. Die Zeitschrift ist Ende 1989 das erste Mal erschienen. Das erste Jahr, für welches Prognosen mit allen 24 Prognosehorizonten vorliegen, ist demnach 1991. Die betrachtete Datenbasis erstreckt sich von 1991 – 2012, womit bei je 24 Prognosen und 7 Staaten insgesamt 3696 Konsensprognosen in die Untersuchung einbezogen werden. Diese liegen in einer dreidimensionalen Panelstruktur vor (Prognosehorizont, Jahr, Land).

Prognosegegenstand ist die Wachstumsrate des BIP in den jeweiligen G7-Staaten. Dabei wird ebenfalls die Zeitschrift *Consensus Forecasts* als Datenquelle herangezogen. Für die Analyse findet diejenige Wachstumsrate Verwendung, die in der Mai-Ausgabe für das jeweils vorher-

gehende Jahr veröffentlicht wurde. Diese Vorgehensweise ist auf folgende Abwägung zurückzuführen: Bedient man sich heutiger Zeitreihen, die von den jeweiligen statistischen Ämtern für die vergangenen Jahrzehnte veröffentlicht werden, kann es durchaus zu Änderungen in der Berechnungsmethodik gekommen sein. Solche Strukturbrüche in den Daten waren für die Prognostiker in aller Regel nicht vorhersehbar, so dass man sie auch nicht daran messen sollte. Greift man im Gegenteil auf Daten zurück, die direkt nach Vollendung des prognostizierten Jahres im Januar des Folgejahres veröffentlicht wurden, so basiert die Zahl zu großen Teilen noch auf Schätzungen. Die Diskussion über die Verwendung der „richtigen“ Vergleichsdaten wird bei empirischen Studien zur Evaluation von Wirtschaftsprognosen regelmäßig geführt, bspw. ausführlich bei Croushore (2006). Die Wahl der Mai-Ausgabe erfolgt analog zu einer Analyse der gleichen Datenbasis von Dovern und Weisser (2008).

Da die Datenbasis seit der Erstausgabe von Consensus Forecasts im Oktober 1989 kontinuierlich wächst und inzwischen eine beachtliche Größe erreicht hat, nimmt die Frequenz der Publikationen zu. Blix (2001), Loungani (2001), Spiwoks (2004), Isiklar u. a. (2006), Heilemann und Klinger (2006), Isiklar und Lahiri (2007), Batchelor (2007), Lahiri und Sheng (2008), Osterloh (2008), Spiwoks u. a. (2008a), Spiwoks u. a. (2008b), Ager u. a. (2009), Spiwoks u. a. (2010), Dovern und Weisser (2011), Gubaydullina u. a. (2011), Dovern u. a. (2012) sowie nochmals Dovern u. a. (2013) haben empirische Studien auf der Basis der Daten von Consensus Economics veröffentlicht. Dabei standen neben der Prognosequalität vor allem die Einigkeit der Prognostiker untereinander, die starke Orientierung der Prognostiker an der Gegenwart zum Zeitpunkt der Prognoseerstellung, die mögliche Verzerrtheit und Effizienz der Prognosen sowie der Umgang der Prognostiker mit neuen Informationen in Form von Prognoserevisionen im Vordergrund.

3 Prognosequalität und Revisionen

Für die Analyse der BIP-Prognosen erfolgt zunächst mit Hilfe des mittleren absoluten Prognosefehlers (*MAP*) die Betrachtung der durchschnittlichen Abweichung der Prognosen *P* von der tatsächlichen Wachstumsrate *BIP*. Hierzu wird zum jeweiligen Zeitpunkt *t* die absolute Abweichung berechnet, aggregiert und anschließend mit der Gesamtzahl der Beobachtungen *T* gewichtet.

$$(1) \quad MAP = \frac{1}{T} \sum_{t=1}^T |P_t - BIP_t|$$

Anschließend erfolgt eine Darstellung der Änderung der Prognosequalität zum jeweils vorhergehenden Prognosehorizont *h*.

$$(2) \quad \Delta MAP = MAP_h - MAP_{h-1}$$

Die Ergebnisse der Berechnungen sind Abbildung 1 und Abbildung 2 zu entnehmen.

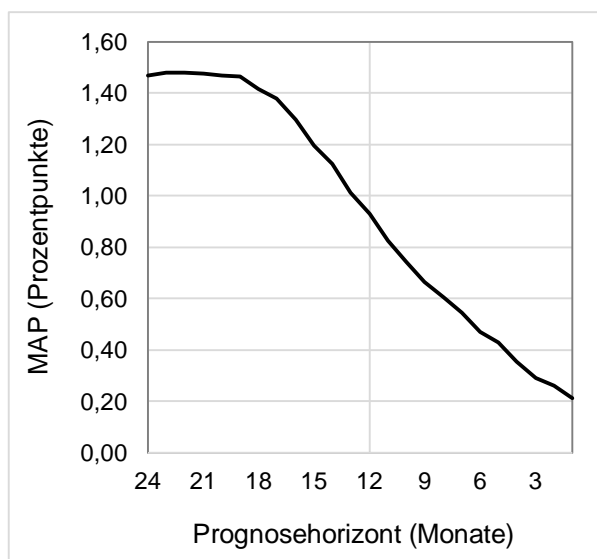


Abbildung 1: Mittlerer absoluter Prognosefehler

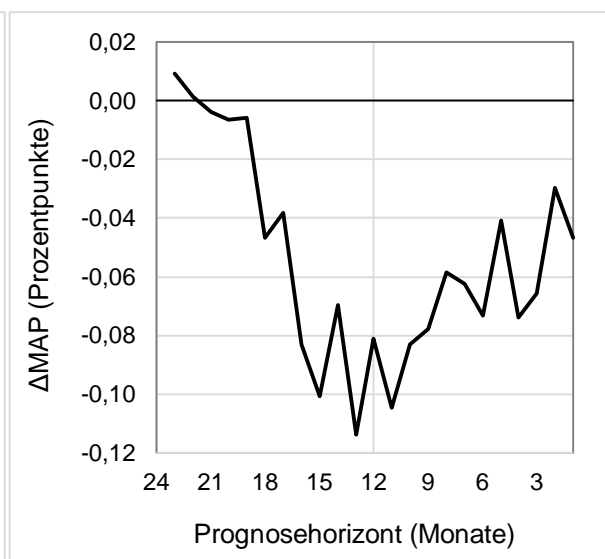


Abbildung 2: Unterschied des Mittleren absoluten Prognosefehlers zum jeweiligen vorherigen Prognosehorizont

Zunächst bestätigt sich der Stand der Forschung insofern, als dass sich die Prognosen mit abnehmenden Prognosehorizont deutlich verbessern. Während das Wirtschaftswachstum im Januar für das jeweilige Folgejahr noch mit einem durchschnittlichen Fehler von über 1,4%-Punkten vorhergesagt wurde, lag dieser im Januar für das jeweils laufende Jahr nur noch bei 0,93%-Punkten.

Die Betrachtung des Qualitätsunterschieds zum jeweils vorhergehenden Prognosehorizont zeigt auf, in welcher Geschwindigkeit sich die Qualität der Konsensprognosen mit der zunehmenden Nähe zum Prognosegegenstand verbessert (Abbildung 2). Im ersten halben Jahr (Prognosehorizont 24 – 19 Monate) verbessern sich die Prognosen nicht und verharren auf einem hohen Fehlerniveau von 1,46 – 1,47 Prozentpunkten. Danach kommt es zu merklichen Verbesserungen. Die deutlichsten Fortschritte in der Prognosequalität werden bei den Prognosen im Oktober (Prognosehorizont 15 Monate; -10 Basispunkte) und Dezember (Prognosehorizont 13 Monate; -11 Basispunkte) für das Folgejahr sowie bei den Prognosen im Februar (Prognosehorizont 11 Monate; -10 Basispunkte) für das laufende Jahr gemessen. Es lässt sich zunächst festhalten, dass sich die Prognosefehler ausgehend von einem hohen Niveau ab einem Prognosehorizont von 18 Monaten verringern.

Um die Qualität besser bewerten zu können, erfolgt nun die Hinzuziehung Maßstabs. Hierzu wird der *MAP* mit der durchschnittlichen Änderung des BIP im Untersuchungszeitraum verglichen. Dies führt zum mittleren absoluten skalierten Prognosefehler (*MAPS*)¹.

$$(3) \quad MAPS = \frac{\frac{1}{T} \sum_{t=1}^T |P_t - BIP_t|}{\frac{1}{T} \sum_{t=1}^T |BIP_{t-1} - BIP_t|}$$

Bei einem Ergebnis von 1 entspricht der *MAP* der durchschnittlichen Änderung der Wachstumsrate des BIP im Untersuchungszeitraum. Der Prognoseaufwand erscheint erst gerechtfertigt, wenn der Wert deutlich unter 1 liegt. Tabelle 1 sind die Ergebnisse der Berechnungen des *MAPS* für die Prognosen des BIP der G7-Staaten bei ausgewählten Prognosehorizonten zu entnehmen.

Tabelle 1: Mittlerer absoluter skaliertes Prognosefehler der Prognosen für das BIP der G7-Staaten bei ausgewählten Prognosehorizonten

PH*	USA	UK	Deutschland	Japan	Frankreich	Kanada	Italien
24	0,87	0,67	1,09	0,87	0,84	0,84	1,06
21	0,86	0,67	1,05	0,89	0,89	0,86	1,06
18	0,78	0,62	1,05	0,85	0,93	0,81	1,02
15	0,63	0,53	0,87	0,72	0,75	0,77	0,81
12	0,51	0,42	0,60	0,63	0,53	0,62	0,60
1	0,11	0,11	0,13	0,15	0,16	0,12	0,12

*PH=Prognosehorizont in Monaten

¹ Hyndman und Koehler (2006)

Lediglich die BIP-Prognosen für Großbritannien weisen bei den größeren Prognosehorizonten einen *MAPS* von deutlich unter 1 auf (0,67). Selbst bei den Prognosen, die im Oktober für das nächste Jahr abgegeben werden (Prognosehorizont 15 Monate), ist der *MAPS* bei fast allen G7-Staaten noch nennenswert hoch. Der Prognosefehler beträgt zu diesem Zeitpunkt im Mittel etwa dreiviertel der durchschnittlichen jährlichen Schwankung der realisierten Wachstumsraten.

Im Mittelpunkt der Studie stehen die Revisionen der Prognosen, d. h. die jeweiligen Korrekturen gegenüber den vorhergehenden Prognosen. Dabei werden zunächst die Mittelwerte der Revisionen (*MRev*) und die Mittelwerte der absoluten Revisionen ermittelt (*MAREv*). Die beiden Maße unterscheiden sich in der unterschiedlichen Handhabung der Saldierung positiver und negativer Revisionen. Sei P die Prognose für das Jahr y im Land i , h der Prognosehorizont und n die Gesamtzahl Prognosen, dann berechnet sich die *MRev* wie folgt:

$$(4) \quad MRev = \frac{1}{n} (P_{y,i,h} - P_{y,i,h-1})$$

Um die gesamte Höhe der Revisionen zu ermitteln, zieht der *MAREv* die absoluten Revisionen zur Ermittlung des arithmetischen Mittels heran:

$$(5) \quad MAREv = \frac{1}{n} |P_{y,i,h} - P_{y,i,h-1}|$$

Für den Prognosehorizont von 24 Monaten gibt es keine vorhergehende Prognose, sodass erst ab einem Prognosehorizont von 23 Monaten Revisionen untersucht werden können. Bei einem Untersuchungszeitraum von 22 Jahren und einem Umfang von 7 Staaten ergeben sich für jeden Prognosehorizont 154 Beobachtungen. Die Darstellung der Ergebnisse der Berechnungen als G7-Mittelwerte erfolgt in Abbildung 3 und Abbildung 4.

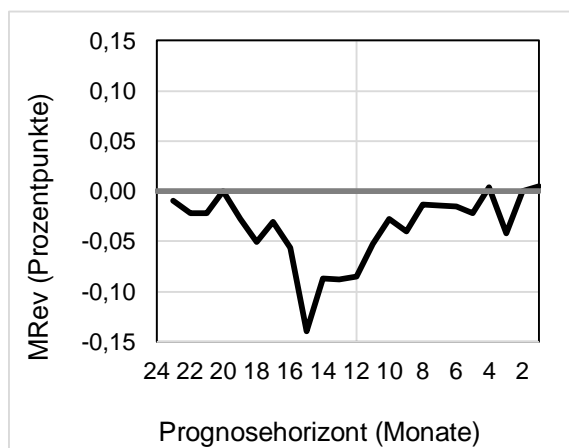


Abbildung 3: Durchschnittliche Prognoserevisionen

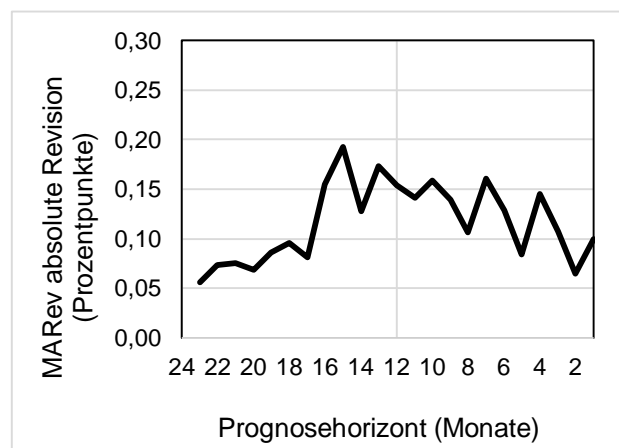


Abbildung 4: Absolute durchschnittliche Prognoserevisionen

Die stärkste Revisionsaktivität findet in den Monaten vor dem Jahreswechsel und dort vor allem im Oktober (Prognosehorizont 15 Monate) statt. Die durchschnittlichen Revisionen sind dabei negativ. Setzt man diese in Relation zu den absoluten Revisionen wird deutlich, dass das Ausmaß der Revisionen zu etwa 75% nach unten erfolgt. Scheinbar werden besonders Wachstumsabschwächungen spät erkannt und machen umfangreiche Revisionen erforderlich. Dieses Phänomen verschwindet nach und nach bei den Prognosen für das laufende Jahr. Zwar werden im letzten Halbjahr immer noch monatlich Revisionen in Höhe von 0,1 Prozentpunkten des BIP-Wachstums vorgenommen, diese verteilen sich jedoch fast gleichmäßig auf positive und negative Revisionen.

Setzt man diese Erkenntnisse in Relation zur Verbesserung der Prognosequalität, so ergibt sich ein konsistentes Bild. Bei den Prognosen im ersten halben Jahr für die Wachstumsraten im Folgejahr erfolgen kaum Revisionen – auch eine Verbesserung der Prognosequalität ist kaum feststellbar. Danach verbessern sich die Prognosen ab Oktober (Prognosehorizont 15 Monate) deutlich – einhergehend mit einer hohen Revisionsaktivität. Diese geht bei den Prognosen für das laufende Jahr zurück, womit auch die Qualitätsverbesserung abnimmt. Das Potenzial für weitere (absolute) Verbesserungen ist zu diesem Zeitpunkt jedoch schon eingeschränkt, da die Prognosefehler bereits vergleichsweise niedrig sind.

Die Betrachtung der durchschnittlichen und absoluten Prognoserevisionen gibt jedoch noch keinerlei Aufschluss darüber, ob diese autokorreliert sind und somit ein Verstoß im Sinne der effizienten Informationsverarbeitung nach Nordhaus (1987) vorliegt. Um dies zu erkennen, muss der Zusammenhang zwischen den Revisionen betrachtet werden. Hierzu wird mittels des

von Nordhaus (1987) vorgeschlagenen Tests überprüft, ob schwache Informationseffizienz bei den Konsensprognosen für die G7-Staaten vorliegt. Hierzu wird eine Regressionsanalyse mittels der Methode der kleinsten Quadrate durchgeführt. Abhängige Variable ist jeweils die Prognoserevision der Prognose des BIP für das Jahr y im jeweiligen Land i , welche sich aus der Differenz zur einen Monat vorher für den gleichen Prognosegegenstand abgegebenen Prognose P ergibt (zum Prognosehorizont $h-1$). Unabhängige Variable ist die vorherige Prognoserevision zwischen den beiden Prognosen der Prognosehorizonte $h-2$ und $h-1$.

$$(6) \quad P_{y,i,h} - P_{y,i,h-1} = \alpha + \beta(P_{y,i,h-1} - P_{y,i,h-2}) + \varepsilon$$

Für das Vorliegen schwacher Effizienz dürfte die vorherige Revision keinen Zusammenhang zur folgenden Prognoserevision aufweisen. Hierfür genügt der Nachweis von $\beta = 0$. Die Berechnung der Standardfehler erfolgt konservativ nach Driscoll und Kraay (1998), da komplexe Korrelationsmuster zwischen den Konsensprognosen für die einzelnen Länder bestehen dürften.

Der längste Prognosehorizont, für den die hier vorgestellte Schätzung des Revisionskoeffizienten durchgeführt werden kann, beträgt 22 Monate und greift auf die Prognosen von Januar - März für das kommende Jahr zurück. Die Revision der Februar-Prognose für das kommende Jahr ist die abhängige Variable und die vorhergehende Revision der Januar-Prognose für das kommende Jahr die unabhängige Variable. Die betrachteten 22 Jahre in Verbindung mit der Eingrenzung auf die G7-Staaten führen zu 154 Beobachtungen mit dem Ergebnis $\beta=0,12$ (p -Wert=0,34; nicht signifikant). Die Schätzung wird für alle 21 niedrigeren Prognosehorizonte wiederholt. Die Höhe des Revisionskoeffizienten β und mit den oberen und unteren Grenzen der 5%-igen Irrtumswahrscheinlichkeit ist Abbildung 5 zu entnehmen.

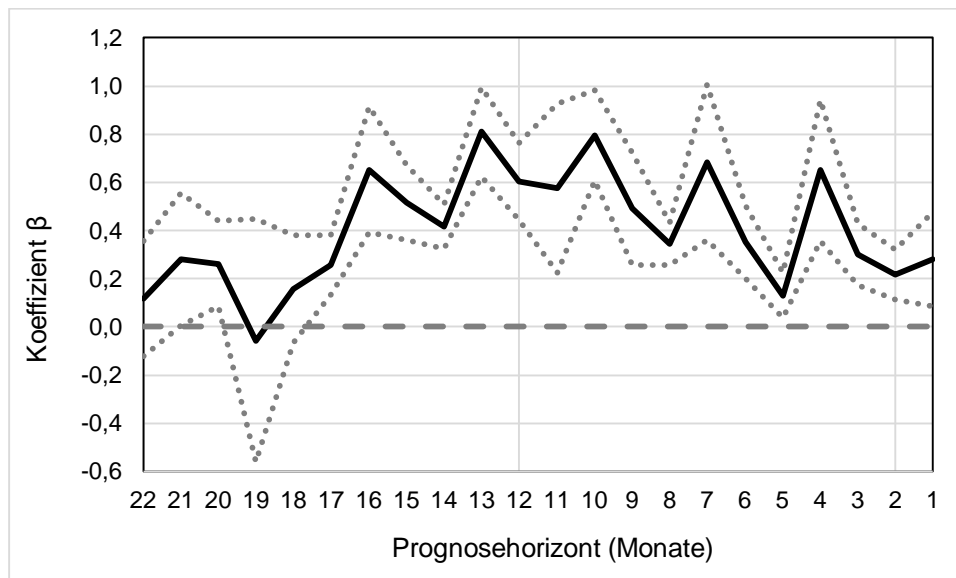


Abbildung 5: Ergebnisse der Schätzung des Revisionskoeffizienten

Für das erste halbe Jahr ist kaum ein signifikanter Zusammenhang zwischen den vorherigen und den folgenden Revisionen der Konjunkturprognosen feststellbar. Bei denjenigen Prognosen, die ab dem September für das Folgejahr abgegeben wurden (Prognosehorizont 16 Monate), steigt der Koeffizient auf über 0,4 an und erreicht bei den Prognosen im Dezember für das kommende Jahr sein Höhepunkt ($\beta=0,8$). Bei den anschließenden Prognosen für das laufende Jahr sinkt der Koeffizient im Mittel wieder – bleibt jedoch zu jedem Zeitpunkt signifikant über 0. Effiziente Informationsverarbeitung nach Nordhaus ($\beta=0$) könnte allenfalls für die Prognosen des ersten halben Jahrs für das Folgejahr gelten (Prognosehorizonte 22-18 Monate).

Das Ergebnis fügt sich mit einem arithmetischen Mittelwert von 0,4 nahtlos in die Ergebnisse anderer Studien ein, wenngleich die Koeffizienten dort nicht mit monatlichen sondern mit Revisionen im Abstand von jeweils einem Quartal geschätzt worden sind. Dovert u. a. (2013) kommen auf horizontübergreifende Werte zwischen 0,37 für Industrie- und 0,4 für Schwellenländer. Sie stellen bei den Prognosehorizonten von 10 und 13 Monaten die höchsten Werte fest (maximal 0,57). Coibion und Gorodnichenko (2010) kommen bei verschiedenen weiteren Datenquellen auf Werte zwischen 0,38 und 0,7.

Die zentrale Untersuchungshypothese lautet:

H1: Je effizienter die Prognoserevisionen im Sinne von Nordhaus erfolgen, desto geringer fallen die Prognosefehler aus.

Die Hypothese muss klar verworfen werden. Bei den hohen Prognosehorizonten von 18 Monaten oder mehr erfolgt die Informationsverarbeitung effizient, aber mit Blick auf die Höhe der Prognosefehler besonders erfolglos. So sind die mittleren absoluten Prognosefehler (Abbildung 1) sowie die mittleren absoluten skalierten Prognosefehler (Tabelle 1) zu diesem Zeitpunkt besonders hoch. Nun ließe sich noch einwenden, dass die Fehler zwar hoch sind, die effiziente Informationsverarbeitung jedoch zu einem schnellen Rückgang führt. Mit Blick auf die Verringerung der Fehler in Abbildung 2 wird jedoch deutlich, dass bei den hohen Prognosehorizonten die geringsten monatlichen Fortschritte zu erkennen sind. Zentrales Ergebnis ist somit, dass es keinen positiven Zusammenhang zwischen der Verbesserung der Prognosequalität auf der einen Seite und einer effizienten Informationsverarbeitung nach Nordhaus auf der anderen Seite gibt. Dies lässt sich möglicherweise mit der vollkommenen Unkenntnis der Prognostiker erklären: „Planlose“ Änderungen der Prognosen in die eine oder andere Richtung sind eben „effizient“, da sich diese in Revisionen in Form eines random-walks manifestieren.

4 Ursachen für den Revisionsbedarf

Der vorige Abschnitt hat gezeigt, dass die bei der Prognose für das BIP-Wachstum des Folgejahres im Februar beginnende Kette von Prognoserevisionen ab einem Prognosehorizont von 17 Monaten keinem random-walk mehr unterliegt. Wie sich ebenfalls herausgestellt hat, erfolgen die Revisionen nicht gleichmäßig, sondern weisen im Mittel einen Schwerpunkt in Form einer starken Korrektur nach unten bei den im Herbst für das Folgejahr abgegebenen Prognosen auf. Betrachtet man die durchschnittliche 24-Monats-Konsensprognose für das Wachstum der G7-Staaten im Vergleich zu den realisierten Wachstumsraten (Abbildung 6), lassen sich Ursachen für den Revisionsbedarf im Zeitverlauf vermuten.

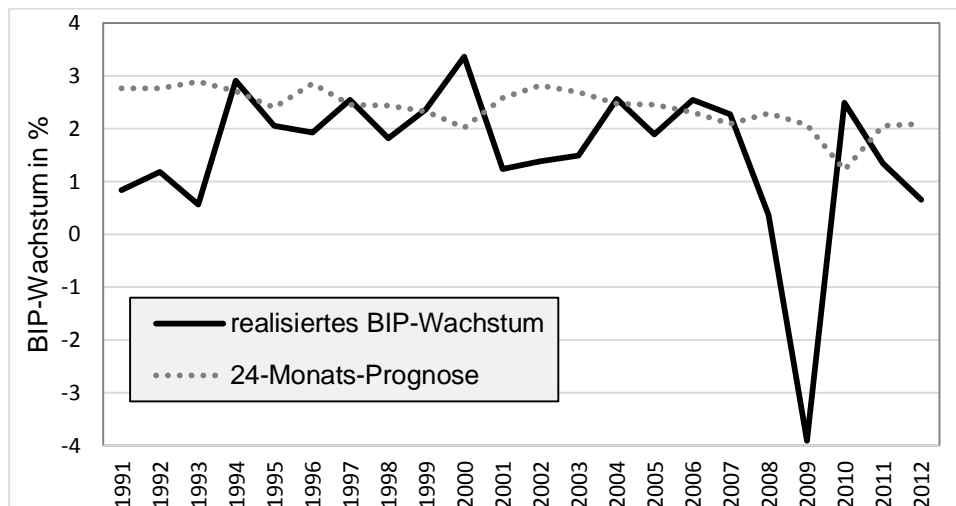


Abbildung 6: Durchschnittliches BIP-Wachstum in den G7-Staaten und entsprechende 24-Monats-Konsensprognose von 1991-2012

Während das durchschnittliche Wachstum in den G7-Staaten starken Schwankungen unterworfen ist, scheinen die 24-Monats-Prognosen weitgehend konstant. Aufbauend auf diesem optischen Eindruck soll folgende Hypothese untersucht werden:

H2: Ein Teil des Revisionsbedarfs resultiert aus der niedrigen Standardabweichung der 24-Monats-Prognosen.

Weiterhin scheinen die Prognosen im Mittel zu hoch zu sein. Neben dem Ausnahmejahr 2009 ist dies auch Anfang der 90er Jahre, 2001-2003 und zuletzt 2012 deutlich der Fall. Jahre, in denen die 24-Monats-Prognosen zu niedrig waren, erscheinen hingegen deutlich weniger aufzutreten. Die Untersuchung der zweiten möglichen Ursache für den Revisionsbedarf erfolgt auf der Basis folgender Hypothese:

H3: Ein Teil des Revisionsbedarfs lässt sich durch die im Mittel deutlich zu hoch abgegebenen Prognosen erklären.

Zunächst richtet sich der Blick auf *H2*. Dabei wird als erstes für jeden G7-Staat und jeden Prognosehorizont die Standardabweichung σ berechnet und anschließend durch die Standardabweichung des eingetretenen BIP-Wachstums dividiert.

$$(7) \quad \sigma - \text{Koeffizient} = \frac{\sigma_{\text{Prognosezeitreihe}}}{\sigma_{\text{BIP-Wachstumsrate}}}$$

Diese Vorgehensweise führt zu einem einfach zu interpretierendem Vergleich der Unterschiede. Werte < 1 zeigen an, dass die Prognosen eine geringere Standardabweichung als die eingetretenen Wachstumsraten aufweisen. Bei Werten > 1 ist die Standardabweichung größer.

Ein Wert von 1 zeigt an, dass die Standardabweichungen der Prognosen und der Wachstumsraten gleich hoch sind. Die folgende Abbildung zeigt den Verlauf der berechneten σ -Koeffizienten für die G7-Staaten nach Prognosehorizonten.

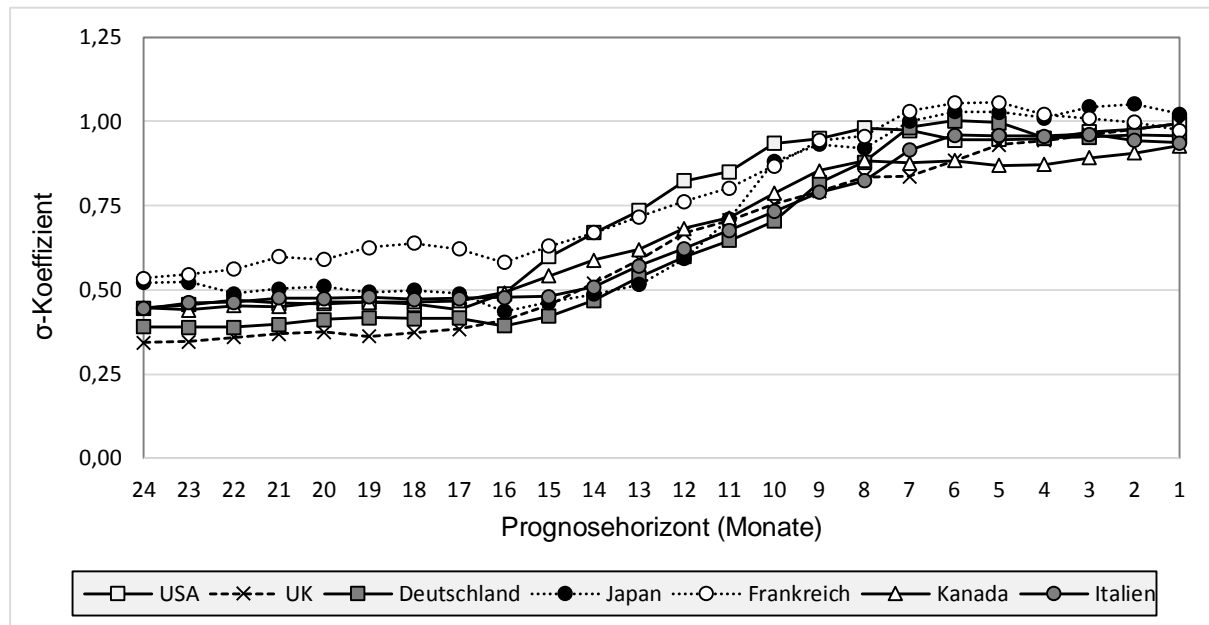


Abbildung 7: Gegenüberstellung der Standardabweichung von Prognosen und Wachstumsrate in den G7-Staaten mittels eines Koeffizienten

Es zeigt sich, dass die Standardabweichungen der Prognosen zunächst nur etwa halb so hoch wie die der tatsächlich eingetretenen Wachstumsraten liegen. Erst ab einem Prognosehorizont von 16 Monaten steigen die Koeffizienten deutlich an. Bei den sehr niedrigen Prognosehorizonten wird in etwa die Standardabweichung der Wachstumsraten in den G7-Staaten erreicht (Werte um 1), da die realisierten Werte im laufenden Jahr immer mehr Bestandteil der Prognosen werden. Wenngleich die Niveaus zwischen den einzelnen Staaten leichte Unterschiede offenbaren, so zeigt sich das Phänomen der zunächst sehr niedrigen Standardabweichungen für alle Staaten gleichermaßen.

Mittels eines F-Tests² wird die Hypothese überprüft, dass die Standardabweichung der Prognosen und des realisierten Wachstums gleich hoch sind. Die Irrtumswahrscheinlichkeit, dass diese Hypothese zu Unrecht verworfen wird, ist Tabelle 2 zu entnehmen.

² F-Test zum Vergleich zweier Varianzen nach Brown und Forsythe (1974). Das verwendete Verfahren ist robust hinsichtlich der Annahme, dass die zugrundeliegenden Zeitreihen normalverteilt sind. Dies lässt sich wegen der geringen Zahl der Beobachtungen pro Jahr und Prognosehorizont ($n=22$) nicht mit Sicherheit feststellen. Das Standardverfahren führt zu insgesamt sehr ähnlichen Resultaten.

Tabelle 2: Irrtumswahrscheinlichkeit, dass die Hypothese gleicher Standardabweichungen von Prognosen und Wachstumsraten in den G7-Staaten zu Unrecht abgelehnt wird.

PH	USA	Deutschland	UK	Japan	Frankreich	Kanada	Italien
24	0,435%***	2,749%**	1,200%**	1,852%**	0,915%***	0,630%***	5,379%*
23	0,607%***	2,912%**	1,207%**	1,971%**	1,379%**	0,524%***	6,646%*
22	0,887%***	2,802%**	1,492%**	1,123%**	1,920%**	0,757%***	6,174%*
21	0,695%***	3,329%**	1,717%**	1,503%**	3,003%**	0,666%***	7,447%*
20	0,672%***	3,616%**	1,846%**	1,464%**	2,921%**	0,865%***	8,105%*
19	0,723%***	3,914%**	1,826%**	1,297%**	6,274%*	0,909%***	8,565%*
18	0,491%***	3,943%**	2,578%**	1,405%**	7,770%*	1,009%**	8,137%*
17	0,368%***	3,968%**	3,502%**	1,209%**	5,906%*	1,137%**	9,266%*
16	0,885%***	3,062%**	5,579%*	0,447%***	4,264%*	2,044%**	10,148%
15	4,452%**	5,657%*	9,308%*	0,773%***	7,879%*	5,330%*	9,902%*
14	9,533%*	9,505%*	16,550%	1,228%**	12,064%	9,107%*	11,755%
13	17,306%	15,104%	24,753%	1,993%**	13,160%	12,460%	14,992%
12	31,365%	20,503%	37,429%	6,138%*	20,893%	17,422%	19,544%
11	41,126%	24,931%	44,999%	12,722%	28,032%	17,761%	23,545%
10	61,661%	33,200%	52,748%	30,815%	38,337%	27,145%	31,407%
9	66,584%	49,530%	60,691%	43,848%	49,720%	35,900%	37,930%
8	71,926%	57,970%	66,581%	43,606%	56,070%	41,993%	43,448%
7	72,890%	74,455%	67,953%	61,739%	73,762%	43,338%	60,177%
6	68,414%	80,355%	80,444%	86,735%	82,179%	52,178%	66,227%
5	72,929%	81,788%	89,972%	88,225%	87,901%	48,032%	69,283%
4	72,082%	85,699%	95,382%	84,083%	96,137%	56,559%	73,268%
3	83,958%	88,841%	99,231%	93,682%	97,395%	67,494%	77,354%
2	88,850%	94,361%	96,189%	92,068%	90,923%	71,157%	79,391%
1	95,988%	95,879%	90,900%	93,402%	85,521%	75,232%	77,647%

PH = Prognosehorizont in Monaten; * = Irrtumswahrscheinlichkeit < 10%, ** = Irrtumswahrscheinlichkeit < 5%, *** = Irrtumswahrscheinlichkeit < 1%;

Es wird unabhängig vom jeweiligen G7-Staat deutlich, dass die Prognosen für das kommende Jahr signifikant geringere Standardabweichungen ausweisen, als die tatsächlich eingetretenen Wachstumsraten. Für die Mehrheit der Staaten liegt der Wert erst ab den Oktober-Prognosen für das kommende Jahr (Prognosehorizont 15 Monate) über dem 5%-Niveau der Irrtumswahrscheinlichkeit. Man kann in diesem Bereich also sicher davon ausgehen, dass die Standardabweichungen der Prognosen und der tatsächlichen Wachstumsraten nicht zufallsbedingt, sondern systematisch voneinander abweichen. Lediglich für Italien liegt das Signifikanzniveau bei den höheren Prognosehorizonten durchweg über 5%, allerdings unter 10%. Bei den Prognosen für das laufende Jahr ändert sich dieses Bild grundlegend – die Standardabweichungen haben sich angenähert. Die Hypothese *H2* lässt sich demnach vorläufig bestätigen, da zumindest bei den Prognosen für das kommende Jahr eine sehr niedrige Standardabweichung im Vergleich zu den realisierten Wachstumsraten vorliegt.

Neben der großen Konstanz der Prognosen legt Abbildung 6 nahe, dass die 24-Monats-Prognosen im Durchschnitt zu hoch ausgefallen sein könnten. Dies hat zur Bildung der dritten Hypothese geführt:

H3: Ein Teil des Revisionsbedarfs lässt sich durch die im Mittel deutlich zu hoch abgegebenen Prognosen erklären.

Zu hohe Prognosen werden an dieser Stelle nicht erstmals thematisiert. Heilemann und Klinger (2006) weisen auf das Phänomen hin und sprechen für die G7-Staaten von „positiven Verzerrungen“ der Prognosen. Für Deutschland sei keine der Rezessionen frühzeitig erkannt worden. Generelle Aussagen lassen sich jedoch nicht treffen, wie der umgekehrte Befund von Blix (2001) zeigt. Er kommt zu dem Schluss, dass die Prognosen in den neunziger Jahren zu niedrig ausgefallen sind – vor allem für die USA.

Zur Überprüfung dieser Hypothese in Bezug auf die Datenbasis dieser Studie wird im Folgenden der Mittlere Prognosefehler (*MP*) ermittelt. *P* ist dabei die Prognose und *BIP* die realisierte Wachstumsrate jeweils zum Zeitpunkt *t*. Die Durchschnittswerte werden unter Zuhilfenahme der Gesamtzahl der Prognosen *T* ermittelt. Anders als beim *MAP* heben sich Über- und Unterschätzungen der Prognostiker gegenseitig auf, so dass dauerhafte Über- oder Unterschätzungen offensichtlich werden.

$$(8) \quad MP = \frac{1}{T} \sum_{t=1}^T P_t - BIP_t$$

Die Ergebnisse im G7-Durchschnitt sind Abbildung 8 zu entnehmen. In Abbildung 9 wird dargestellt, wie viel Prozent bezogen auf die Anzahl der Prognosen zu hoch waren.

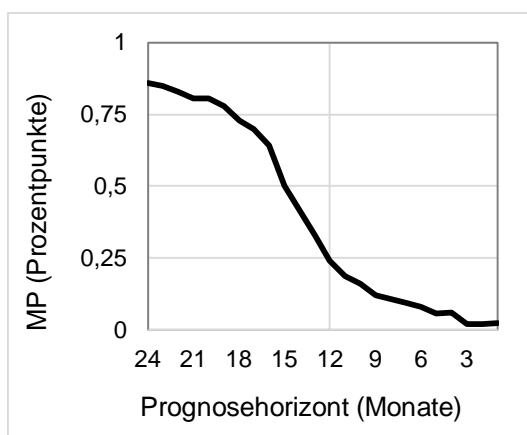


Abbildung 8: Mittlerer Prognosefehler

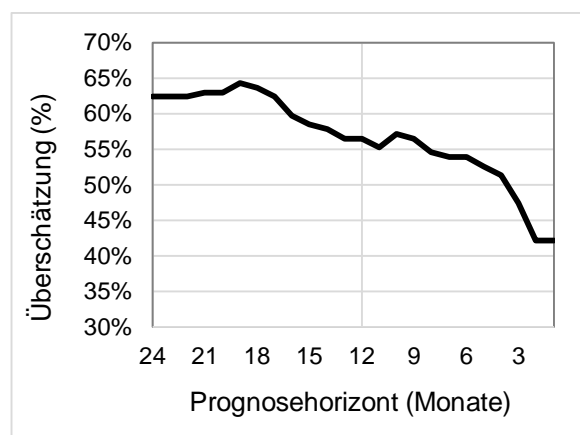


Abbildung 9: Häufigkeit der Überschätzung des BIP-Wachstums in Prozent

Im Mittel ist das Wachstum in den G7-Staaten erheblich überschätzt worden. Diejenigen Prognosen, die im ersten halben Jahr für das Wachstum des Folgejahres abgegeben worden sind, weisen eine Überschätzung von über 0,75%-Punkten auf. Der Prognosehorizont von 24 Monaten geht mit einer Überschätzung von 0,86%-Punkten einher. Zieht man den im 2. Kapitel ermittelten absoluten Prognosefehler von 1,47%-Punkten in die Bewertung mit ein, so lässt sich sagen, dass knapp 80% der Höhe der Prognosefehler Folge von Überschätzungen sind. In Abbildung 9 wird deutlich, dass zunächst über 60% der Prognosen zu hoch ausgefallen sind. Der Erwartungswert in einem Szenario effizienter Prognosen beträgt 50% und wird erst bei den Prognosen für das laufende Jahr erreicht. Somit lässt sich die 3. Hypothese ebenfalls vorläufig bestätigen, d. h. dass ein Teil des Revisionsbedarfs durch die im Mittel deutlich zu hohen Prognosen zu erklären ist.

Die Daten zeigen also, dass

1. Anfang jeden Jahres sehr ähnliche Prognosen für das Folgejahr abgegeben werden
2. und dass diese im Durchschnitt deutlich zu hoch ausfallen.

Das erste Phänomen ist in einer Welt der rationalen Erwartungsbildung nur erklärbar, wenn kaum Informationen und / oder Kenntnisse über die volkswirtschaftlichen Zusammenhänge vorliegen. In einer solchen Situation wäre es rational, den zukünftigen Konjunkturverlauf bspw. anhand der vergangenen durchschnittlichen Wachstumswahlen des betrachteten Wirtschaftsraums zu approximieren. Dies impliziert den niedrigsten Erwartungswert bezüglich der Höhe des Revisionsbedarfs.

Auch das zweite Phänomen ist unter Umständen mit rationalen Erwägungen erklärbar. So ist es durchaus fraglich, ob sich der vorsichtige Prognostiker am arithmetischen Mittel der vergangenen realisierten Wachstumsraten orientieren sollte. Jahren mit eher durchschnittlichen Wachstumsraten stehen deutliche Ausreißer nach unten gegenüber, wie der Vergleich der Mittelwerte und des jeweiligen Median der G7-Staaten im Untersuchungszeitraum veranschaulicht.

Tabelle 3: Arithmetisches Mittel und Median des BIP-Wachstums der G7-Staaten im Vergleich

	USA	UK	Deutschland	Japan	Frankreich	Kanada	Italien
Arithm. Mittel	2,54	1,51	1,38	1,01	1,47	2,20	0,68
Median	2,75	2,10	1,60	1,20	1,60	2,60	1,10
Differenz	-0,21	-0,22	-0,59	-0,19	-0,13	-0,40	-0,42

σ = Standardabweichung des BIP-Wachstums / der Prognosen für den Untersuchungszeitraum 1991-2012.

In allen Staaten liegt der Mittelwert unter dem Median. Die Jahre mit sehr niedrigen Wachstumsraten (G7-Durchschnitt 2009: -3,91; 2008: 0,37; 1993: 0,56) sind erheblich weiter vom arithmetischen Mittel aller Wachstumsraten entfernt als die Jahre mit hohen Wachstumsraten (G7-Durchschnitt 2000: 3,37; 1994: 2,91; 2004: 2,56). Auch ohne Berücksichtigung des „Ausnahmejahres“ 2009 liegt der Median in allen Staaten über dem arithmetischen Mittel. Bei Prognosen unter Unsicherheit kann es durchaus rational sein, sich eher am Median der vergangenen Jahre zu orientieren als am arithmetischen Mittel. Für einzelne Jahre mit sehr schwachen Wachstumsraten fällt der Prognosefehler dann umso größer aus. In den meisten Jahren jedoch ist er entsprechend niedriger.

Es ist also denkbar, dass das in der Literatur diskutierte Phänomen der mangelnden Effizienz von Prognoserevisionen vorwiegend auf den erheblichen Revisionsbedarf im Fall von Rezessionen zurückzuführen ist. In Abbildung 10 und Abbildung 11 erfolgen analog zum 3. Abschnitt eine Darstellung der mittleren (absoluten) Prognoserevisionen. Nun jedoch werden die Ergebnisse für die G7-Staaten in jeweils eine Hälfte mit den höheren und eine Hälfte mit den niedrigeren Wachstumsraten geteilt. Es erfolgt also ein Vergleich der Revisionen für die Jahre mit hohen und für die Jahre mit niedrigen realisierten Wachstumsraten.

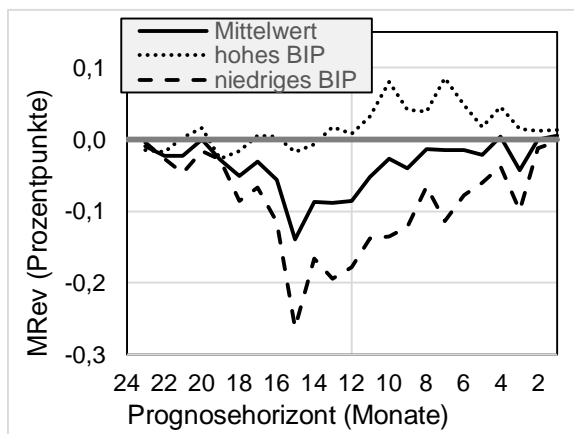


Abbildung 10: Durchschnittliche Prognoserevisionen bei hohen und niedrigen Wachstumsraten

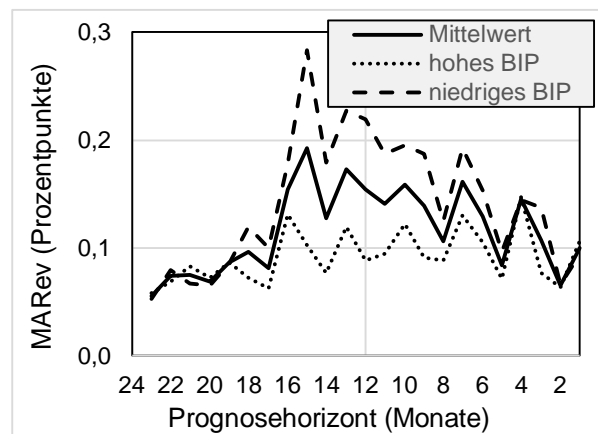


Abbildung 11: Absolute durchschnittliche Prognoserevisionen bei hohen und niedrigen Wachstumsraten

Es offenbaren sich erhebliche Unterschiede zwischen den Jahren mit den höheren und den niedrigeren Wachstumsraten. Zunächst ist in Abbildung 10 zu erkennen, dass sich die Revisionen für die Jahre mit höheren Wachstumsraten bei den Prognosen für das kommende Jahr gegenseitig aufheben und im Mittel nahe 0 liegen. Anders stellt sich dies für die Prognosen für Jahre mit den niedrigeren Wachstumsraten dar. Ab einem Prognosehorizont von 18 Monaten erfolgen

sehr deutliche Korrekturen nach unten. Mit einer durchschnittlichen Abwärtsrevision von 0,26 Prozentpunkten findet dieser Prozess bei den Prognosen im Oktober (Prognosehorizont 15 Monate) seinen Höhepunkt. In Abbildung 11 wird deutlich, dass der Revisionsbedarf für die Jahre mit niedrigen Wachstumsraten zeitweise mehr als doppelt so hoch ist.

5 Fazit

Konjunkturprognosen, die für ein bestimmtes Jahr abgegeben werden, werden bis zum jeweiligen Jahresende regelmäßig revidiert. Eine aus der Theorie der rationalen Erwartungsbildung abgeleitete Forderung an die Prognostiker ist es, dass sämtliche neue Informationen sofort vollumfänglich in die neue Prognose eingearbeitet werden. Demnach müssten die Revisionen nach Nordhaus (1987) einem Random-walk unterliegen. Diese Studie bestätigt zunächst den aktuellen Stand der Forschung dahingehend, dass sich auch bei einer monatsweisen Betrachtung von Revisionen für Konsensprognosen bezüglich des BIP-Wachstums in den G7-Staaten in der Regel kein random-walk einstellt, die Revisionen also nicht effizient sind.

Der Fokus der Untersuchung liegt auf einem differenzierten Blick auf den zeitlichen Verlauf und dem Zusammenhang zwischen der Höhe der Prognosefehler und der Effizienz der Prognoserevision. Es zeigt sich, dass diejenigen Prognosen, die im ersten Halbjahr für das BIP-Wachstum des Folgejahres abgegeben werden, effizient im Sinne der Theorie der rationalen Erwartungsbildung sind. Die Prognosefehler liegen zu diesem Zeitpunkt konstant auf hohem Niveau und nehmen kaum ab. Danach erfolgt bei sich verkürzendem Prognosehorizont Monat für Monat eine Verbesserung der Prognosen. Dies geschieht jedoch nicht effizient im Sinne der Theorie der rationalen Erwartungsbildung, da die Revisionen mit ihrem jeweiligen Vormonatswert korreliert sind. Als zentrales Untersuchungsergebnis lässt sich also festhalten, dass eine effiziente Informationsverarbeitung im Sinne von Nordhaus keinesfalls mit geringeren Prognosefehlern einhergeht.

Die Studie weist darüber hinaus auf zwei mögliche Ursachen für den Revisionsbedarf hin. Die anfänglichen Prognosen weisen eine geringe Standardabweichung auf – es werden also für jedes Jahr zunächst ähnlich hohe Prognosen abgegeben. Der Prognosegegenstand, die jährliche Wachstumsrate des BIP, zeigt diesbezüglich erheblich höhere Abweichungen. Mit rückläufigem Prognosehorizont erhöht sich die Standardabweichung und gleicht sich immer mehr der Standardabweichung der Wachstumsraten an. Die zweite Ursache ist die Höhe der Prognosen,

da diese deutlich über der durchschnittlichen Wachstumsrate in den G7-Staaten liegt. Dies impliziert im Mittel einen Revisionsbedarf der Prognosen nach unten. Eine tiefergehende Analyse ergibt, dass sich bei den Prognosen für das kommende Jahr lediglich für die Jahre mit niedrigen Wachstumsraten ein einseitiger und hoher Revisionsbedarf ergibt.

Am Ende ist es fraglich, ob der Fokus auf die Effizienz von Prognoserevisionen einen sinnvollen Beitrag für die empirische Analyse von Konjunkturprognosen liefern kann. Anders als bei Prognosen in den Naturwissenschaften (bspw. Sonnenfinsternis) ist die konjunkturelle Entwicklung zum Zeitpunkt der Prognoseabgabe (höchstwahrscheinlich) nicht durch Gesetzmäßigkeiten determiniert, die es „nur“ noch durch immer differenziertere Methoden aufzudecken gilt. Wenn sich Rezessionen Schritt für Schritt entwickeln sollten und bei jedem Schritt Unklarheit über den weiteren Verlauf herrscht, ist auch eine schrittweise Anpassung der Prognosen die einzig mögliche und einzig sinnvolle Vorgehensweise.

Literatur

- Ager, P. / Kappler, M. / Osterloh, S. (2009): „The accuracy and efficiency of the Consensus Forecasts: A further application and extension of the pooled approach“, *International Journal of Forecasting*, Bd. 25, Nr. 1, S. 167–181.
- Batchelor, R. (2007): „Bias in macroeconomic forecasts“, *International Journal of Forecasting*, Bd. 23, Nr. 2, S. 189–203.
- Blix, M. (2001): „How good is the forecasting performance of major institutions?“, *Sveriges Riksbank economic review*, Bd. 12, Nr. 3, S. 38–68.
- Brown, M. B. / Forsythe, A. B. (1974): „Robust Tests for the Equality of Variances“, *Journal of the American Statistical Association*, Bd. 69, Nr. 346, S. 364–367.
- Coibion, O. / Gorodnichenko, Y. (2010): „Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts“, *National Bureau of Economic Research Working Paper Series*, Bd. No. 16537.
- Croushore, D. (2006): Forecasting with real-time macroeconomic data, in: Graham, E. / Granger, C. W. J. / Timmermann, A. (Herausgeber), Reihe Handbook of Economic Forecasting, Bd. 1.
- Dovern, J. / Fritsche, U. / Loungani, P. / Tamirisa, N. (2013): „Information Rigidities in Economic Growth Forecasts: Evidence from a Large International Panel“, *IMF Working Paper*, Bd. 13, Nr. 56.
- Dovern, J. / Fritsche, U. / Slacalek, J. (2012): „Disagreement Among Forecasters in G7 Countries“, *Review of Economics & Statistics*, Bd. 94, Nr. 4, S. 1081–1096.
- Dovern, J. / Weisser, J. (2008): „Are they really rational? Assessing professional macro-economic forecasts from the G7-countries.“, *Kiel Working Paper, IFW*, Nr. 1447.
- Dovern, J. / Weisser, J. (2011): „Accuracy, unbiasedness and efficiency of professional macro-economic forecasts: An empirical comparison for the G7“, *International Journal of Forecasting*, Bd. 27, Nr. 2, S. 452–465.
- Driscoll, J. C. / Kraay, A. C. (1998): „Consistent Covariance Matrix Estimation With Spatially Dependent Panel Data“, *Review of Economics & Statistics*, Bd. 80, Nr. 4, S. 549–560.
- Fama, E. (1970): „Efficient Capital Markets, A Review of Theory and Empirical Work“, *Journal of Finance*, Bd. 25, S. 383–417.
- Gubaydullina, Z. / Hein, O. / Spiwoks, M. (2011): „The Status Quo Bias of Bond Market Analysts“, *Journal of Applied Finance & Banking*, Bd. 1, Nr. 1, S. 31–51.
- Heilemann, U. / Klinger, S. (2006): „Zu wenig Wettbewerb?: Zu Stand und Entwicklung der Genauigkeit makroökonomischer Prognosen“. In *Wirtschaftspolitik im Systemwettbewerb, Schriften des Vereins für Socialpolitik*, Berlin, S. 225–257.

- Hyndman, R. J. / Koehler, A. B. (2006): „Another look at measures of forecast accuracy“, *International Journal of Forecasting*, Bd. 22, Nr. 4, S. 679–688.
- Isiklar, G. / Lahiri, K. (2007): „How far ahead can we forecast? Evidence from cross-country surveys“, *International journal of forecasting*, Bd. 23, Nr. 2, S. 167–187.
- Isiklar, G. / Lahiri, K. / Loungani, P. (2006): „How Quickly Do Forecasters Incorporate News? Evidence from Cross-Country Surveys“, *Journal of Applied Econometrics*, Bd. 21, Nr. 6, S. 703–725.
- Lahiri, K. / Sheng, X. (2008): „Evolution of forecast disagreement in a Bayesian learning model“, *Journal of Econometrics*, Bd. 144, Nr. 2, S. 325–340.
- Loungani, P. (2001): „How accurate are private sector forecasts? Cross-country evidence from consensus forecasts of output growth“, *International journal of forecasting*, Bd. 17, Nr. 3, S. 419–432.
- Lucas, R. E. (1972): „Expectations and the neutrality of money“, *Journal of Economic Theory*, Bd. 4, Nr. 2, S. 103–124.
- Mankiw, N. G. / Reis, R. / Wolfers, J. (2003): „Disagreement About Inflation Expectations“, *NBER Macroeconomics Annual*, Bd. 18.
- Muth, J. F. (1961): „Rational Expectations and the Theory of Price Movements“, *Econometrica*, Bd. 29, Nr. 3, S. 315–335.
- Nordhaus, W. D. (1987): „Forecasting Efficiency: Concepts and Applications“, *The review of economics and statistics*, Bd. 69, Nr. 4, S. 667–674.
- Osterloh, S. (2008): „Accuracy and Properties of German Business Cycle Forecasts“, *Applied Economics Quarterly*, Bd. 54, Nr. 1, S. 27–57.
- Sims, C. A. (2003): „Implications of rational inattention“, *Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information*, Bd. 50, Nr. 3, S. 665–690.
- Spiwoks, M. (2004): „External Triggered Herding bei Rentenmarkt-Analysten“, *Financial Markets and Portfolio Management*, Bd. 18, Nr. 1, S. 58–83.
- Spiwoks, M. / Bedke, N. / Hein, O. (2008): „Forecasting the past - The case of U.S. interest rate forecasts“, *Financial Markets and Portfolio Management*, Bd. 22, S. 357–379.
- Spiwoks, M. / Bedke, N. / Hein, O. (2010): „Topically Orientated Trend Adjustment and Autocorrelation of the Residuals - An Empirical Investigation of the Forecasting Behavior of Bond Market Analysts in Germany“, *Journal of Money, Investment and Banking*, Nr. 14, S. 16–35.
- Tversky, A. / Kahneman, D. / Choice, R. (1981): „The framing of decisions“, *Science*, Bd. 211, S. 453–458.

Woodford, M. (2001): „Imperfect Common Knowledge and the Effects of Monetary Policy“, *National Bureau of Economic Research Working Paper Series*, Bd. No. 8673.

Kapitel 5

Planspiel Kapitalmarktprognose

*Ein empirischer Vergleich der Prognosekompetenz von Amateuren
und Experten*

mit Kilian Bizer und Markus Spiwoks

Erschienen als:

Studie der Sonderforschungsgruppe Institutionenanalyse (sofia), Nr. 13-2, Darmstadt 2013.

Abstract

In dieser Studie werden die Ergebnisse eines Planspiels vorgestellt, bei dem Anfänger über einen längeren Zeitraum hinweg Aktienkurse einer empirischen Zeitreihe vorhergesagt haben. Ihnen standen neben vergangenen Kursdaten auch Fundamentaldaten in Form von Presseartikeln zur Verfügung. Im Resultat sind die Prognosen nicht besser oder schlechter als empirische Prognosen der Experten: Sie sind schlechter als die naive Prognose, entsprechen nicht den Kriterien der rationalen Erwartungsbildung und weisen ebenso starke gegenwarts-orientierte Verlaufsanpassungen auf. Zudem zeigt sich die von den Teilnehmern angegebene Unsicherheit bei der Prognoseabgabe als Gradmesser der Divergenz der Prognosen untereinander, so wie es für die Prognosen echter Analysten auch angenommen wird. Zusammengefasst lässt sich sagen, dass es offensichtlich nicht spezieller Institutionen im Umfeld beruflicher Kapitalmarktanalysten bedarf, um diese Phänomene verursachen.

1 Einleitung

Erfolgreiche Kapitalmarktprognosen sind von herausragender Bedeutung, beispielsweise für Manager aktiver Wertpapierfonds, im Bereich der Fristentransformation bei Banken oder auch für Anlage- und Investitionsentscheidungen von Unternehmen und Privatpersonen. Dennoch sind sie in der Regel von schlechter Qualität. Zudem weisen sie spezifische Charakteristika auf, wie beispielsweise systematische Fehler oder gegenwartsorientierte Verlaufsanpassungen. Neben den zahlreichen Evaluationen von Expertenprognosen, die zu dieser Feststellung führen¹, werden immer wieder einzelne Aspekte mit Laborexperimenten überprüft, die jedoch den üblichen Beschränkungen wie dem knappen Zeitbudget und der Frage nach der Übertragbarkeit der Ergebnisse auf die Realität unterworfen sind.²

Diese Studie greift auf Ergebnisse einer ungewöhnlichen Datenerhebungsmethode³ zurück. Studierende prognostizieren über ein ganzes Semester hinweg in Form eines Planspiels einen Aktienkurs. Anders als bei herkömmlichen Planspielen wird hier jedoch auf einen modellhaften Algorithmus, der die Inputdaten (Einflussfaktoren auf den Aktienkurs) mit den Outputdaten (Aktienkurse) verknüpft, verzichtet. Stattdessen greift das vorliegende Planspiel auf einen tatsächlichen Aktienkursverlauf der 1980er Jahre zurück. Die Studierenden, die an diesem Planspiel teilnehmen, werden mit realen Presseartikeln dieser Zeit über die tatsächlichen Einflussfaktoren (Umsätze, Gewinne, Beschäftigung, Übernahmen, Wechselkurse, gesamtwirtschaftliches Wachstumstempo, Konsumentenstimmung etc.) informiert. Allerdings wurden alle Informationen so aufbereitet, dass weder das tatsächliche Unternehmen noch der tatsächliche Zeitraum erkennbar war. Auf diese Weise konnte eine starke Realitätsnähe des Planspiels erreicht werden. Etwaige Verzerrungen durch die Wahl eines herkömmlichen Planspiel-Algorithmus wurden somit vermieden.

Die Studierenden geben nicht nur Kursprognosen ab, sondern auch Einschätzungen ihrer bei der Prognoseabgabe empfundenen Sicherheit.

¹ siehe Kapitel 2

² Für eine Übersicht zu experimentellen Studien im Bereich der Wirtschafts- und Kapitalmarktprognosen siehe Leitner und Leopold-Wildburger (2011).

³ Teile des Designs und vor allen Dingen die umfangreiche Recherche historischer Presseartikel ist der Bearbeitung einer Diplomarbeit an der Universität Göttingen zu verdanken. Ursprünglich sollte das Spiel an Schulen gespielt werden. Dem Bearbeiter Kristian Horn und der Betreuerin Dr. Zulia Gubaydullina gebührt unser Dank.

Neben den empirischen Presseartikeln und den wichtigsten Unternehmenskennzahlen stehen den Studierenden auch die Kursdaten der Zeit vor dem jeweiligen Prognosezeitpunkt zur Verfügung, was die Prognoseaufgabe deutlich zeitaufwendiger aber realitätsnäher als bei Laborexperimenten gestaltet. Sie arbeiten über die gesamte Laufzeit in festen Gruppen zusammen. Ihren Anreiz zur erfolgreichen Prognoseabgabe beziehen sie aus dem Abschluss einer Lehrveranstaltung, in der Sie die Grundzüge der fundamentalen Aktienanalyse lernen.

Ziel dieser Studie ist ein Vergleich der Prognosen dieser Anfängergruppen mit den Charakteristika empirischer Kapitalmarktprognosen von Experten. Gemeinsamkeiten und Unterschiede werden identifiziert und diskutiert, insbesondere vor dem Hintergrund des unterschiedlichen institutionellen Umfelds, in dem die Prognosen abgegeben werden. Folgende Phänomene werden einer näheren Betrachtung unterzogen, wobei die detaillierte Erläuterung der einzelnen Phänomene im 2. Kapitel erfolgt:

- Die Prognosequalität in Form der Höhe der Prognosefehler in Bezug auf die naive Prognose⁴.
- Die Form der Erwartungsbildung der Prognostiker, insbesondere die Frage, ob systematische Prognosefehler vorliegen.
- Der Gegenwartsbezug der Prognosen und die Frage, ob das Ausmaß der künftigen Änderungen realistisch eingeschätzt wird.
- Die Frage, ob die Unsicherheit der Prognostiker mit einem divergenten Prognoseverhalten der Gruppen einhergeht.

Das zweite Kapitel dient der Schilderung der angesprochenen Phänomene, der Anknüpfung an den aktuellen Stand der Forschung sowie der Formulierung der Forschungshypothesen. Die ausführliche Erläuterung und kritische Hinterfragung des Planspieldesigns erfolgt im 3. Kapitel. Das 4. Kapitel dient der Darstellung der Ergebnisse, wobei die zu deren Erreichung benötigten Verfahren ebenfalls im 4. Kapitel erläutert werden. Im 5. Kapitel werden die Ergebnisse zusammengefasst und mögliche Anknüpfungspunkte für weiterführende Forschungsansätze dargestellt. Der umfangreiche Anhang⁵ dieser Studie enthält alle Informationen, die den Stu-

⁴ Die naive Prognose ist die Annahme, dass der zukünftige Kurs unverändert bleibt.

⁵ Im Rahmen dieser Dissertation wird aus Platzgründen ein verkürzter Anhang präsentiert. Der vollständige Anhang ist der bei der Sonderforschungsgruppe Institutionenanalyse veröffentlichten Studie (Nr. 13-2) zu entnehmen. Ein Download ist möglich unter <http://www.sofia-darmstadt.de/studien.0.html>

dierenden während des Planspiels zur Verfügung standen. Diese unterteilen sich in die Fragebögen, in denen die Gruppen ihre Angaben zu den Prognosen und der jeweiligen Sicherheit machen (Anhang B) sowie die Presseartikel und Unternehmenskennzahlen, auf deren Basis die Prognosen abgegeben werden konnten (Anhang C).

2 Stand der Forschung und Hypothesenbildung

In diesem Abschnitt werden verschiedene Aspekte von Kapitalmarktprognosen auf Basis der vorhandenen Literatur erläutert. Darauf aufbauend werden jeweils Untersuchungshypothesen formuliert. Abschnitt 2.1 zeigt auf, wie es um die Treffsicherheit von empirischen Kapitalmarktprognosen bestellt ist. In 2.2 geht es um die Frage, welche Formen der Erwartungsbildung in der Literatur diskutiert werden und inwiefern Kapitalmarktprognosen von Experten dem Kriterium der rationalen Erwartungsbildung entsprechen. In Bezug auf diese Studie ist besonders der Aspekt der systematischen Prognosefehler von Relevanz. In Abschnitt 2.3 geht es um den starken Gegenwartsbezug von Kapitalmarktprognosen. In Abschnitt 2.4 wird erläutert, wie sich die Uneinigkeit der Prognostiker mit ihrem eigenen Vertrauen in die Prognosequalität in Verbindung bringen lässt. Abschnitt 2.5 fasst das Kapitel 2 zusammen.

2.1 Treffsicherheit von Aktienkursprognosen

Das Ziel von Aktienkursprognosen sollte es sein, erfolgreiche Anlageempfehlungen zu ermöglichen. Als Benchmark für den Erfolg solcher Prognosen hat sich der Vergleich zu sogenannten naiven Prognose etabliert. Die Idee der naiven Prognose geht auf den französischen Mathematiker Laplace (1814) zurück. In einer Situation, in der nichts über den zukünftigen Verlauf einer Zeitreihe bekannt ist, gilt nach Laplace das Prinzip des unzureichenden Grundes. In diesem Fall besteht die einzig rationale Prognosestrategie darin, von einem gleichbleibenden Wert auszugehen, der sogenannten naiven Prognose. Um Anlageempfehlungen auf Basis einer Prognose umzusetzen zu können, sollte die Qualität der naiven Prognose übertroffen werden.

Empirische Untersuchungen zur Qualität von Aktienkursprognosen sind nicht besonders häufig. Spiwoks und Hein (2007) führen das darauf zurück, dass Aktienanalysten vorwiegend Gewinnprognosen oder Anlageempfehlungen abgeben. Sie kommen wie zuvor schon Spiwoks (2004a) bei einer Untersuchung von Aktienindexprognosen des Zentrums für Europäische Wirtschaftsforschung (ZEW), die auf Basis monatlicher Umfragen von 400 Kapitalmarktexperten generiert wurden, zu dem Schluss, dass beinahe alle betrachteten Zeitreihen schlechter

als die naiven Prognosen abschneiden. Dies schließt neben Prognosen für die wichtigsten Aktienindizes auch Prognosezeitreihen für das Zinsniveau der größten Industrieländer als auch die Wechselkurse zwischen den wichtigsten Währungen mit ein.

Mit der halbjährlich erhobenen Livingstone Survey⁶ wertet Söderlind (2010) eine der am häufigsten⁷ betrachteten Datenquellen aus. Er kommt zu dem Schluss, dass die Konsensprognosen der zahlreichen namhaften US-amerikanischen Ökonomen schlechter als die naiven Prognosen und einfache standardisierte Prognosemodelle abschneiden. Eine weitere Datenquelle stellt eine Umfrage des Handelsblatts dar, die seit Anfang der 90er Jahre immer am Jahresende durchgeführt wird. Zahlreiche namhafte Banken geben ihre Einschätzung zum Zinsniveau, dem Wechselkurs EUR/USD sowie dem Kurs des DAX am Ende des folgenden Jahres ab. Benke (2006) kommt zu dem Schluss, dass sich diese Prognosen nicht für die Umsetzung von aktiven Anlagestrategien eignen. Sie entsprechen der Qualität eines Münzwurfs.

Wenngleich die Aktienkursprognosen in den vorgestellten Studien mit sehr unterschiedlichen Verfahren generiert worden sind, wird die schlechte Qualität konstant festgestellt. Für Prognosen anderer Kapitalmarktdaten wie dem Zinsniveau und den Wechselkursen ist die Datenlage erheblich umfangreicher und es gibt Hunderte von Studien. Der Grundtenor der schlechten Qualität ist anders als bei Konjunktur- und Inflationsprognosen jedoch im Grundsatz immer derselbe. Hervorzuheben ist eine experimentelle Studie von Leitner und Schmidt (2006), bei der die Probanden Prognosen für den Verlauf einer empirischen Wechselkurszeitreihe ausschließlich auf der Basis vergangener Kursdaten abgegeben haben. Im Ergebnis haben die Prognosen im Durchschnitt zwar schlechter als die naive Prognose abgeschnitten, waren überraschenderweise jedoch besser als alle zum Vergleich herangezogenen empirischen Prognosen verschiedener professioneller Prognostiker.

Für die Prognosen im Planspiel wäre eine bessere Prognosequalität als die naive Prognose im Hinblick auf den Stand der Forschung eine große Überraschung, zumal es sich bei den Prognostikern um Studierende ohne Prognoseerfahrung handelt. Sie erstellen die Prognosen auf Basis eines Bruchteils der empirischen Informationsbasis und müssen zudem eine bestimmte Zeitvorgabe einhalten. Die erste Forschungshypothese lautet:

⁶ Die Livingstone Survey beinhaltet die Ergebnisse einer halbjährlich durchgeführten Befragung zahlreicher US-amerikanischer Ökonomen. Diese sind sowohl bei Banken als auch in der Industrie, an Universitäten und bei der Regierung tätig. Für die Studie von Söderlind (2010) wurden Prognosen des S&P Industrial Indexes (1952-1990) sowie der S&P 500 (1990-2007) ausgewertet.

⁷ Für eine Übersicht zu älteren Studien basierend auf der gleichen Datenbasis siehe Croushore (1997).

H1: Die Prognosen der Teilnehmer schneiden schlechter als empirische Kapitalmarktprognosen von Experten ab.

2.2 Erwartungsbildungstheorie

Unabhängig von der Höhe der Prognosefehler im Vergleich zur naiven Prognose sind die Fragen des Zustandekommens der einzelnen Prognosen und die Art und Weise des Scheiterns der Prognoseanstrengungen im Allgemeinen Gegenstand zahlreicher Untersuchungen. Die am häufigsten diskutierte theoretische Grundlage hierfür liefert die Hypothese rationaler Erwartungsbildung von Muth (1961). Deren Entwicklung war eine Reaktion auf die bis dahin von Keynes (1936) geprägten Sicht der Erwartungsbildung von Wirtschaftsakteuren. Bereits gut 40 Jahre vor der Formulierung der Prospect Theory⁸ durch Kahneman und Tversky (1979) hat Keynes auf die Bedeutung der kognitiven Grenzen der Akteure bei der Erwartungsbildung hingewiesen. Die Erwartungsbildung auf der Basis mathematisch-statistisch korrekter Berechnungen erschien ihm realitätsfremd, was im Konflikt zu den später für die neoklassischen Gleichgewichtsmodelle benötigten Annahmen über die Erwartungsbildung der Wirtschaftsakteure stand.

Muth (1961) hat die Hypothese der rationalen Erwartung formuliert und vor allem Lucas (1972) hat ihr zum Durchbruch in der Makroökonomie verholfen. Der wesentliche Unterschied zu der bis dahin vorherrschenden Sicht der Erwartungsbildung besteht darin, dass neben der vergangenen Entwicklung auch alle weiteren Informationen über die Zukunft in die Erwartungsbildung einbezogen werden. So antizipieren die Wirtschaftsakteure nach dieser These auch strukturelle Änderungen der Volkswirtschaft durch Eingriffe der Politik. Es wird von „rationaler“ Erwartungsbildung gesprochen, weil die Akteure auf Basis aller ihnen zur Verfügung stehenden Informationen die bestmögliche Prognose abgeben.

Anders als Keynes sieht die Hypothese rationaler Erwartungsbildung keinerlei Spielraum für irrationales Verhalten, was bedeutet, dass Emotionen und kognitive Grenzen etc. keine Auswirkungen auf die Erwartungen der Akteure haben sollten. Keinesfalls jedoch bedeuten Erwartungen auf rationaler Basis völlige Voraussicht über die zukünftigen Entwicklungen. Es müssen lediglich alle nach dem ökonomischen Prinzip tragbaren Anstrengungen unternommen worden sein, diese zu erreichen.⁹

⁸ Die Prospect Theory wird im Deutschen auch „Neue Erwartungstheorie“ genannt.

⁹ Svindland (1983), S. 333.

Basierend auf den theoretischen Überlegungen von Muth und Lucas müssen für Wirtschaftsprognosen folgende Bedingungen erfüllt sein, um sie als „rational“ gelten zu lassen:

- Die Prognosequalität ist nicht schlechter als die (in Abschnitt 2.1 erläuterte) naive Prognose, da sonst die kostenlose Alternative der naiven Prognose vorzuziehen wäre.
- Die Prognosefehler entsprechen einem stochastischen Störterm und weisen somit keinerlei Systematik auf.
- Alle relevanten und öffentlich verfügbaren Informationen werden adäquat ausgewertet und sind Bestandteil der Prognosen.¹⁰ Dies schließt die eigenen Prognosefehler mit ein.

Zur Überprüfung dieser Bedingungen sind verschiedene Testverfahren entwickelt worden. Für den Vergleich mit der naiven Prognose werden vorwiegend der mittlere relative absolute Prognosefehler bezogen auf die naive Prognose (*MRAPnP*) und der Theilsche Ungleichheitskoeffizient¹¹ (Abschnitt 4.1.2) berechnet. Häufig findet auch der Test auf Forecast Encompassing nach Diebold und Mariano (1995) Anwendung. Das Vorliegen systematischer Prognosefehler wird in der Regel mit dem Test auf Unverzerrtheit¹² überprüft (Abschnitt 4.2.2). Eine Beurteilung der Informationsauswertung erfolgt vorwiegend mittels des Effizienztests¹³, welcher in dieser Studie allerdings unberücksichtigt bleibt.

Die im vorigen Abschnitt vorgestellte Livingston Survey ist auf die Rationalität der Prognostiker im Hinblick auf ihre Prognosen des S&P Industrial Index überprüft worden. Lakonishok (1980) zeigt, dass die Konsensprognose¹⁴ aller Teilnehmer systematische Prognosefehler aufweist. Nachdem Figlewski und Wachtel (1981) darauf aufmerksam machen, dass eine Untersuchung der Konsensprognosen keinesfalls Rückschlüsse auf die Rationalität der Einzelprognosen zulassen, kommt Pearce (1984) bei einer Analyse der individuellen Prognosen ebenfalls zu dem Ergebnis, dass umfassende systematische Prognosefehler vorliegen. Dokko und Edelstein (1989) widersprechen den Resultaten. Sie weisen darauf hin, dass die Prognosen nicht gleichzeitig an einem Stichtag abgegeben werden und die tagesaktuellen Schwankungen der

¹⁰ Vgl. hierzu Nordhaus (1987). Bei der strengen Form der rationalen Erwartungsbildung müssten auch nicht-öffentliche Informationen berücksichtigt werden. Nordhaus schlägt daher die schwache Form der rationalen Erwartungsbildung vor, bei der nur öffentlich Informationen berücksichtigt werden müssen.

¹¹ Vgl. Theil (1961).

¹² Vgl. Mincer und Zarnowitz (1969).

¹³ Vgl. Holden und Peel (1990).

¹⁴ Bei einer Konsensprognose handelt es sich um den arithmetischen Mittelwert der Einzelprognosen.

Börsen berücksichtigt werden müssten. Mit einer angepassten Methodik finden Sie die Hinweise auf die systematischen Prognosefehler nicht mehr.

Widersprüchliche Resultate liefern auch experimentelle Ansätze zur Erforschung der rationalen Erwartungsbildung bei Prognosen. Während Garner (1982) keine rationale Erwartungsbildung bei den Prognosen der Probanden feststellt, sieht Beckman (1992) bei einfachen Prognosen mit kurzem Horizont keine Anzeichen für Irrationalität. Er und auch Becker und Bolle (1996) sehen die hohe Komplexität des zugrundeliegenden Systems zur Entwicklung der Zeitreihen als Ursache für die Irrationalität. Sie widersprechen damit Keane und Runkle (1990), die die Ursachen hierfür eher in den zur Prognose führenden Daten und deren fortwährenden Revisionen sehen. Brennscheidt (1993) weist in einer Serie von Experimenten nach, dass die Probanden die Prognosen in der Regel zu hoch ansetzen und das Ausmaß der Änderungen systematisch unterschätzen. Leitner und Schmidt (2006) kommen bei einem Experiment zu Wechselkursprognosen von Anfängern zum gleichen Schluss wie Beckman (1992). Bei den 1-Monats-Prognosen sind die Prognosen rational, bei längeren Prognosehorizonten nicht.

Die Mehrzahl der Laborexperimente befasst sich darüber hinaus vorwiegend mit der Frage, ob spezifische zusätzliche Informationen von den Probanden ausreichend berücksichtigt worden sind.¹⁵ Dies bietet sich bei dem in dieser Studie verwendeten offenen Design nicht an und ist daher auch nicht Gegenstand der Auswertung. Die Überprüfung auf systematische Komponenten bei den Prognosefehlern ist jedoch ohne weiteres möglich (Abschnitt 4.2.2). Vor dem Hintergrund des komplexen Prognoseproblems und der zusätzlichen Herausforderung durch eine nicht-stationäre Zeitreihen wären Prognosen ohne systematische Fehler eher eine Überraschung.

H2: Die Prognosefehler weisen systematische Komponenten auf und entsprechen somit nicht den Kriterien der rationalen Erwartungsbildung.

Bei einer vorläufigen Bestätigung der Hypothese stellt sich die Frage, ob die Planspielteilnehmer einfachere Formen der Erwartungsbildung aufweisen (Abschnitt 4.2.3). Es wird geprüft, ob sie Trends extrapolieren (extrapolative Erwartungsbildung) oder zumindest ihre eigenen Prognosefehler zum Anlass für eine Korrektur ihrer Prognosen nehmen (adaptive Erwartungsbildung).

¹⁵ Für eine Übersicht vgl. Leitner und Leopold-Wildburger (2011).

2.3 Gegenwartsorientierte Verlaufsanpassung

Die Erstellung von Wirtschaftsprognosen erfolgt immer vor dem Hintergrund eines Informationsstands, der sehr stark von den jeweiligen Gegebenheiten zum Zeitpunkt der Prognoseabgabe geprägt ist. In der Literatur wird debattiert, ob dies dazu führen kann, dass die Prognosen nur noch die Gegenwart widerspiegeln und keinen Zukunftsbezug aufweisen. Zunächst Manzur (1988) und Allen und Taylor (1990) stellen für Analysteneinschätzungen von Wechselkursen für Prognosehorizonte von 4 Wochen und weniger fest, dass die Prognosen der tatsächlichen Entwicklung in Höhe des Prognosehorizontes nachlaufen. Die Prognostiker scheinen die zukünftige Entwicklung vom Ausmaß her zu unterschätzen und bleiben mit ihren Prognosen sehr nah an der Gegenwart zum Zeitpunkt der Prognoseerstellung. Andres und Spiwoks (1999) nennen dieses Phänomen gegenwartsorientierte Verlaufsanpassung (GOVA) und stellen ein Kriterium auf, anhand dessen sich die GOVA nachweisen lässt.¹⁶

In der Folgezeit wurde das Phänomen für Prognosen des Zinsniveaus¹⁷ sowie von Wechsel- und Aktienkursen¹⁸ nachgewiesen. Offen sind die Ursachen für das Phänomen. Spiwoks (2004b) sieht eine spezielle Form des Herdenverhaltens¹⁹ als mögliche Ursache, bei der sich die Analysten am jeweils aktuellen Wert zum Zeitpunkt der Prognoseerstellung orientieren. Die Komplexität und Unsicherheit bei der Prognoseaufgabe könnte auch zur Anwendung von Heuristiken führen.²⁰ So bringen Bofinger und Schmidt (2003) die Ankerheuristik²¹ als mögliche Ursache in die Debatte ein, bei der die Prognostiker unbewusst vom jeweiligen Gegenwartswert bezüglich ihrer Prognoseentscheidung beeinflusst werden. Campbell und Sharpe (2007) meinen dieses Phänomen bei Zinsprognosen nachweisen zu können. Gubaydullina u. a. (2011) halten es für möglich, dass bei Zinsprognosen ein Status quo bias²² vorliegt, also eine kognitive Grenze

¹⁶ Zur Messung, ob eine GOVA vorliegt, schlagen sie die Bildung eines Koeffizienten aus den Bestimmtheitsmaßen der Zusammenhänge zwischen den Prognosen und den tatsächlich eingetretenen Werten sowie zwischen den Prognosen und den Werten zum Zeitpunkt der Prognoseerstellung vor. Die Erläuterung dieses sogenannten GOVA-Koeffizienten erfolgt in Abschnitt 4.3.2.

¹⁷ Spiwoks (2002a) und Spiwoks u. a. (2010) für Deutschland, Scheier und Spiwoks (2006) für Großbritannien, Bedke (2008) sowie Spiwoks u. a. (2008) für die USA und Spiwoks u. a. (2009) für die Schweiz.

¹⁸ Bofinger und Schmidt (2003) für mehrere Wechselkurse und Spiwoks (2004a) für Aktienindexprognosen.

¹⁹ Keynes (1936) hat gezeigt, dass es für Finanzmarktanalysten durchaus rational sein kann, sich bei seinen Einschätzungen an seinen Berufskollegen zu orientieren. Er argumentiert so, dass es weniger schlimm sei, in der Gruppe falsch zu liegen, als dies bei einer Außenseitereinschätzung der Fall wäre.

²⁰ Kahneman und Tversky (1979) zeigen mit der „Prospect Theory“, dass der Mensch bei Entscheidungen häufig auf Daumenregeln zurückgreift, die ihm die Entscheidung erleichtern. Dabei kommt es jedoch häufig zu Entscheidungen, die dem Rationalitätsprinzip im ökonomischen Sinne widersprechen.

²¹ Die Ankerheuristik (Anchoring bias) ist erstmals von Tversky und Kahneman (1974) beschrieben worden.

²² Der Begriff des Status quo bias ist von Samuelson und Zeckhauser (1988) in die wissenschaftliche Debatte eingeführt worden.

der Prognostiker, die sie daran hindert, das Ausmaß der zukünftigen Änderungen zu erfassen. Sie weisen nach, dass die Prognostiker systematisch zu geringe Änderungsraten der Zinssätze prognostizieren.

Da die GOVA bei Kapitalmarktprognosen mit großer Regelmäßigkeit festgestellt wird, wäre dies für die Prognosen der Studierenden ebenfalls nicht verwunderlich. Es ist gut möglich, dass die Prognoseentscheidungen durch die Ankerheuristik oder das Vorliegen eines Status quo bias beeinflusst werden.

Die Hypothese lautet:

H3: Die Prognosen weisen gegenwartsorientierte Verlaufsanpassung auf.

2.4 Uneinigkeit und Unsicherheit der Prognostiker

Für die Nutzer von Prognosen ist es von herausragender Bedeutung, deren ex-ante Zuverlässigkeit zu kennen. So verwenden beispielsweise sowohl die Bank von England als auch die Schwedische Reichsbank seit Mitte der 90er Jahre Diagramme, die Konfidenzintervalle um ihre offiziellen Prognosen herum darstellen. In Deutschland ist die Debatte um die Sicherheit der Prognosen besonders auf dem Höhepunkt der Finanzkrise geführt worden, als der damalige DIW-Chef Klaus Zimmermann seine Kollegen dazu aufforderte, auf die Abgabe von Prognosen vorerst zu verzichten, da den vorhandenen Modellen angesichts der außergewöhnlichen Situation keinerlei Wert mehr beizumessen sei. Zu einer Kultur der gleichzeitigen Anzeige der (subjektiven) Prognosesicherheit bei der Veröffentlichung von Prognosen hat die damalige Debatte in Deutschland jedoch nicht geführt.

Der erste grundlegende Aufsatz zu dem Thema stammt von Zarnowitz und Lambros (1987). Sie schlagen vor, das Ausmaß der Streuung der Einzelprognosen zu einem Ereignis als Äquivalent für die Zuverlässigkeit der Prognosen zu verwenden. Unter den Ökonomen herrscht eine nicht abgeschlossene Debatte darüber, ob die Uneinigkeit der Prognostiker ein geeignetes Maß für die Unsicherheit der Prognosen darstellt. Bomberger (1996) sowie Giordani und Söderlind (2003) sprechen der Hypothese ihre theoretische Fundierung ab. Lahiri und Sheng (2010) halten die Uneinigkeit der Prognostiker für ein geeignetes Instrument zur Bestimmung der Unsicherheit, wenn nämlich stabile Verhältnisse herrschen und die prognostizierten Variablen kaum von exogenen Schocks beeinflusst werden. Darüber hinaus nimmt die Verlässlichkeit mit der Länge des Prognosehorizonts ab.

Speziell an den Kapitalmärkten stellt sich die Frage, ob eine erhöhte Unsicherheit zu einer trägeren Informationsverarbeitung und somit einer ineffizienten Preisfindung führt. Zhang (2006) schlägt neben der Einigkeit der Analysten weitere Möglichkeiten wie die Firmengröße und die Cash-Flow Volatilität als mögliche Anzeige für den Grad der Sicherheit vor. Vor allem nach schlechten Nachrichten konnten die verschiedenen Stellvertretervariablen für die Unsicherheit zuverlässig ein zu träges Marktverhalten prognostizieren. Die Uneinigkeit der Analysten bei ihren Einschätzungen zu den Aktienkursen war hierfür jedoch kein passender Indikator.

Im Rahmen des im nächsten Kapitel erläuterten Designs des Planspiels erfolgt eine Abfrage der subjektiv empfundenen Sicherheit der Planspielgruppen. Damit lässt sich folgende Hypothese überprüfen:

H4: Die subjektiv empfundene Unsicherheit der Gruppen stimmt mit der anhand des Kriteriums Prognosedivergenz objektiv messbaren Unsicherheit überein.

2.5 Zusammenfassung

Unter Berücksichtigung der empirischen und experimentellen Literatur in Kombination mit dem im nächsten Abschnitt erläuterten Untersuchungsdesign werden folgende Hypothesen überprüft:

H1: Die Prognosen der Teilnehmer schneiden schlechter als empirische Kapitalmarktprognosen von Experten ab.

H2: Die Prognosefehler weisen systematische Komponenten auf und entsprechen somit nicht den Kriterien der rationalen Erwartungsbildung.

H3: Die Prognosen weisen gegenwartsorientierte Verlaufsanpassung auf.

H4: Die subjektiv empfundene Unsicherheit der Gruppen stimmt mit der anhand des Kriteriums Prognosedivergenz objektiv messbaren Unsicherheit überein.

Die Ergebnisse und deren ausführliche Diskussion sind Bestandteil von Kapitel 4. Dabei erfolgt auch ein Vergleich der Ergebnisse mit anderen Studien.

3 Durchführung des Planspiels

In diesem Kapitel wird die Vorgehensweise bei der Durchführung des Planspiels geschildert. Das allgemeine Design zur Erhebung der Prognosedaten wird in Abschnitt 3.1 erläutert. Die Besonderheit der Untersuchungsmethode ist die umfangreiche Informationsbasis, auf deren Grundlage die Teilnehmer die Prognosen abgeben. Diese wird in Abschnitt 3.2 geschildert. Die gewählte Untersuchungsmethode ist eine Mischung zwischen ökonomischem Labor- und Feldexperiment. Eine entsprechende Diskussion hierzu findet sich in Abschnitt 3.3. Das Planspiel ist im Sommersemester 2012 an der Ostfalia in Wolfsburg durchgeführt worden. Eine Schilderung des Ablaufs mit beschreibenden Statistiken erfolgt in Abschnitt 3.4. In Abschnitt 3.5 wird das 3. Kapitel zusammengefasst.

3.1 Design

Die im 2. Kapitel aufgezeigten Charakteristika, welche sich bei der Evaluation von Kapitalmarktprognosen zeigen, werden auf ihren Bestand im Kontext eines Prognoseplanspiels mit Studierenden überprüft. Dem liegt die Idee zugrunde, dass bestimmte Institutionen im Umfeld der Kapitalmarktanalysten auf das Prognoseverhalten dahingehend wirken könnten, dass sich die Charakteristika nur bei Prognosen von Experten herausbilden.

Bei ökonomischen Experimenten besteht das Grunddesign zumeist darin, die Wirkung einer bestimmten Maßnahme dadurch zu überprüfen, dass man alle Faktoren außer der untersuchten Variable konstant hält. Auf ein solches Design, welches in der Regel eine Unterscheidung in eine Kontroll- und eine Behandlungsgruppe vorsieht, wird bewusst verzichtet. Es geht nicht darum, einen speziellen Faktor bei der Erstellung von Kapitalmarktprognosen einer genauen Prüfung zu unterziehen. Dem Design liegt ein sehr allgemeines Verständnis eines Experiments zugrunde, dass sich als methodisch angelegte Untersuchung zur empirischen Datengewinnung beschreiben lässt. Ein Vergleich erfolgt nicht zwischen Behandlungs- und Kontrollgruppe, sondern zwischen den im Planspiel generierten Daten und Daten, die bereits im Rahmen der Literatur bei Evaluationen von Kapitalmarktprognosen veröffentlicht worden sind.

Das Experiment ist folgendermaßen gestaltet: Die Durchführung erfolgt im Rahmen eines Planspiels, welches mit dem Titel „Stock Forecasting“ an einer Hochschule beworben wird. Die

Studierenden erhalten für die Teilnahme Credit Points²³, womit ihnen ein Anreiz zur Teilnahme gegeben wird. Da die Studierenden in der Regel eine Vielzahl an Möglichkeiten bezüglich der Belegung von Wahlpflichtfächern haben, ist die Rekrutierung der Teilnehmer dahingehend verzerrt, dass sie sich für ein Fach mit einem Planspiel im Allgemeinen und Aktienkursprognosen im Speziellen interessieren. Die Teilnahme ist freiwillig und die bei ökonomischen Experimenten üblichen monetären Zahlungen, die ein Anreizsystem über die Erbringung einer Studienleistung hinaus implementieren würden, werden nicht geleistet. Um die Motivation während des Planspiels hochzuhalten, werden regelmäßig die Zwischenergebnisse der besten 3 Gruppen veröffentlicht. Außerdem wird den Teilnehmern bei der Erläuterung der Regeln mitgeteilt, dass die Qualität ihrer Prognosen einen Einfluss auf die Benotung haben wird.²⁴ Selten u. a. (2003) zeigen, dass auch ein nicht-monetäres Anreizsystem in Form von Noten im Rahmen einer Hochschulveranstaltung zu großem Engagement der Teilnehmer an einem Experiment führen kann.

In einer Einführungspräsentation werden die Teilnehmer über die Grundzüge der fundamentalen Aktienanalyse informiert (ca. 45 Minuten). Anschließend erfolgt eine Erläuterung der Spielregeln. Es wird ihnen mitgeteilt, dass es Ziel des Spiels ist, möglichst präzise Prognosen über den künftigen Aktienkursverlauf eines deutschen Industrieunternehmens abzugeben. Zum Abschluss der ersten Veranstaltung erfolgt eine zufällige Aufteilung der Teilnehmer in Gruppen von 3-4 Studierenden.

Das Planspiel wird rundenbasiert durchgeführt, wobei jede Runde in der gleichen Gruppenzusammensetzung und mit den gleichen Rahmenbedingungen abläuft. Die Gruppen erhalten jeweils zu Rundenbeginn:

- Ein Chart mit dem Aktienkurs des Unternehmens in den letzten 6 Spielrunden (Anhang B).

²³ Bei Credit Points handelt es sich um eine Verrechnungseinheit für erfolgreich abgeschlossene Studienleistungen. Für die Teilnahme am Planspiel und eine Hausarbeit zur fundamentalen Aktienanalyse erhalten die Studierenden 3 Credit Points. Das gesamte Bachelorstudium setzt an der Ostfalia erfolgreich abgeschlossene Studienleistungen von 180 Credits Points in 6 Semestern voraus. Die Benotung der Fachs erfolgt anhand einer Hausarbeit zur Fundamentalanalyse.

²⁴ Als Bewertungsmaßstab sollte die Prognosequalität der professionellen Kapitalmarktanalysten dienen. Nach Abschluss des Planspiels wird deutlich, dass die Prognosen der Teilnehmer einem vergleichbaren Qualitätsniveau entsprechen (Abschnitt 4.1). Somit schneiden Sie im Sinne der Benotung hervorragend ab, da man von Anfängern kaum eine bessere Qualität als von Experten erwarten kann. Eine individuell differenzierte Benotung erfolgte dann durch die Erstellung von Hausarbeiten im Themenbereich der fundamentalen Aktienanalyse.

- 5-6 Seiten an fundamentalen Informationen (Presseartikel + Unternehmenskennzahlen, Anhang C).
- Einen Fragebogen, auf dem die Prognosen für den Aktienkurs des betrachteten Unternehmens in 3 sowie in 12 Spielrunden (=Monaten) eingetragen werden sollen (siehe Anhang B).

Der Aktienkurs und sämtliche Informationen basieren auf realen Gegebenheiten. Dies ist den Teilnehmern bekannt. Sämtliche Informationen werden jedoch soweit verfremdet, dass eine Bestimmung des Zeitraums und des Unternehmens nicht möglich ist. Die ausführliche Erläuterung der Informationsbasis erfolgt in Abschnitt 3.2.

Der zu prognostizierende Aktienkurs basiert auf dem Verlauf der Volkswagen-Aktie in der zweiten Hälfte der 1980er Jahre. Der Zeitraum und die gewählte Aktiengesellschaft erscheinen aus mehreren Gründen sinnvoll:

- Die Größe und Bekanntheit des Volkswagen-Konzerns sorgten für eine ausreichende Zahl an Presseartikeln, die in jeder Spielrunde des Experiments zur Verfügung gestellt werden.
- Der Zeitraum liegt vor der umfassenden Digitalisierung und Bereitstellung öffentlicher Informationen durch das Internet. Durch die Eingabe einzelner Fragmente der ausgegebenen Presseartikel in die gängigen Suchmaschinen lassen sich keine Rückschlüsse auf den Untersuchungszeitraum bzw. die betrachtete Aktiengesellschaft ziehen.
- Die Aktie wird regelmäßig bei hoher Liquidität gehandelt und weist nennenswerte Kursänderungen im Untersuchungszeitraum auf (siehe Abbildung 1).

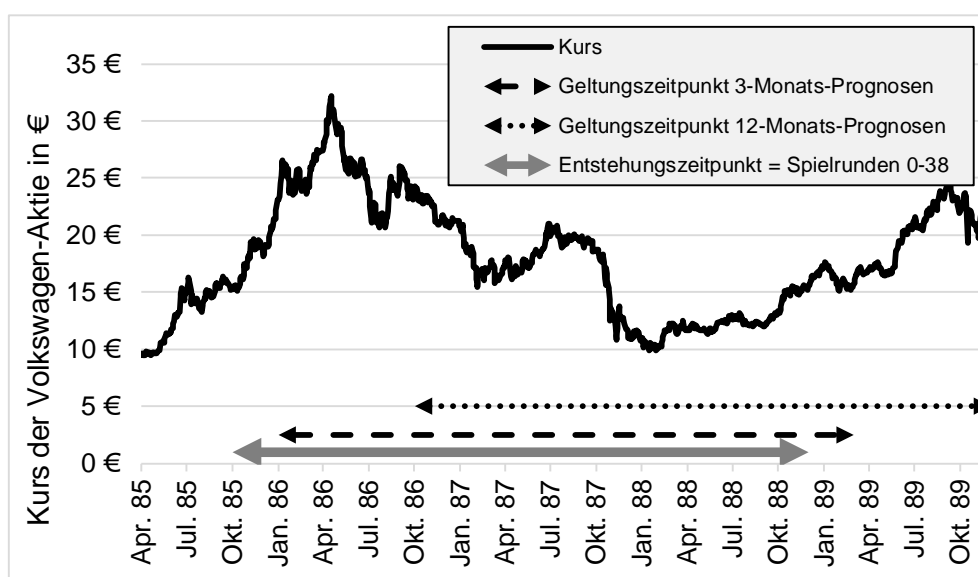


Abbildung 1: Kursverlauf der Volkswagen-Aktie (Euro) im Untersuchungszeitraum basierend auf Daten von Thomson Reuters Datastream

Jede Spielrunde endet mit der gleichzeitigen Abgabe der ausgefüllten Fragebögen durch alle Gruppen. Erhoben werden jeweils vier Informationen: Prognostizierter Aktienkurs in 3 Spielrunden, prognostizierter Aktienkurs in 12 Spielrunden und die jeweilige Sicherheit der Gruppe hinsichtlich der Prognoseabgabe auf eine Likert-Skala von 1 (sehr unsicher) bis 6 (sehr sicher). Die Zeitvorgabe pro Spielrunde muss dabei nicht konstant sein und kann je nach Umfang der Fundamentalinformationen und Erfahrung der Spieler variieren. Die zur Verfügung stehende Zeit wird jeweils am Anfang der Spielrunde kommuniziert.

3.2 Informationsbasis

Die Tätigkeit von Finanzmarktanalysten besteht darin, aus der Vielzahl der zur Verfügung stehenden Informationen die für die zukünftigen Werte relevanten zu identifizieren und die richtigen Schlüsse zu ziehen. Die zwei großen Strömungen bei der Analyse speziell von Aktienkursen lassen sich in Technische Analyse und Fundamentalanalyse aufteilen. Die Planspielteilnehmer erhalten Informationen für beide Analyseformen. Es ist den Studierenden also möglich, ihre Einschätzungen im Rahmen einer der beiden Analyseformen vorzunehmen oder beide zu verbinden. Für die Chartanalyse steht den Teilnehmern in jeder Runde ein Linienchart mit den letzten 6 Spielrunden zur Verfügung (Beispiel siehe Abbildung 2, alle Charts siehe Anhang B – Kursinformationen und Fragebögen).

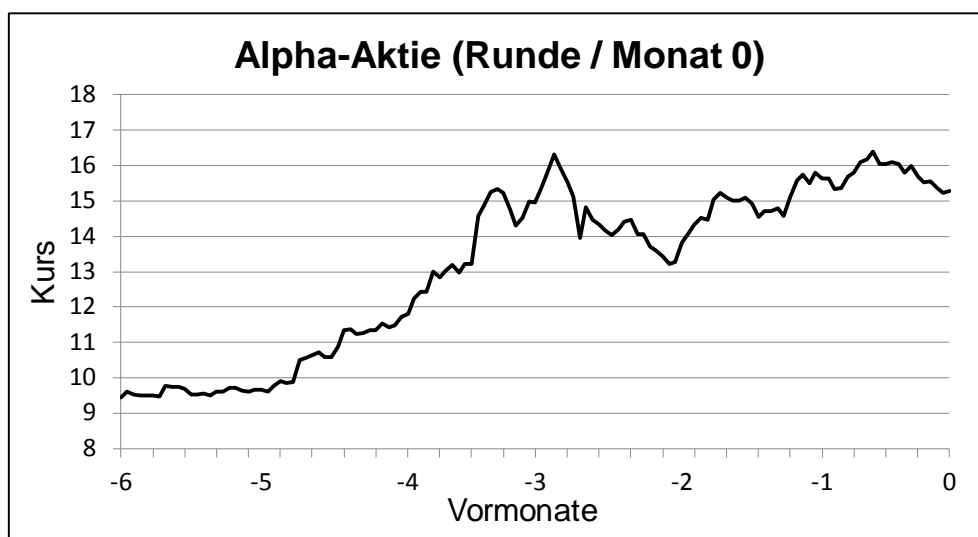


Abbildung 2: Kursinformation für die Planspielteilnehmer am Beispiel der ersten Spielrunde

Die Darreichungsform der Informationen kann die Ergebnisse eines ökonomischen Experiments beeinflussen. Dies ist bei der Kursinformation besonders zu berücksichtigen. Die Teilnehmer werden daher darüber informiert, dass die Skalierung keine Rückschlüsse auf den künftigen Kursverlauf erlaubt. Der Linienchart ist in allen 39 Spielrunden nach folgenden Regeln erstellt worden, wobei dies eine unbewusste Beeinflussung nicht endgültig ausschließt:

- Es werden die letzten 6 Spielrunden / Monate angezeigt.
- Die horizontalen Hilfslinien haben den Abstand von 1 Euro Kurswert.
- Das Kursmaximum und das Kursminimum sind immer mindestens 1 Euro und maximal 2 Euro vom oberen und unteren Ende des Liniencharts entfernt.

Die vielleicht wichtigste Information lässt sich ebenfalls aus dem Linienchart entnehmen: Die Planspielteilnehmer sollen den Kurswert – und nicht bspw. die Änderungsrate – in 3 und 12 Monaten / Spielrunden prognostizieren. Somit ist die Information, von welchem aktuellen Kurswert aus die Prognose abgegeben wird, von besonderer Relevanz.

Deutlich umfangreicher ist das zweite Informationspaket, bei denen es sich um einen historischen Pressespiegel handelt. Die Zeitungsartikel jeder Spielrunde entstammen immer dem Monat vor dem letzten bekannten Aktienkurs. Im Startmonat 0 ist der letzte bekannte Wert des Liniencharts in Abbildung 1 der Kurs der VW-Aktie vom 1.10.1985, womit die Zeitungsartikel allesamt dem September 1985 entstammen. Sie sind in verschiedenen Zeitungen und Zeitschriften erschienen und lassen sich in die Kategorien „Unternehmen Volkswagen“, „Konjunktur Inland“ und „Weltwirtschaft“ einteilen. Die Auswahl der Artikel ist nach den Kriterien

- Abdeckung der drei Kategorien
- Erfassung der wichtigsten Geschehnisse während des der Prognose vorausgehenden Monats
- Vermeidung von Redundanz
- und einer möglichst hohen Kompaktheit erfolgt.

Sämtliche Presseartikel wurden von den Spielleitern überarbeitet, um mögliche Rückschlüsse auf den Zeitraum und das Unternehmen zu verhindern. So wurden sämtliche Zeitangaben durch eine fortlaufende Zeitachse ersetzt. Jeder weitere gespielte damalige Monat entsprach genau einer weiteren Runde des Planspiels. Im weiteren Verlauf dieser Studie ist demnach von 3- bzw. 12-Monats-Prognosen die Rede, wobei sich die Begriffe 3- bzw. 12-Runden-Prognosen synonym verwenden lassen. Das Unternehmen Volkswagen wurde in „Alpha“ umbenannt. Außer

den USA und Deutschland wurden alle anderen Staatennamen durch „Wunderland“ ersetzt, um Rückschlüsse auf den Zeitraum und das Unternehmen zu erschweren. Ebenfalls maskiert wurden die Tochterunternehmen durch Begriffe wie „Gamma“ oder „Delta“. Generell gilt, dass alle Wortumwandlungen über das Planspiel hinweg einheitlich verwendet worden sind.

Der Umfang der Texte beläuft sich auf 4-6 Seiten pro Spielrunde. Alle zwölf Spielrunden haben die Gruppen zusätzlich Unternehmenskennzahlen erhalten.

3.3 Methodische Einordnung

Das Untersuchungsdesign lässt sich weder eindeutig als Labor- noch als Feldexperiment einordnen. Eine wiederkehrende Kritik an den üblichen Laborexperimenten ist die Beschränkung der Versuchspersonen auf Studierende. Daraus hat sich die Kategorie der „artefactual field experiments“ entwickelt, bei der je nach Fragestellung die Probanden aus ihren gewohnten Handlungsfeldern rekrutiert werden, um die externe Validität der Ergebnisse zu erhöhen.²⁵ Das hier vorgestellte Design geht untypischerweise den umgekehrten Weg.

Gegen eine Kategorisierung des Planspiels als Laborexperiment spricht die mangelnde Kontrolle aller maßgeblichen Einflussfaktoren. So ist eine mögliche Kommunikation zwischen den verschiedenen Gruppen während des Spiels nicht explizit verboten. Auch können die Räume verlassen sowie Smartphones und Notebooks benutzt werden. Zwischen den wöchentlich stattfindenden Veranstaltungsterminen könnte es zu zahlreichen spielrelevanten Handlungen der Teilnehmer kommen (Aneignung von zusätzlichem Fachwissen, Absprachen zwischen den Gruppen etc.), ohne dass die Spielleiter davon erfahren würden. Weiterhin spricht gegen die Einordnung als Laborexperiment die Unkenntnis der Planspielteilnehmer darüber, dass ihre Prognosen zu wissenschaftlichen Zwecken ausgewertet werden.²⁶ Die Durchführung des Planspiels in gewöhnlichen Seminarräumen über ein ganzes Semester hinweg ist darüber hinaus nicht mit einmaligen Terminen in einem Experimentallabor vergleichbar.

²⁵ Der Begriff ist von Harrison und List (2004) eingeführt worden. Zu einer ausführlichen Diskussion der verschiedenen Formen ökonomischer (Feld-)Experimente und deren Vorteile vgl. auch List (2011).

²⁶ Unter dem Eindruck der Vorgänge während der Zeit der nationalsozialistischen Regierung wurde 1947 der Nürnberger Kodex verabschiedet, der eine freiwillige Zustimmung der Versuchspersonen als zwingende Voraussetzung für wissenschaftliche Experimente mit Menschen zum Inhalt hat. Levitt und List (2009) sehen dies bei ökonomischen Experimenten nicht als zwingende Voraussetzung an, insbesondere wenn die Experimente einer ethischen Überprüfung unterzogen werden und den Probanden kein Schaden entsteht. Bei dem hier durchgeführten Planspiel wurden die Studierenden zum Abschluss der Veranstaltung über das spezifische Erkenntnisinteresse der Dozenten informiert. Darüber hinaus wurde eine fachliche Einordnung sowie eine gemeinsame Diskussion über die Forschungsfrage durchgeführt.

Eine Einordnung als Feldexperiment kommt jedoch auch nicht in Frage. Die Probanden werden mit einer neuen Fragestellung konfrontiert und können somit nicht auf Erfahrungswissen zurückgreifen. Zudem werden Kapitalmarktprognosen für gewöhnlich in einem völlig anderen institutionellen Rahmen erstellt. Die Prognoseabgabe wird auch in der Praxis auf Basis einer Gruppendiskussion getroffen, jedoch wird diese auf eine erheblich umfangreichere Informationsbasis und mathematische Schätzverfahren aufbauen. Das Anreizsystem im Planspiel ist zudem nicht vergleichbar mit den Anreizen, denen Finanzanalysten ausgesetzt sind.

Durch die festen Zeitvorgaben, die Kontrolle über den Ablauf und die Studierenden als Probanden überwiegt der Laborcharakter, eine Mischung mit den methodischen Elementen der Feldexperimente ist jedoch durchaus vorhanden. Eine Zusammenfassung der wesentlichen Merkmale des Experiments ist Tabelle 1 zu entnehmen.

Tabelle 1: Übersicht über die Ausgestaltung des Experiment differenziert nach den Merkmalen für Feldexperimente nach Carpenter u. a. (2005)

Merkmal	Ausgestaltung
Subject pool	Studierende, zufällige Gruppenzuteilung
Information and experience	Kaum Erfahrung hinsichtlich der Analyse von Aktienkursen, homogener Kenntnisstand, Information über Planspiel – nicht Experiment!
Commodity	Prognosen: Genauigkeit, Erwartungsbildung, Gegenwartsbezug, Unsicherheit
Task and institutional rules	Aktienkursprognosen einer empirischen Zeitreihe auf der Grundlage vergangener Kurse und Fundamentaldaten, Ablauf in Spielrunden bei situativ angepasster Zeitvorgabe, keine festen Regeln für Kommunikation
Stakes	Credits für Teilnahme, Motivation durch Wettbewerb, Note in Abhängigkeit vom Prognoseerfolg
Environment	Durchführung in Seminarräumen, Übergabe der Informationen und Abgabe in Papierform

3.4 Ablauf

Das Untersuchungsdesign wurde im Rahmen eines Wahlpflichtfachs an der Ostfalia am Campus Wolfsburg mit dem Titel „Projekt: Aktienanalyse“ im Sommersemester 2012 angewendet. Teilgenommen haben Studierende des Bachelorstudiengangs Betriebswirtschaftslehre im 2. Semester. Sie verfügen zu diesem Zeitpunkt kaum über Kenntnisse der Aktienanalyse. Zudem sollte der Kenntnisstand durch die Beschränkung auf das 2. Semester weitgehend homogen gewesen sein.

Das Planspiel fand über das gesamte Semester hinweg wöchentlich statt. Eine Übersicht über den Verlauf ist Tabelle 2 zu entnehmen. Bei 9 Veranstaltungen wurden Planspielrunden gespielt, wobei die erste Veranstaltung vornehmlich der Erläuterung der Grundlagen der fundamentalen Aktienanalyse und der Spielregeln vorbehalten war. Am Abschluss dieser Veranstaltung wurde die erste Spielrunde gespielt.

Tabelle 2: Übersicht über die Planspieltermine

Veranstaltung	Datum	Teilnehmer	Gruppen	Ø pro Gruppe	Spielrunden
1	08.03.2012	37	10	3,70	1
2	15.03.2012	33	9	3,67	4
3	29.03.2012	30	10	3,00	4
4	12.04.2012	31	10	3,10	5
5	19.04.2012	28	9	3,11	5
6	26.04.2012	29	10	2,90	5
7	03.05.2012	26	10	2,60	5
8	10.05.2012	28	9	3,11	5
9	24.05.2012	22	9	2,44	5

Ø = Arithmetischer Mittelwert

Bemerkenswert ist der Rückgang der Teilnehmerzahl von 37 in der ersten Veranstaltung auf 22 in der letzten Veranstaltung (siehe auch Abbildung 3). Der leichte Rückgang im Semesterverlauf ist nicht ungewöhnlich und auch bei anderen Hochschulveranstaltungen üblich. Planspielbezogene Ursachen dürften hierfür nicht in Frage kommen, wenngleich sie auch nicht gänzlich auszuschließen sind. Der starke Rückgang in der letzten Veranstaltung lässt sich nach Rücksprache mit den anwesenden Teilnehmern auf den späten Zeitpunkt im Semester zurückführen. Viele Studierende befinden sich zu diesem Zeitpunkt schon mitten in der Klausurvorbereitung und waren scheinbar nicht mehr bereit, Zeit in das Wahlpflichtfach zu investieren.

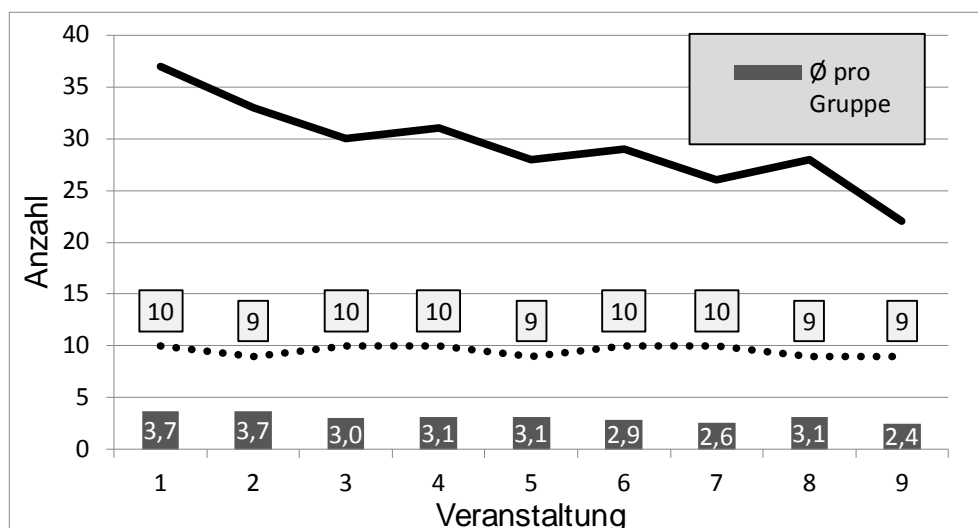


Abbildung 3: Verlauf der Teilnehmer- und Gruppennzahlen

Die Zahl der Gruppen blieb einigermäßen konstant. Die Zahl der Teilnehmer pro Gruppe war rückläufig. Da pro Runde recht umfangreiche Unternehmens- und Konjunkturinformationen

ausgewertet werden konnten, durften Gruppen mit nur noch einem Teilnehmer mit einer anderen Kleingruppe (1-2 Teilnehmer) kooperieren. Sie durften die gleichen Prognosen wie die Kooperationsgruppe abgeben, konnten aber auch eine eigene Einschätzung auf Basis der gemeinsamen Informationsauswertung treffen. Tabelle 3 informiert über die Zusammenarbeit der Gruppen und die gleichzeitige Häufigkeit deckungsgleicher Angaben auf den Fragebögen.

Tabelle 3: Übersicht über die Zusammenarbeit von Gruppen

Veranstaltung	Runden	Betr. Gruppe	Zusammenarbeit	Übereinstimmung bei Antworten der kollaborierenden Gruppen in %			
				3-M-P	12-M-P	3-M-S	12-M-S
4	5	3 und 6	3 und 6 (jeweils alleine) haben zusammengearbeitet	40%	20%	40%	80%
5	5	6	Zusammenarbeit mit 9 (2 Teilnehmer)	20%	0%	40%	0%
6	5	8	Zusammenarbeit mit 6 (2 Teilnehmer)	0%	40%	20%	0%
9	5	1 und 11	Wollten nicht zusammenarbeiten	-	-	-	-
Arithmetisches Mittel				30%	30%	33,3%	26,7%

3-M-P: 3-Monats-Prognosen; 12-M-P: 12-Monats-Prognosen; 3-M-S: Sicherheit bei der 3-Monats-Prognose; 12-M-S: Sicherheit bei der 12-Monats-Prognose

In insgesamt 6 Fällen bestand eine Gruppe an einem Spieltag nur aus einem Mitglied. In einem Fall haben zwei solcher Minigruppen kollaboriert. Bei der 9. Veranstaltung wollten die Mitglieder die Ihnen gereichten Informationen alleine auswerten und schlugen das Angebot der möglichen Zusammenarbeit aus. Bei zwei Veranstaltungen kam es zur Zusammenarbeit einer Minigruppe mit einer ebenfalls kleinen Gruppe (je 2 Mitglieder). In keinem der Fälle kam es zu völlig deckungsgleichen Prognosen oder zu übereinstimmenden Angaben zur dabei empfundenen Sicherheit. Unter Berücksichtigung der Tatsache, dass insgesamt 370 Fragebögen abgegeben worden sind, erscheint es daher nicht wahrscheinlich, dass die durch dieses Vorgehen betroffenen 30 Fragebögen (8,1%) zu einer nennenswerten Beeinflussung des Untersuchungsergebnisses führen konnten.

Die Zeitvorgabe pro Spielrunde ist der Abbildung 4 zu entnehmen und richtete sich vor allem nach dem Umfang der Presseartikel. Da die Teilnehmer mit der Zeit scheinbar in der Lage waren, die Prognosen schneller zu erstellen, ließ sich die Zeitvorgabe in den späteren Spielrunden etwas knapper gestalten.

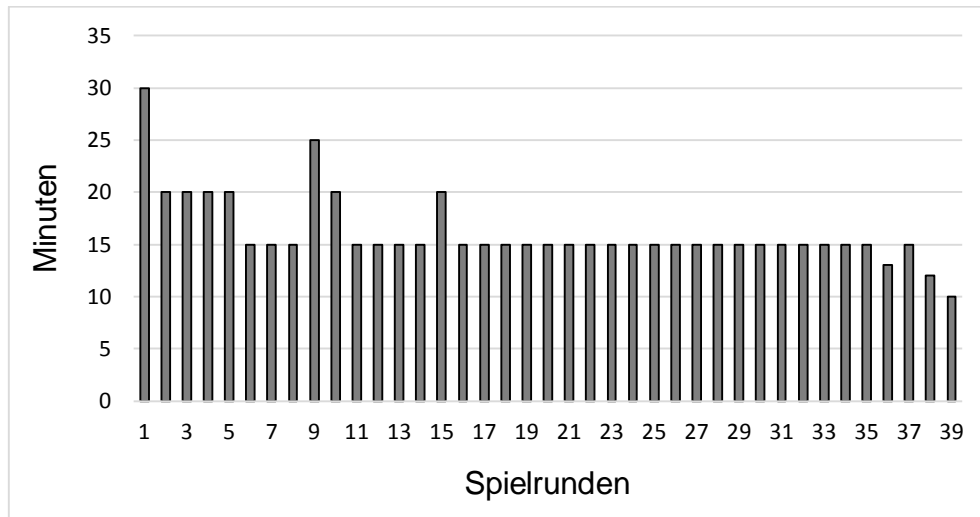


Abbildung 4: Zeitvorgabe nach Spielrunden

Da nicht von allen Gruppen zu jeder Veranstaltung mindestens ein Teilnehmer anwesend war, variiert die Zahl der abgegebenen Prognosen je nach Gruppe. Bei den meisten Gruppen erfolgte jedoch über die gesamte Laufzeit des Planspiels die kontinuierliche Abgabe von Prognosen (Tabelle 4).

Tabelle 4: Teilnahme der Gruppen nach Spielrunden

Veranstaltung	Spielrunden	1	2	3	4	5	6	7	8	9	10
1	0	x	x	x	x	x	x	x	x	x	
2	1-4	x	x		x	x	x	x	x	x	x
3	5-8	x	x	x	x	x	x	x	x	x	x
4	9-13	x	x	x	x	x	x	x	x	x	x
5	14-18	x	x	x	x	x	x	x	x		x
6	19-23	x	x	x	x	x	x	x	x	x	x
7	24-28	x	x	x	x	x	x	x	x	x	x
8	29-33	x	x	x	x	x	x	x	x		x
9	34-38	x	x	x	x	x	x	x	x		x
Anzahl Fragebögen	-	39	39	35	39	39	39	39	39	24	38

7 Gruppen haben alle Prognosen abgegeben. Die wenigsten Prognosen kamen von Gruppe 9. Da die Gruppen die Fragebögen immer vollständig ausgefüllt haben ist die Zahl der pro Gruppe erfassten Fragebögen mit der Zahl der 3- und 12-Runden-Prognosen und der angegebenen Unsicherheiten in der weiteren Analyse identisch.

3.5 Zusammenfassung

Die in Kapitel 2 vorgestellten Untersuchungshypothesen haben größtenteils gemein, dass Phänomene, die sich bei Kapitalmarktprognosen von Finanzanalysten beobachten lassen, auf ihre Replizierbarkeit hin überprüft werden sollen. Es könnte durchaus sein, dass spezifische Institutionen im beruflichen Umfeld der Analysten für das Auftreten der Phänomene verantwortlich sind, wobei insbesondere das vorherrschende Anreizsystem in Frage kommt. Das Design des Planspiels hat zum Ziel, ein Vergleich zu Kapitalmarktprognosen in einem anderen Umfeld zu ermöglichen. Hierzu nehmen Studierende an einem Planspiel zur Kapitalmarktprognose teil. Es läuft über ein ganzes Semester bei maximal 5 Spielrunden pro Veranstaltung. Sie prognostizieren in Gruppen den empirischen Aktienkurs der Alpha-Aktie (Volkswagen AG Ende der 80er Jahre) mit Hilfe von vergangenen Kursdaten und Presseartikeln. Das Informationsmaterial wird so überarbeitet, dass ein Rückschluss auf das Unternehmen und den Zeitraum ausgeschlossen ist.

Zusammengefasst ermöglicht das Untersuchungsdesign eine allgemeine Institutionenanalyse. Stellen sich bei den Prognosen der Planspielteilnehmer die gleichen Phänomene wie bei den echten Analysten ein, so scheinen die analystenspezifischen Institutionen, die nicht im Planspiel abgebildet werden, hierfür keine notwendige Voraussetzung zu sein. Fallen die Ergebnisse des Planspiels anders als in der Realität aus, so könnten im beruflichen Umfeld der Analysten spezifische Institutionen wirken, die durch eine Beobachtung im Feld oder eine Isolierung im Labor näher untersucht werden sollten.

4 Auswertung

In diesem Kapitel werden die mittels des Planspiels gewonnenen Prognosedaten ausgewertet. Tabelle 5 gibt einen Überblick über die im 2. Kapitel aufgestellten Hypothesen und den Abschnitt, in dem die jeweilige Überprüfung stattfindet.

Tabelle 5: Überblick über die Struktur der Auswertung

Hypothese	Abschnitt
<i>H1: Die Prognosen der Teilnehmer schneiden schlechter als empirische Kapitalmarktprognosen von Experten ab.</i>	4.1
<i>H2: Die Prognosefehler weisen systematische Komponenten auf und entsprechen somit nicht den Kriterien der rationalen Erwartungsbildung.</i>	4.2
<i>H3: Die Prognosen weisen gegenwartsorientierte Verlaufsanpassung auf.</i>	4.3
<i>H4: Die subjektiv empfundene Unsicherheit der Gruppen stimmt mit der anhand des Kriteriums Prognosedivergenz objektiv messbaren Unsicherheit überein.</i>	4.4

Im Abschnitt 4.5 erfolgt eine Zusammenfassung der Auswertung.

4.1 Qualität

Die im Abschnitt 2.1 hergeleitete erste Hypothese lautet:

H1: Die Prognosen der Teilnehmer schneiden schlechter als empirische Kapitalmarktprognosen von Experten ab.

Für den Nutzer von Kapitalmarktprognosen ist die Qualität sicherlich die wichtigste Eigenschaft der Prognosezeitreihen. Um von einer guten Qualität sprechen zu können, müssen die Prognosen einen zusätzlichen Wert gegenüber einem Verzicht ihrer Nutzung aufweisen.

Zwingende Voraussetzung ist zunächst eine Stationarität gleicher Ordnung zwischen den Prognosezeitreihen und den tatsächlichen Aktienkursen. Diese wird im Abschnitt 4.1.1 untersucht. Anschließend erfolgt eine differenzierte Betrachtung der Prognosefehler. Dabei wird untersucht, ob die Gruppen eher zu hoch oder zu niedrig prognostiziert haben, wie hoch die absoluten Fehler sind und ob die Prognosen besser oder schlechter als die naive Prognose abschneiden (Abschnitt 4.1.2). Dies ist für die Frage entscheidend, ob sich ein positiver Nutzen für den Nachfrager der Prognose geben kann, da ihm die naive Prognose jederzeit ohne Aufwand als Alternative zur Verfügung steht. In Abschnitt 4.1.3 werden die Resultate mit den Ergebnissen empirischer Studien verglichen. In 4.1.4 schließlich erfolgt eine Einschätzung, ob die Anfänger in dieser Studie wie im Rahmen der ersten Hypothese formuliert schlechter als die Experten abschneiden.

4.1.1 Stationarität

Zeitreihen an den Kapitalmärkten sind in der Regel nicht stationär, d. h. sie unterliegen einem Trend. Die entsprechende Überprüfung des Aktienkurses der VW AG im Prognosezeitraum sowie der Prognosezeitreihen der Planspielgruppen ist aus folgenden Gründen geboten:

1. Für nicht-stationäre Zeitreihen ist eine Transformation in stationäre Zeitreihen häufig unumgänglich, da diese Gegenstand weiterer ökonomischer Untersuchungen sein sollen. Hierfür sind zeitkonstante Eigenschaften der Zeitreihen eine notwendige Bedingung, um die Identifikation von vermeintlichen Zusammenhängen ausschließen zu können, die ausschließlich auf der Nicht-Stationarität der Zeitreihen basieren.
2. Im Hinblick auf die Prognosequalität ist ein Vergleich der Stationarität zwischen den Prognosezeitreihen und den Aktienkursen aufschlussreich, da ein unterschiedlicher Grad der Stationarität schon ein Ausschlusskriterium für gute Prognosen darstellen dürfte.²⁷

Getestet wird das Vorliegen von Stationarität mittels des Augmented Dickey-Fuller-Tests (ADF-Test).²⁸ Die zugrundeliegende Nullhypothese lautet: Die Zeitreihe verfügt über eine Einheitswurzel und ist somit nicht-stationär. Die in Tabelle 6 dargestellten p-Werte drücken die Irrtumswahrscheinlichkeit für den Fall aus, dass diese Nullhypothese zu Unrecht abgelehnt wird.

Tabelle 6: Ergebnisse des ADF-Tests auf Stationarität der Zeitreihen

		Prognosezeitreihen der einzelnen Gruppen – p-Werte*									
PH	Kurs	1	2	3	4	5	6	7	8	9	10
3	0,638	0,568	0,524	0,376	0,676	0,599	0,450	0,453	0,902	0,568	0,571
12	0,578	0,403	0,536	0,073	0,751	0,782	0,338	0,666	0,400	0,646	0,643

*p-Wert = Irrtumswahrscheinlichkeit für die fälschliche Ablehnung der Hypothese, dass die Zeitreihen stationär sind.

Bis auf einen Fall handelt es sich mit hoher Wahrscheinlichkeit nicht um stationäre Zeitreihen. Diese lassen sich durch die Transformation in Zeitreihen mit ihren ersten Differenzen möglicherweise in stationäre Zeitreihen umformen. Die Transformation funktioniert folgendermaßen:

²⁷ Cheung und Chinn (1998) zeigen diese Möglichkeit für die Prognosen von Wechselkursen und Schwarzbach u. a. (2012) für Renditeprognosen von Staatsanleihen auf.

²⁸ Dickey und Fuller (1979).

$$(1) \quad \Delta P_t = P_t - P_{t-1} \quad \text{und} \quad \Delta A_t = A_t - A_{t-1}$$

P ist dabei die jeweilige Prognose und A der Aktienkurs zum Zeitpunkt t .

Nach dieser Transformation der Zeitreihen in ihre ersten Differenzen wird der Test erneut durchgeführt. Danach sind sowohl alle Prognosezeitreihen als auch die tatsächlichen eingetretenen Aktienkurse als stationär anzusehen (1%-Signifikanzniveau). Die Stationarität gleicher Ordnung als Voraussetzung für eine gute Prognosequalität ist gegeben.

Ursächlich für die Stationarität gleicher Ordnung ist aber nicht zwingend die Antizipation des künftigen Kursniveaus durch die Planspielteilnehmer. Es ist ebenso möglich, dass die Prognosen dem jeweiligen Aktienkurs zum Zeitpunkt der Prognoseerstellung folgen, wie es bei Kapitalmarktprognosen von Experten in der Regel der Fall ist (siehe Abschnitt 4.3). In diesem Fall ist es ebenfalls zu erwarten, dass die Prognosezeitreihen die Stationaritätseigenschaften der prognostizierten Zeitreihe annehmen.

4.1.2 Höhe der Prognosefehler

In Abbildung 5 werden die 3-Monats-Prognosen der 10 Gruppen (graue Linien) und der tatsächliche Aktienkurs (schwarze Linie) dargestellt. Da eine Spielrunde genau einem Monat des Kursverlaufs entspricht, erhält bspw. eine 3-Monats-Prognose aus Spielrunde 10 ihre Gültigkeit in Runde 13.

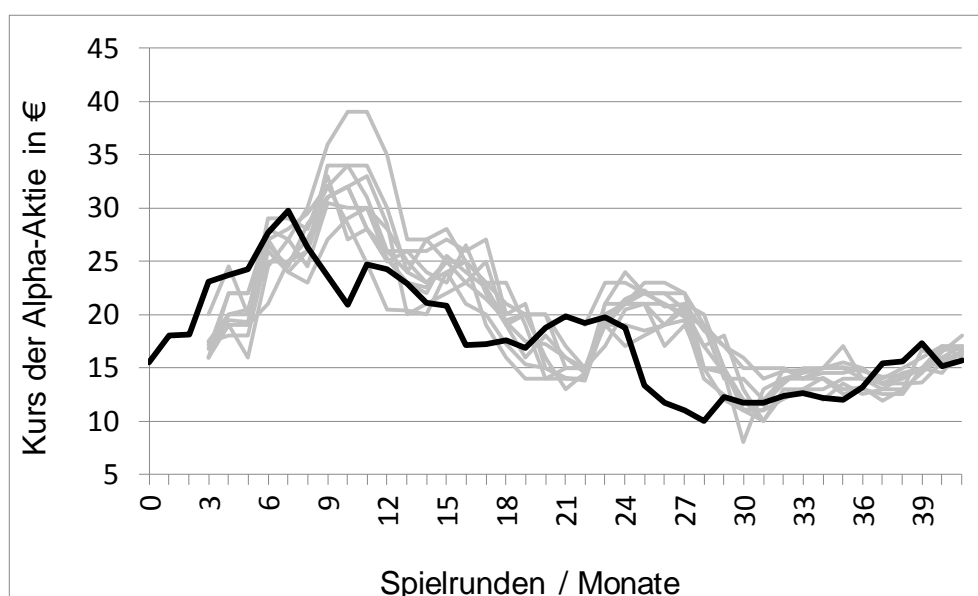


Abbildung 5: 3-Monats-Prognosen zum Geltungszeitpunkt

Die grafische Auswertung deutet darauf hin, dass die Prognosen wesentliche Wendepunkte des Kursverlaufs nicht erfasst haben. Der Höhepunkt des Aktienkurses in den Runde 6 und 7 sagen die Planspielteilnehmer erst für einige Runden später vorher. Der starke Kurssturz nach Runde 24 ist ebenfalls nicht vorhergesagt worden. Der zeitliche Versatz zwischen Kursverlauf und Prognose in Höhe des Prognosehorizontes deutet auf einen starken Gegenwartsbezug der Prognosen hin wie er auch bei Kapitalmarktprognosen der Experten regelmäßig gemessen wird. Genauere Ausführungen hierzu folgen in Abschnitt 4.3.

Ein etwas anderes Bild zeigt sich für die 12-Monats-Prognosen (Abbildung 6).

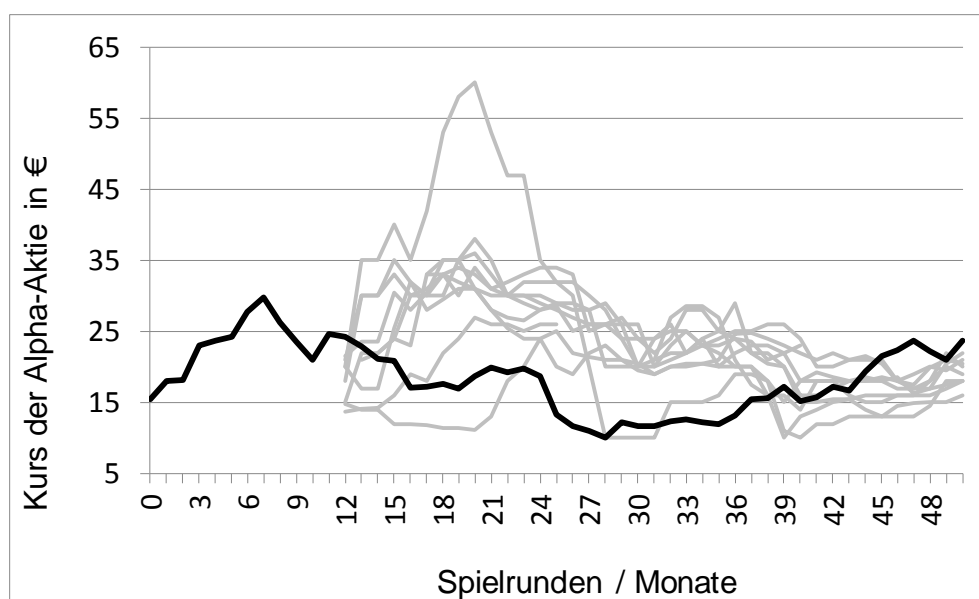


Abbildung 6: 12-Monats-Prognosen zum Geltungszeitpunkt

Auffällig ist die deutliche Überschätzung des Kursverlaufs, die bis auf den Anfang und das letzte Viertel des Planspiels außerordentlich ausgeprägt zu sein scheint. Eine Gruppe rechnete zwischenzeitig sogar mit einer Verdoppelung des Kursniveaus. Besonders für die erste Hälfte des Planspiels zeigt sich zudem eine deutlich höhere Abweichung der Prognosen untereinander, als dies bei den 3-Monats-Prognosen der Fall ist. Beide Phänomene zeigen sich auch rechnerisch, wie es im weiteren Verlauf dieses Abschnitts deutlich gemacht wird.

Die Frage, ob es zu einer Über- oder Unterschätzung des Kursverlaufs gekommen ist, lässt sich durch Berechnung des Mittleren Prognosefehlers (MP) beantworten. Sei P die Prognose und A der tatsächlich eingetretene Wert zum jeweiligen Gültigkeitszeitpunkt t . Die aufsummierten Über- und Unterschätzungen werden zudem mit der Gesamtzahl der Beobachtungen T gewichtet. Der MP berechnet sich wie folgt:

$$(2) \quad MP = \frac{1}{T} \sum_{t=1}^T P_t - A_t$$

Mit dem mittleren absoluten Prognosefehler (*MAP*) wird darüber hinaus die Frage beantwortet, wie weit die Prognosen im Durchschnitt daneben lagen, ohne dass sich Über- oder Unterschätzungen gegenseitig aufheben. Der *MAP* berechnet sich wie folgt:

$$(3) \quad MAP = \frac{1}{T} \sum_{t=1}^T |P_t - A_t|$$

Die Ergebnisse der 3- und 12-Monats-Prognosen aller Gruppen in Bezug auf die vorgestellten einfachen Prognosegütemaße sind Tabelle 7 zu entnehmen.

Tabelle 7: Mittlerer und absoluter mittlerer Prognosefehler

Maß / Gruppe	1	2	3	4	5	6	7	8	9	10	Ø	σ
MP - 3 Mon. (€)	1,01	1,22	1,31	2,04	1,24	1,50	1,84	2,31	2,63	2,44	1,75	0,55
MP - 12 Mon. (€)	5,48	1,78	4,91	7,55	6,43	6,48	5,89	4,76	8,33	8,56	6,02	1,89
MAP - 3 Mon. (€)	3,23	3,63	2,78	4,34	3,37	3,72	3,99	4,83	4,14	4,52	3,85	0,60
MAP - 12 Mon. (€)	7,85	6,36	6,80	11,92	8,76	8,24	9,59	7,34	10,09	9,79	8,67	1,62

Ø = arithmetischer Mittelwert der 10 Gruppen ohne Gewichtung nach der Anzahl Beobachtungen; σ = Standardabweichung der 10 Ergebnisse

Die *MP* aller Gruppen haben ein positives Vorzeichen, was bedeutet, dass es im Mittel von allen Gruppen zu Überschätzungen des prognostizierten Kursverlaufs kam. Diese waren bei den 12-Monats-Prognosen mit einem arithmetischen Mittel von 6,02 € weitaus ausgeprägter als bei den 3-Monats-Prognosen (1,75 €). Dass es auch zu häufigen Unterschätzungen kam, lässt sich an den deutlich höheren *MAP* ablesen. Bei den 3-Monats-Prognosen lagen die Gruppen im Durchschnitt 3,75 € daneben, bei den 12-Monats-Prognosen um 8,67 €

Zwischen den Gruppen sind erhebliche Unterschiede bei der Prognosequalität festzustellen. Der *MAP* der 3-Monats-Prognosen liegt zwischen 2,78 € (Gruppe 3) und 4,83 € (Gruppe 8). Die Standardabweichung beträgt 60 Cent. Auch bei den 12-Monats-Prognosen liegt die Gruppe mit dem höchsten Prognosefehler (Gruppe 4: 11,92 €) im Durchschnitt fast doppelt so weit daneben wie die Gruppe mit dem niedrigsten *MAP* (Gruppe 2: 6,36 €). Bei den 12-Monats-Prognosen liegt die Standardabweichung bei 1,62 €

Um die Ergebnisse des Planspiels vergleichbar zu machen werden relative Prognosegütemaße benötigt. Bei dem mittleren relativen absoluten Prognosefehler (*MRAP*) werden die *MAP* aus (3) durch den jeweiligen tatsächlich eingetretenen Kurs *A* zum Zeitpunkt *t* dividiert.

$$(4) \quad MRAP = \frac{1}{T} \sum_{t=1}^T \frac{|P_t - A_t|}{A_t}$$

Diese Normierung hat den Vorteil, dass die Prognosefehler mit anderen Studien vergleichbar sind. Eine Multiplikation mit 100 führt zur prozentualen Abweichung der Prognosen von den tatsächlich eingetretenen Werten. Ein erheblicher Nachteil des *MRAP* ist die enorme Ungenauigkeit bei Vergleichswerten nahe 0. Dieses Problem kommt bei den hier betrachteten Aktienkursen nicht zum Tragen. Die berechneten *MRAP* sind Tabelle 8 zu entnehmen.

Tabelle 8: Mittlerer relativer absoluter Prognosefehler

Maß / Gruppe	1	2	3	4	5	6	7	8	9	10	Ø	σ
<i>MRAP</i> - 3 Mon.	19%	23%	17%	25%	20%	23%	24%	29%	25%	28%	23%	04%
<i>MRAP</i> - 12 Mon.	52%	43%	46%	68%	58%	57%	63%	47%	68%	67%	57%	09%

Ø = arithmetischer Mittelwert der 10 Gruppen ohne Gewichtung nach der Anzahl Beobachtungen; σ = Standardabweichung der 10 Ergebnisse

Die 3-Monats-Prognosen lagen im Durchschnitt um 23,4 % daneben. Bei den 12-Monats-Prognosen liegt der *MRAP* bei knapp 60%. Dies ist eine ganz erhebliche Abweichung. Sehr große Unterschiede zwischen den Gruppen ergeben sich nicht.

Eine noch aussagekräftigere Möglichkeit ist der Vergleich der *MAP* mit dem Prognosefehler der sogenannten naiven Prognose. Der naiven Prognose liegt die Idee von Laplace (1814) zu Grunde, dass in einer Situation der völligen Ungewissheit ebenso viel für steigende wie für fallende Werte spricht. In dieser Situation lässt sich ebenso gut an dem aktuellen Wert als Prognose festhalten. Eine Prognosezeitreihe, die schlechter als die naive Prognose abschneidet, zeigt in aller Deutlichkeit, dass die Prognoseaufgabe nicht erfolgreich bewältigt werden konnte. Die Qualität der naiven Prognose lässt sich bei Kapitalmarktprognosen sehr einfach erreichen, da die prognostizierten Werte in der Regel tagesaktuell vorliegen.

Um die Qualität einer Prognosezeitreihe in diesem Sinne in einer einzigen Zahl zu veranschaulichen, wird der mittlere relative absolute Prognosefehler bezogen auf die naive Prognose (*MRAPnP*) berechnet. Dabei wird der *MAP* aus (3) durch den Prognosefehler der naiven Prognose dividiert. Dieser berechnet sich aus der Differenz aus dem tatsächlich eingetretenen Wert

A zum Zeitpunkt t und dem Wert zum Zeitpunkt der Prognoseerstellung, der an einer um dem Prognosehorizont h zurückverschobenen Stelle der Zeitreihe ablesbar ist.

$$(5) \quad MRAPnP = \frac{1}{T} \sum_{t=1}^T \frac{|P_t - A_t|}{|A_{t-h} - A_t|}$$

Ein *MRAPnP* Wert von 1 bedeutet, dass die Prognosefehler der Prognosezeitreihe und der naiven Prognose gleich hoch sind. Bei einem Wert von kleiner als 1 hat die Prognosezeitreihe besser, bei einem Wert von größer als 1 schlechter abgeschnitten. Der *MRAPnP* lässt durch seine Normierung somit nicht nur einen Vergleich zwischen verschiedenen Studien zu, er bietet auch selber eine klare Aussage über die Qualität der untersuchten Prognosezeitreihen.

In der gleichen Art und Weise lässt sich der Theilsche Ungleichheitskoeffizient neuer Art (U_2) interpretieren.²⁹ Der wesentliche Unterschied zum *MRAPnP* besteht darin, dass einzelne besonders hohe Prognosefehler stärker ins Gewicht fallen als viele kleine Prognosefehler. Der U_2 berechnet sich wie folgt:

$$(6) \quad U_2 = \frac{\sqrt{\frac{1}{T-h} \sum_{t=1}^T \left(\frac{P_t - A_{t-h}}{A_{t-h}} - \frac{A_t - A_{t-h}}{A_{t-h}} \right)^2}}{\sqrt{\frac{1}{T-h} \sum_{t=1}^T \left(\frac{A_t - A_{t-h}}{A_{t-h}} \right)^2}}$$

Der Prognosefehler der naiven Prognose ist für alle Gruppen mit der gleichen Zahl an Beobachtungen gleich. 7 der 10 Gruppen haben über alle 39 Spielrunden Prognosen abgegeben. Der Prognosefehler der naiven Prognosen für alle Spielrunden liegt für die 3-Monats-Prognosen bei 3,06 € und für die 12-Monats-Prognose bei 6,58 €. Die berechneten *MRAPnP* und U_2 sind Tabelle 9 zu entnehmen.

Tabelle 9: Prognosefehler bezogen auf die naive Prognose

Maß / Gruppe	1	2	3	4	5	6	7	8	9	10	Ø	Md.
MRAPnP - 3 Mon.	1,06	1,19	0,91	1,42	1,10	1,22	1,31	1,58	1,35	1,48	1,26	1,26
MRAPnP - 12 Mon.	1,19	0,97	1,03	1,81	1,33	1,25	1,46	1,11	1,53	1,49	1,32	1,29
U_2 - 3-Mon.	1,03	1,24	1,00	1,36	1,04	1,21	1,26	1,25	1,26	1,51	1,22	1,25
U_2 - 12- Mon.	1,07	0,97	0,88	1,66	1,17	1,14	1,27	1,32	1,33	1,37	1,22	1,22

Ø = arithmetischer Mittelwert der 10 Gruppen ohne Gewichtung nach der Anzahl Beobachtungen; Md. = Median

²⁹ Vgl. Theil (1961), S. 32. Die ältere Variante stellt einen Koeffizienten aus dem *RMSE* der Prognosen und dem *RMSE* der naiven Prognosen da und führt zu Ergebnissen analog dem *MRAPnP*.

Sowohl bei den 3-Monats-Prognose (Gruppe 3: $MRAPnP$ 0,91) als auch bei den 12-Monats-Prognosen (Gruppe 2: $MRAPnP$ 0,97) ist es jeweils nur einer Gruppe gelungen, eine besser Prognosezeitreihe als die naive Prognose zu generieren. Da es jeweils eine andere Gruppe ist und der Wert zudem immer noch nahe bei 1 liegt, liegt die Vermutung nahe, dass die beiden Gruppen nur zufällig besser abgeschnitten haben. Die mittels des U_2 gemessenen Werte sind im Mittel etwas niedriger. Bei den 3-Monats-Prognosen findet sich keine Gruppe mit einem Wert <1 wieder, bei den 12-Monats-Prognosen immerhin 2. Dabei bestätigt sich der Wert von Gruppe 2 mit 0,97. Gruppe 3 weist einen U_2 von 0,88 auf, gegenüber dem Wert von 1,03 der mittels des $MRAPnP$ gemessen wurde. Stabil ist das vermeintlich bessere Ergebnis also nicht.

Die Durchschnittswerte aller Gruppen liegen sehr deutlich über 1, unabhängig vom verwendetet Prognosegütemaß. Insgesamt lässt sich also sagen, dass die Prognoseaufgabe nicht erfolgreich bewältigt werden konnte. Hätten die Gruppen einfach den ihnen bekannten Gegenwartswert als Prognose eingesetzt, wäre das Ergebnis für sie in den meisten Fällen besser ausgefallen.

4.1.3 Vergleich zu Kapitalmarktprognosen von Experten

Die erste Hypothese lautet:

H1: Die Prognosen der Teilnehmer schneiden schlechter als empirische Kapitalmarktprognosen von Experten ab.

Um sich hierüber ein Urteil bilden zu können erfolgt ein Vergleich mit den Ergebnissen empirischer Studien. Diese werden in den Vergleich mit einbezogen, wenn folgende Kriterien erfüllt sind:

- es handelt sich um eine empirische Analyse von Kapitalmarktprognosen
- bei denen es sich im speziellen um Punktprognosen handelt
- die einen Horizont von 1-12 Monaten aufweisen
- und eines der Prognosegütemaße $MRAPnP$ oder U_2 zur Berechnung der Prognosequalität nutzen, da erst der Vergleich zur naiven Prognose eine Qualitätsbewertung über verschiedene Studien hinweg ermöglicht.

Insbesondere die Beschränkung auf die beiden Prognosegütemaße schränkt die Zahl der Studien stark ein. Der in den meisten Studien verwendete root mean square error lässt jedoch keine Vergleiche zwischen verschiedenen Untersuchungen zu, da er skalenabhängig ist. Die entsprechenden Studien sind Tabelle 10 zu entnehmen.

Tabelle 10: Übersicht zu Studien mit Vergleichsmöglichkeit der Prognosequalität

Autoren / Jahr	Gegenstand	# Zeitreihen (Länge)	PH in Mon.	$\bar{U}_2 /$ <i>MRAPnP</i> (% < 1)
Spiwoks (2002a)	Zinsniveau Deutschland	gleicher Datensatz, aber aktueller: Spiwoks / Bedke / Hein 2010		
Bofinger und Schmidt (2003)	Devisenkurs: EUR/USD	8 (48-50)	1/3/6 /12/24	1,33 (0%)
Spiwoks (2004a)	Aktienindizes: DJ, FTSE, DAX,...	6 (99)	6	1,06 (17%)
Scheier und Spiwoks (2006)	Zinsniveau Großbritannien	73 (63-180)	3/12	1,24/1,25* (16%)
Spiwoks u. a. (2008)	Zinsniveau USA	132 (50-171)	3/12	1,13 (23,5%)
Spiwoks u. a. (2009)	Zinsniveau Schweiz	40 (ca. 80)	3/12	1,27 (18%)
Spiwoks u. a. (2010)	Zinsniveau Deutschland	112 (39-203)	3/12	1,11 (33%)

PH = Prognosehorizont; * Da bis auf eine Studie alle Studien den U_2 als Prognosegütemaß verwenden wird zur Berechnung der Mittelwerte aller Studien ebenfalls der Wert U_2 verwendet

Die Ergebnisse der Studien sind in ihrer Aussagekraft nicht direkt vergleichbar. Es werden unterschiedliche Prognosegegenstände bei verschiedenen Prognosehorizonten und Volkswirtschaften betrachtet. Die Zahl der eingehenden Prognosen schwankt zwischen knapp 400 bei Bofinger und Schmidt (2003) und knapp 18.000 bei Spiwoks u. a. (2010). Die Zahl der untersuchten Zeitreihen schwankt zwischen 6 bei Spiwoks (2004a) und 132 bei Spiwoks u. a. (2008). Gemeinsam haben alle Studien, dass die Mehrzahl der untersuchten Prognosezeitreihen der professionellen Analysten schlechter als die naive Prognose abschneiden.

 Tabelle 11: Gegenüberstellung des U_2 empirischer Studien und der Planspielteilnehmer

Wert	Empirische Studien					Planspiel
	\bar{U}_2 ungew.	\bar{U}_2 gew. n. # Zeitreihen	Median	Min	Max	\bar{U}_2
U_2	1,19	1,16	1,19	1,06	1,33	1,22
% besser als naiv	17,9%	23,7%	17,5%	0%	33%	10%

\bar{U}_2 = arithmetischer Mittelwert; gew. = gewichtet; # = Anzahl

In Tabelle 11 werden die in den Studien ermittelten U_2 den Werten des Planspiels gegenübergestellt. Nimmt man den ungewichteten arithmetischen Mittelwert der sechs Studien so ergibt sich ein U_2 von 1,19. Bei einer Gewichtung nach der Zahl der Zeitreihen beträgt der Wert 1,16. Der in dem Planspiel durch die 20 Zeitreihen (10 Gruppen á 2 Prognosehorizonte) ermittelte

Wert von 1,22 liegt zwar etwas höher, die empirischen Analysen der Hälfte der Studien haben jedoch noch schlechtere Werte ergeben. 10% der Gruppen haben besser als die naive Prognose abgeschnitten, was unter dem ungewichteten arithmetischen Mittel der empirischen Studien von 17,9% liegt. Dies kann zufallsbedingt sein, da 3 Gruppen nur ganz knapp über dem Wert von 1 liegen. Es hätten somit ohne weiteres auch 25% der Gruppen besser als die naive Prognose abschneiden können. Eine solche Ungenauigkeit dürfte zumindest bei zwei Dritteln der Studien nicht vorliegen, da sie mindestens über die doppelte Anzahl an ausgewerteten Zeitreihen verfügen.

Insgesamt lässt sich bei aller Vorsicht hinsichtlich der Vergleichbarkeit festhalten, dass ganz erhebliche Ähnlichkeiten bei der Prognosequalität zwischen den Analysten im Feld und den Planspielteilnehmern bestehen. Systematisch schlechter sind die Prognosen der Planspielteilnehmer nicht.

4.1.4 Zusammenfassung

Die Prognosezeitreihen verfügen ebenso wie der zugrundeliegende Aktienkurs über eine Stationarität 2. Ordnung. Dies ist eine notwendige Bedingung für eine gute Prognosequalität, sagt jedoch noch nichts über das Vorliegen dieser aus. Die Auswertung mittels der üblichen Prognosegütemaße ergibt, dass alle Gruppen den künftigen Aktienkurs im Durchschnitt überschätzt haben und mit ihren Prognosen nicht unerheblich daneben liegen. Bei den 12-Monats-Prognosen sind diese Effekte etwa doppelt so hoch wie bei den 3-Monats-Prognosen.

Hätten die Gruppen einfach den jeweiligen Gegenwartswert in die Zukunft extrapoliert, wären die Prognosen im Mittel deutlich besser gewesen, wie der Vergleich zur naiven Prognose zeigt. Damit sind die Planspielteilnehmer ohne vorherige Prognoseerfahrung auf einem Level mit den professionellen Prognostikern, wie abschließend der Vergleich mit empirischen Analysen von Kapitalmarktprognosen zeigt, bei denen ebenfalls ein Vergleich zur naiven Prognose mit der gleichen Methodik vorgenommen wurde.

Die erste Hypothese („*H1: Die Prognosen der Teilnehmer schneiden schlechter als empirische Kapitalmarktprognosen von Experten ab.*“) muss abgelehnt werden. Die Studierenden verfügen über weniger Expertenwissen, eine geringere Informationsbasis sowie höheren Zeitdruck. Dennoch schneiden sie im Großen und Ganzen genauso gut bzw. schlecht ab wie sich dies bei empirischen Prognosezeitreihen messen lässt, die in einem völlig anderen institutionellen Umfeld entstanden sind.

4.2 Erwartungsbildung

Die im Abschnitt 2.2 hergeleitete Hypothese lautet:

H2: Die Prognosefehler weisen systematische Komponenten auf und entsprechen somit nicht den Kriterien der rationalen Erwartungsbildung.

Zunächst erfolgt eine graphische Einschätzung der Prognosefehler mittels Prognose-Realisations-Diagramme in Abschnitt 4.2.1. Die rechnerischen Ergebnisse der Überprüfung dieser Hypothese mittels des Tests auf Unverzerrtheit werden in Abschnitt 4.2.2 präsentiert. In Abschnitt 4.2.3 wird auf das Vorliegen der adaptiven und extrapolativen Erwartungsbildung getestet. Eine Diskussion der Ergebnisse erfolgt in Abschnitt 4.2.4. Abschnitt 4.2.5 fasst dieses Kapitel zusammen.

4.2.1 Prognose-Realisations-Diagramm

Für eine qualitative Beurteilung der Prognosefehler bietet sich das von Theil (1961) entwickelte Prognose-Realisations-Diagramm an. Hierzu werden zunächst die relativen Prognosefehler ρPF und die realisierten relativen Änderungen der Aktienkurse der Alpha-Aktie ρA berechnet. P ist dabei wieder die Prognose, A der tatsächliche Aktienkurs, t die Zeitvariable und h der Prognosehorizont. A_t ist der realisierte Kurs zum Zeitpunkt der Prognosegeltung und A_{t-h} ist der Kurs zum Zeitpunkt der Prognoseerstellung.

$$(7) \quad \rho PF = \frac{P_t - A_{t-h}}{A_{t-h}} \quad \text{und} \quad \rho A = \frac{A_t - A_{t-h}}{A_{t-h}}$$

Anschließend werden die für alle Teilnehmergruppen berechneten Werte in einem Punktdiagramm zusammengeführt, wobei auf der x-Achse die realisierten relativen Kursänderungen und auf der y-Achse die dazugehörigen relativen Prognosen abgetragen werden (Abbildung 7 und Abbildung 8).

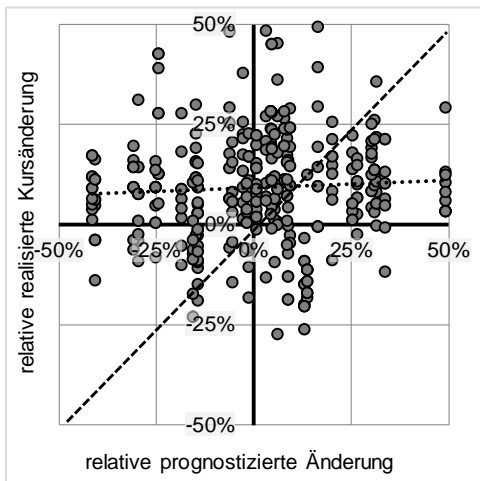


Abbildung 7: Prognose-Realisations-Diagramm für die 3-Monats-Prognosen

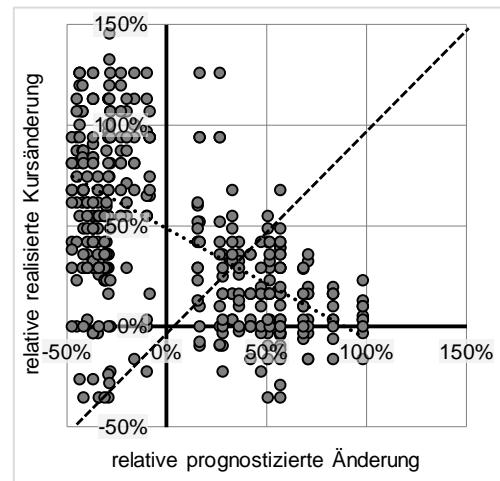


Abbildung 8: Prognose-Realisations-Diagramm für die 12-Monats-Prognosen

Die Diagonale (gestrichelte Linie) zeigt den Bereich der möglichen perfekten Prognosen an, also eine Übereinstimmung von realisierter und prognostizierter Kursänderung. Bei einer guten Prognosezeitreihe müssten die Werte also nahe an der Diagonalen liegen. Will man systematische Prognosefehler in Form regelmäßiger Über- oder Unterschätzungen ausschließen, sollte eine Gleichverteilung der Werte um die Diagonale herum feststellbar sein. Dies lässt sich mittels der Steigung der Regressionsgraden (gepunktete Linie) überprüfen, die in diesem Fall 1 sein sollte.

Eine solche Gleichverteilung lässt sich für die Prognosen der Planspielteilnehmer jedoch ausschließen. Bei den 3-Monats-Prognosen ist die starke Ansammlung von Werten links der Diagonalen optisch gut erkennbar (Abbildung 7). Bei realisierten Kursverlusten haben sie häufig mit Kursgewinnen gerechnet (2. Quadrant). Zudem sind kleine Kursgewinne erheblich über- und große Kursgewinne erheblich unterschätzt worden. Die Trendlinie ist mit 0,04 leicht positiv, jedoch weit von 1 entfernt.

Die Streuung bei den 12-Monats-Prognosen ist noch erheblich größer. Auffällig sind die vielen Wendepunktfehler im 2. Quadranten, bei denen mit Kursgewinnen gerechnet wurde obwohl es zu Rückgängen kam. Zudem sind sehr hohe Steigerungen erwartet worden, die so nicht eingetroffen sind. Einige Extremwerte jenseits der 200% werden aus Platzgründen nicht dargestellt. Rechts der Diagonalen sind im 1. Quadranten zahlreiche hohe Kurssteigerungen deutlich zu niedrig prognostiziert worden. Die Steigung der Trendgraden ist mit -0,56 eindeutig negativ, womit der Schluss zulässig ist, dass es erhebliche systematische Komponenten in den Prognosefehlern gibt.

4.2.2 Test auf Unverzerrtheit

Der am weitesten verbreitete Test auf das Vorliegen rationaler Erwartungen bei Prognosen ist der sogenannte Test auf Unverzerrtheit mittels der Mincer-Zarnowitz-Regression.³⁰ Dabei wird überprüft, ob die Prognosefehler eine Systematik aufweisen. Dies würde im Widerspruch zur Theorie der rationalen Erwartungen stehen.

Als abhängige Variable dient der tatsächlich eingetretene Wert A . Geschätzt werden die Konstante α , der Koeffizient β der entsprechenden Prognose P sowie der Fehlerterm u zum Zeitpunkt t .

$$(8) \quad A_t = \alpha + \beta P_t + u_t$$

Prognosen gelten als unverzerrt, wenn die Gemeinschaftshypothese von $\alpha = 0$ und $\beta = 1$ mit geringer Irrtumswahrscheinlichkeit bestätigt werden kann. Dies wird mit einem Wald-Test überprüft. Eine weitere Bedingung ist die Abwesenheit von Autokorrelation bei den geschätzten Werten des Fehlerterms u , was mit einem Durbin-Watson-Test überprüft wird. Liegt nach diesen Kriterien eine Prognosezeitreihe im Sinne der rationalen Erwartungsbildung vor, so bedeutet dies nach Granger und Newbold (1973) jedoch nicht, dass die Prognosen perfekt sind. Sie weisen lediglich keine systematischen Fehler auf.

Unter anderem Webb (1987) weist auf mögliche Probleme der Anwendung dieses Testverfahrens bei nicht-stationären Zeitreihen hin. Da die Ergebnisse besser interpretierbar sind wird zunächst das Standardverfahren nach (8) verwendet. Die Ergebnisse der Berechnung eines angepassten Testverfahrens sind in Anhang A zu finden, entsprechen in deren Aussage aber denen des Standardverfahrens in Tabelle 12.

Tabelle 12: Ergebnisse des Tests auf Unverzerrtheit der einzelnen Gruppen

Gruppe	# Beob.	3-Monats-Prognosen				12-Monats-Prognosen			
		α	β	D-W-Stat.	p-Wert H: $\alpha=0$ und $\beta=1$	α	β	D-W-Stat.	p-Wert H: $\alpha=0$ und $\beta=1$
1	39	5,31 (2,11)	0,67 (0,11)	0,48	0,004***	19,25 (2,59)	-0,10 (0,11)	0,17	0,000***
2	39	5,02 (2,74)	0,67 (0,14)	0,33	0,019**	24,30 (2,93)	-0,38 (0,15)	0,24	0,000***

³⁰ Mincer und Zarnowitz (1969).

3	35	5,77 (1,76)	0,61 (0,09)	0,48	0,000***	16,96 (3,36)	-0,01 (0,15)	0,15	0,000***
4	39	8,81 (1,71)	0,46 (0,08)	0,40	0,000***	15,32 (1,24)	0,07 (0,04)	0,18	0,000***
5	39	5,63 (2,07)	0,64 (0,10)	0,47	0,000***	19,48 (2,62)	-0,10 (0,11)	0,18	0,000***
6	39	5,55 (2,60)	0,64 (0,13)	0,35	0,004***	25,13 (3,28)	-0,34 (0,14)	0,22	0,000***
7	39	6,96 (2,21)	0,56 (0,11)	0,49	0,000***	19,44 (2,21)	-0,10 (0,09)	0,17	0,000***
8	24	4,89 (4,77)	0,67 (0,21)	0,27	0,040**	24,84 (3,90)	-0,37 (0,17)	0,30	0,000***
9	39	6,44 (2,04)	0,56 (0,09)	0,42	0,000***	20,09 (2,67)	-0,12 (0,10)	0,18	0,000***
10	38	6,68 (2,88)	0,55 (0,14)	0,36	0,000***	22,68 (3,26)	-0,23 (0,13)	0,20	0,000***

α und β : Höhe der konstanten und des Koeffizienten des Tests auf Unverzerrtheit (Standardfehler in Klammern). p-Wert H: $\alpha=0$ und $\beta=1$: Irrtumswahrscheinlichkeit für die Gültigkeit der Hypothese, dass die Prognosefehler systematische Komponenten aufweisen (* $p<0,1$; ** $p<0,05$; *** $p<0,01$). Durbin-Watson-Statistik: 0 = perfekt positive Autokorrelation, 2 = keine Autokorrelation, 4 = perfekt negative Autokorrelation. Der kritische Wert (5%-Signifikanzniveau) liegt je nach Zahl der Beobachtungen zwischen 1,45 und 1,54.

Alle untersuchten Prognosezeitreihen weisen systematische Prognosefehler im Sinne des Tests auf Unverzerrtheit von Mincer und Zarnowitz (1969) auf. Die Durbin-Watson-Statistik zeigt für alle Prognosezeitreihen eine deutliche positive Autokorrelation erster Ordnung auf, wenngleich hierfür die Nicht-Stationarität der zugrundeliegenden Zeitreihen verantwortlich sein kann. Das angepasste Testverfahren (Anhang A) kommt, wenn auch etwas weniger eindeutig, zu dem gleichen Ergebnis. Es sind also systematische Bestandteile in den Residuen messbar. Bei den Zeitreihen der Gruppen 3 und 8 gibt es Datenlücken, welche jedoch bei dem hier angewandten Testverfahren keine Auswirkungen haben dürfte.³¹

Die Gemeinschaftshypothese einer Konstanten von 0 und eines Koeffizienten von 1 wird für alle Prognosezeitreihen sicher abgelehnt. Auf eine mögliche Ursache für die systematischen Prognosefehler weisen vor allem bei den 12-Monats-Prognosen die niedrigen Koeffizienten β hin. Es ist kein Zusammenhang zwischen den Prognosen und den tatsächlich eingetreten Kursen nachweisbar. Weiterhin ist die Überlappung der Prognosehorizonte problematisch, was insbesondere bei den 12-Monats-Prognosen zum Tragen kommen könnte. Ob die 12-Monats-Prognosen systematische Fehler aufweisen, können die Gruppen erst ab Spielrunde 13 erfahren.

³¹ Zur Verwendung der Durbin-Watson-Statistik bei Zeitreihen mit Datenlücken siehe Bhargava (1989).

Um die Rationalität der Prognosen in beiden Hälften des Planspiels bewerten zu können, ist eine Panelbetrachtung der Prognosedaten erforderlich, da eine Halbierung der Zeitreihen auf Gruppenbasis zu einer zu geringen Zahl an Beobachtungen führt (Anhang A). Die Unterschiede der Schätzergebnisse zwischen der ersten und zweiten Planspielhälfte sind marginal. Keinesfalls kann ein Lernprozess der Gruppen im Planspiel festgestellt werden. Die Prognosen in der zweiten Planspielhälfte sind genauso mit systematischen Fehlern behaftet wie in der ersten Hälfte. Auffällig ist das extrem niedrige Bestimmtheitsmaß der Schätzungen. Dies ist Folge des kaum nachweisbaren Zusammenhangs zwischen den Prognosen und den tatsächlich eingetretenen Werten.

4.2.3 Test auf extrapolative und adaptive Erwartungsbildung

Bei der extrapolativen Erwartungsbildung orientieren sich die Prognostiker ausschließlich an der vergangenen Kursentwicklung. Bei einer 3-Monats-Prognose würde man je nach Form der extrapolativen Erwartungsbildung davon ausgehen, dass sich die Änderungsrate der letzten 3 Monate verstärkt, umkehrt oder genauso fortsetzt. Getestet wird dies wieder mit einem Schätzmodell mittels linearer Regression analog zu Leitner und Schmidt (2006).

Als abhängige Variable dient die für die Zukunft zum Zeitpunkt t prognostizierte Änderung θP . Hierzu wird vom prognostizierten Kurs P zum Geltungszeitpunkt t der Kurs A zum Zeitpunkt der Prognoseerstellung $t-h$ subtrahiert. Als unabhängige Variable dient die Änderung des Aktienkurses θA vor h Monaten bis zum Zeitpunkt der Prognoseabgabe $t-h$.

$$(9) \quad \theta P_t = P_t - A_{t-h} \quad \text{und} \quad \theta A_t = A_{t-h} - A_{t-2h}$$

Die im Abschnitt 4.1.1 erläuterte Stationaritätsproblematik zeigt sich bei den mittels (9) berechneten Zeitreihen, weshalb deren jeweils erste Differenzen Δ Grundlage für die Schätzung sind. Geschätzt werden die Konstante α , der Koeffizient β sowie der Fehlerterm u zum jeweiligen Zeitpunkt t .

$$(10) \quad \Delta \theta P_t = \alpha + \beta \Delta \theta A_t + u_t$$

Für die Interpretation spielt der Koeffizient β die entscheidende Rolle. Bei einem Wert von > 0 rechnen die Prognostiker damit, dass sich der vergangene Trend fortsetzt. Ist der Wert dabei größer als 1 rechnen sie mit einer Verstärkung und zwischen 0 und 1 mit einer Abschwächung.

Bei einem negativen Koeffizienten rechnen sie damit, dass sich der Trend umkehrt. Der Auffassung von Leitner und Schmidt (2006), dass ein β von 0 mit der statischen Erwartung gleichzusetzen ist, können wir hingegen nicht folgen. Zwar würde sich die statische Erwartung in einem β von 0 widerspiegeln, umgekehrt gilt dies jedoch nicht zwingend. Besteht nämlich kein Zusammenhang zwischen der prognostizierten Änderungsrate und dem jeweils aktuellen Trend, so dürfte ebenso ein β von 0 geschätzt werden.

Problematisch ist die Implementierung des Tests bei den Prognosen für den Aktienkurs in 12 Spielrunden bzw. Monaten. Den Teilnehmern liegt jeweils nur der Aktienkurs der letzten 6 Spielrunden vor, so dass sich der Test bei den 12-Monats-Prognosen auf die Extrapolation des Trends aus den letzten 6 Spielrunden beschränkt.

In Tabelle 13 werden die Ergebnisse des Tests auf extrapolative Erwartungen dargestellt.

Tabelle 13: Ergebnisse des Tests auf extrapolative Erwartungen der einzelnen Gruppen

Gruppe	# Beob.	3-Monats-Prognosen Trend: h = 3			12-Monats-Prognosen Trend: h = 6		
		α	β	R ²	α	β	R ²
1	38	0,01	-0,12	0,03	0,11	-0,22	0,06
2	38	0,00	-0,13	0,04	0,11	-0,54***	0,27
3	33	-0,12	-0,26*	0,10	-0,12	-0,28**	0,15
4	38	0,03	0,03	0,00	-0,04	0,28	0,03
5	38	-0,03	-0,02	0,00	-0,11	-0,20*	0,08
6	38	0,03	-0,12	0,02	0,01	-0,24	0,05
7	38	0,01	-0,02	0,00	-0,06	-0,05	0,00
8	22	-0,11	-0,26*	0,14	0,26	-0,50**	0,28
9	38	-0,09	0,14	0,04	-0,09	-0,06	0,00
10	37	0,01	-0,37***	0,18	-0,05	-0,48***	0,19
KP	38	0,00	-0,09	0,03	0,01	-0,22**	0,16

Erläuterung der Berechnung in (9) und (10). Signifikanzniveaus: *p<0,1; **p<0,05; ***p<0,01. KP = Konsensprognose.

Bei den 3-Monats-Prognosen rechnen die Gruppen 3, 8 und 10 mit einer gegenläufigen Entwicklung (mean revision³²) zum vorherrschenden Trend der letzten 3 Monate. Ansonsten scheint das Modell der extrapolativen Erwartungen bei den Prognosen kurzer Frist keinen Erklärungsansatz zu bieten. Anders stellt sich dies bei den 12-Monats-Prognosen dar, bei denen auf einen Einbezug des Trends aus den letzten 6 Monaten getestet wurde. Die Hälfte der Gruppen weist einen signifikanten Zusammenhang zum vorherrschenden Trend auf. Sie rechnen dabei mit einer Umkehrung des Trends. Möglicherweise ist dies dem Verlauf des Aktienkurses in der Mitte des Planspiels geschuldet. Mit einer Fortsetzung des längeren Abwärtstrends haben die Teilnehmer scheinbar nicht gerechnet.

Bei der adaptiven Erwartungsbildung nehmen die Prognostiker ihre eigenen Prognosefehler zur Kenntnis und readjustieren ihre Prognosen entsprechend. Voraussetzung hierfür ist, dass sie ihre eigenen Prognosefehler überhaupt erst einmal zur Kenntnis nehmen. Dies stellt beim Planspiel keine Selbstverständlichkeit dar, da die Gruppen ihre Fehler hierfür selbst berechnen

³² Bei der mean revision bzw. der Regression zum Mittelwert handelt es sich um die bei zahlreichen Phänomenen zu beobachtende Tatsache, dass auf einen extremen Wert fast immer ein Wert näher am Mittelwert folgt. Die Entdeckung ist in der zweiten Hälfte des 19. Jahrhunderts gemacht worden geht auf den britischen Wissenschaftler Francis Galton zurück. Die Bedeutung dieses Phänomen bei Aktienkursen ist seit Jahrzehnten umstritten. So kommt bspw. Mukherji (2011) bei einer umfangreichen Analyse zu dem Schluss, dass sich der Effekt in den letzten knapp einhundert Jahren abgeschwächt hat, vor allem bei den Titeln kleinerer Firmen aber immer noch eine Rolle spielt. Spierdijk u. a. (2012) weisen bei einer ebenfalls sehr umfangreichen Analyse auf die besondere Ausprägung des Phänomens nach exogenen Schocks in Zeiten großer Unsicherheit hin.

müssten. Eine gruppenbezogene konkrete Information haben sie von den Spielleitern hierzu nicht erhalten.

Für den Test auf das Vorliegen adaptiver Erwartungsbildung wird überprüft, ob sich die Prognose in Abhängigkeit vom letzten bekannten Prognosefehler ändert. Als abhängige Variable dient die Differenz der in (9) erläuterten Prognose der Änderung des Aktienkurses für den Zeitpunkt t und der Prognose für den Zeitpunkt der Abgabe dieser Prognose $t-h$. Es handelt sich also nicht um die Prognose selbst, sondern um die Änderung im Vergleich zur letzten Prognose, deren Erfolg sich messen lässt. Die unabhängige Variable ist der jeweils letzte messbare Prognosefehler, der sich aus der Differenz zwischen der tatsächlichen Änderung und der prognostizierten Änderung ergibt. Auch in diesem Fall wird wegen der Stationaritätsproblematik mit ersten Differenzen Δ geschätzt.

$$(11) \quad \Delta\theta P_t - \Delta\theta P_{t-h} = \alpha + \beta (\Delta\theta A_{t-h} - \Delta\theta P_{t-h}) + u_t$$

Auch bei dieser Schätzung ist der Koeffizient β entscheidend. Liegt er bei 1, so ist im Durchschnitt eine Anpassung in Höhe des jeweils letzten Prognosefehlers vorgenommen worden. Für das Vorliegen adaptiver Erwartungsbildung spräche ein signifikant positiver Koeffizient bei einer Konstanten α von 0.

Problematisch ist der Verlust an Beobachtungen bei diesem Testverfahren. So ist bei den 3-Monats-Prognosen für die ersten 3 Spielrunden keine Berechnung möglich, da den Gruppen noch kein Prognosefehler bekannt ist. Bei den 12-Monats-Prognosen liegt der erste Prognosefehler naturgemäß erst nach 12 Spielrunden vor. Dies schränkt die Zuverlässigkeit der Schätzungen bei den 12 Monats-Prognose massiv ein.

In Tabelle 14 werden die Ergebnisse des Tests auf adaptive Erwartungen dargestellt.

Tabelle 14: Ergebnisse des Tests auf adaptive Erwartungen der einzelnen Gruppen

Gruppe	# Beob.	3-Monats-Prognosen			# Beob.	12-Monats-Prognosen		
		α	β	R ²		α	β	R ²
1	35	-0,02	0,23	0,06	26	-0,22	0,09	0,01
2	35	-0,07	0,25	0,05	26	-0,57	-0,82***	0,44
3	30	0,01	0,29	0,06	21	0,43	-0,27*	0,18
4	35	-0,02	0,43***	0,11	26	-0,19	0,75**	0,24
5	35	-0,03	0,33**	0,12	26	-0,20	-0,03	0,00
6	35	-0,02	0,22	0,05	26	-0,11	0,04	0,00
7	35	-0,07	0,42***	0,20	26	-0,12	0,28	0,10
8	16	-0,72	-0,28	0,05	7	-	-	-
9	35	-0,05	0,43***	0,27	26	-0,37	0,09	0,01
10	34	-0,01	0,37*	0,11	25	-0,81	-0,25	0,06
KP	39	-0,06	-0,03	0,00	26	-0,32	-0,17	0,09

Erläuterung der Berechnung in (11). Signifikanzniveaus: *p<0,1; **p<0,05; ***p<0,01. KP = Konsensprognose.

Bei den 3-Monats-Prognosen ist bei 5 Gruppen zu erkennen, dass ihre vergangenen Prognosefehler positiv mit den jeweils aktuellen Prognosen zusammenhängen. Adaptive Erwartungsbildung könnte in diesen Fällen also vorliegen, zumal auch die Konstante α nicht signifikant von 0 abweicht. Bei den 12-Monats-Prognosen weist Gruppe 4 einen positiven Zusammenhang auf. Es liegen jedoch nur 26 Beobachtungen vor. Bei zwei Gruppen ist der Zusammenhang negativ.

4.2.4 Diskussion

In diesem Kapitel wurde auf drei verschiedene Formen der Erwartungsbildung getestet. Tabelle 15 fasst die Ergebnisse dieser Test zusammen.

Tabelle 15: Zusammenfassung der Tests auf die verschiedenen Formen der Erwartungsbildung

Gruppe	3-Monats-Prognosen			12-Monats-Prognosen		
	rational	extrapolativ	adaptiv	rational	extrapolativ	adaptiv
1	nein	-	-	nein	-	-
2		-	-		gegenläufig	-
3		gegenläufig	-		gegenläufig	-
4		-	ja		-	ja
5		-	ja		gegenläufig	-
6		-	-		-	-
7		-	ja		-	-
8		gegenläufig	-		gegenläufig	-
9		-	ja		-	-
10		gegenläufig	ja		gegenläufig	-
KP	-	-	gegenläufig	-		

KP = Konsensprognose

Die rationale Erwartungsbildung liegt bei keiner der Gruppen vor. Bei den Prognosen kurzer Frist scheint entweder ein sich umkehrender Trend (extrapolativ) oder die Anpassung an den jeweils letzten Prognosefehler (adaptiv) eine Rolle zu spielen. Bei den 12-Monats-Prognosen wird überwiegend mit einer Umkehrung des Trends der letzten 6 Monate gerechnet (extrapolativ). Einzig Gruppe 4 weist adaptive Erwartungsbildung auf.³³

Die Ergebnisse sind insgesamt konsistent mit den in Abschnitt 2.2 geschilderten Ergebnissen anderer empirischer und experimenteller Studien. Insbesondere das Ausbleiben der rationalen Erwartungsbildung wird in aller Regelmäßigkeit festgestellt. Am ehesten methodisch vergleichbar ist eine Studie von Leitner und Schmidt (2006) zu Wechselkursprognosen, bei der ebenfalls ein Vergleich experimenteller und empirischer Daten erfolgt. Deren experimentellen Daten sind Resultat einer deutlich weniger aufwendigen Informationsauswertung der Probanden, da sie lediglich über vergangene Kursdaten als Entscheidungsgrundlage verfügten. Bei ihnen kann rationale Erwartungsbildung bei Prognosen mit einem Horizont von größer 1 als einem Monat ebenfalls nicht festgestellt werden. Während die Experten mit einer Umkehrung des Trends rechnen, haben die Anfänger eine Verstärkung erwartet und lagen damit besser.

Die Erklärung für das Verhalten der Profis ist nach Leitner und Schmidt (2006) in ihrer Erwartung eines langfristigen Gleichgewichts bei den Wechselkursen zu suchen, da dies gängige Lehrmeinung sei. Es ist jedoch fraglich, ob das abweichende Verhalten der Anfänger überraschend ist. Diese verfügten ausschließlich über die vergangenen Kursinformationen, wodurch die Prognose eines sich fortsetzenden Trends als sehr naheliegend darstellt. Zu diskutieren ist auch der methodische Ansatz, da ausschließlich die Konsensprognosen untersucht worden sind. Hier könnten sich verschiedene Formen der Erwartungsbildung gegenseitig aufheben.

4.2.5 Zusammenfassung

Die Prognosen weisen keine rationale Erwartungsbildung im Sinne von Muth (1961) und Lucas (1972), wie die entsprechenden Tests im Abschnitt 4.2.2 zeigen. Die Hypothese („Die Prognosefehler weisen systematische Komponenten auf und entsprechen somit nicht den Kriterien der rationalen Erwartungsbildung“) kann daher vorläufig bestätigt werden.

³³ Betrachtet man die Ergebnisse dieser Gruppe im Detail, so fällt auf, dass sie ihre Prognosen am Anfang des Planspiels viel zu hoch angesetzt haben. Als sie gemerkt haben, dass die Kurse in der Mitte des Planspiels fielen, anstatt wie prognostiziert stark zu steigen, haben sie ihre Prognosen stark nach unten revidiert (adaptive Erwartungsbildung).

Im weiteren Verlauf des Kapitels wurde überprüft, ob einfachere Formen der Erwartungsbildung vorliegen. Die diesbezüglichen Tests in Abschnitt 4.2.3 ergeben ein gemischtes Bild bei den 3-Monats-Prognosen. Die Hälfte der Gruppen lassen ihre Prognosefehler in die Erwartungsbildung mit einfließen, während 3 Gruppen eine Umkehrung des Trends des Aktienkurses erwarten. Bei den 12-Monats-Prognosen war die Erwartung eines sich umkehrenden Trends bei der Hälfte der Gruppen von Bedeutung. Die Gruppen scheitern also mit unterschiedlichen Ausprägungen der Erwartungsbildung an der Prognoseaufgabe.

4.3 Gegenwartsbezug

Die im Abschnitt 2.3 hergeleitete erste Hypothese lautet:

H3: Die Prognosen weisen gegenwartsorientierte Verlaufsanpassung auf.

Lehnen die Prognostiker ihre Prognosen sehr stark an der jeweiligen Entwicklung zum Zeitpunkt der Prognoseerstellung an, so können sie ihren zukunftserfassenden Charakter verlieren.³⁴ Es ist daher von Interesse zu ermitteln, ob dieses bei Kapitalmarktprognosen von Experten regelmäßig auftretende Phänomen auch bei den Planspielgruppen zu beobachten ist.

Zunächst wird in Abschnitt 4.3.1 untersucht, mit welchem Zeitpunkt der Aktienkurszeitreihe die abgegebenen Prognosen am stärksten korrelieren. Es wäre denkbar, dass ein noch stärkerer Bezug zur Vorperiode oder einem Zeitpunkt in der Zukunft vor oder nach dem eigentlichen Geltungszeitpunkt besteht. In Abschnitt 4.3.2 erfolgt die Berechnung des GOVA-Koeffizienten, bei dem ein direkter Vergleich zwischen den Prognosen auf der einen und dem Geltungs- und Entstehungszeitpunkt auf der anderen Seite stattfindet. In Abschnitt 4.3.3 erfolgt eine Untersuchung dahingehend, ob die Studierenden die Spannweite der zukünftigen Änderungen unterschätzt haben. Ein Vergleich zwischen den berechneten Koeffizienten und anderen Studien erfolgt in Abschnitt 4.3.4. In Abschnitt 4.3.5 wird ein abschließendes Urteil darüber gefällt, ob die Prognosen wie im Rahmen der dritten Hypothese formuliert, eine gegenwartsorientierte Verlaufsanpassung aufweisen.

³⁴ Vgl. Andres und Spiwoks (1999).

4.3.1 Korrelation zum Aktienkurs

Um diese Hypothese zu überprüfen, ist zunächst eine Betrachtung der Kreuzkorrelationen zwischen den Prognosen und dem Aktienkurs zum Zeitpunkt der Prognoseerstellung aufschlussreich. Dabei wird für jede Prognosezeitreihe überprüft, ob sie die höchste Korrelation zum Aktienkurs in die eine oder andere Richtung zeitverzögert erreicht. Bei einer präzisen Prognosezeitreihe wäre zu erwarten, dass die höchste Korrelation mit einer Zeitverzögerung in Höhe des Prognosehorizonts auftritt.

Tabelle 16 ist der jeweils höchste Korrelationskoeffizient gemeinsam mit der Zeitverzögerungsstufe dargestellt.

Tabelle 16: Höchste Kreuzkorrelationen zwischen Prognosen und Aktienkurs

Horizont		Prognosezeitreihen der einzelne Gruppen									
		1	2	3	4	5	6	7	8*	9	10
3-Mon.	Verzögerung	0	0	0	0	0	0	0	0	0	0
	Korrelation	0,96	0,85	0,95	0,90	0,94	0,95	0,93	0,92	0,94	0,91
	Korr. Geltung	0,72	0,63	0,76	0,69	0,71	0,63	0,65	0,56	0,70	0,56
12-Mon.	Verzögerung	0	-4	0	0	0	1	0	-6	0	2
	Korrelation	0,93	-0,82	0,88	0,84	0,93	0,88	0,92	-0,52	0,91	0,85
	Korr. Geltung	-0,14	-0,38	-0,02	0,26	-0,16	-0,38	-0,18	-0,40	-0,19	-0,29

Korr. = Korrelationskoeffizient nach Pearson. * Gruppe 8 weist lediglich 24 Beobachtungen auf. Bei der Untersuchung der LAG-Variablen fallen weitere Beobachtungen weg, wodurch das Ergebnis für diese Gruppe nicht aussagekräftig ist.

Der starke Gegenwartsbezug wird bei allen 3-Monats-Prognosen deutlich. Die Korrelation zum prognostizierten Aktienkurs ist immer dann am höchsten, wenn man den Kurs zum Zeitpunkt der Prognoseerstellung zu Grunde legt (Verzögerung = 0). Vergleicht man die Prognosen mit dem um 3 Monate verzögerten Kurs, also demjenigen Kurs, der eigentlich prognostiziert werden sollte, so ist die Korrelation in jedem Fall niedriger.

Etwas weniger eindeutig ist das Ergebnis bei den 12-Monats-Prognosen. Bei den Gruppen 2 und 8 ist der Betrag des Korrelationskoeffizienten dann am höchsten, wenn die Prognosen mit dem Aktienkurs mehrere Spielrunden vor der Prognoseabgabe verglichen werden. Der Korrelationskoeffizient ist allerdings negativ. Die beiden Gruppen haben also die gegenteilige Kursbewegung prognostiziert. Dies ist konsistent mit dem Ergebnis aus Abschnitt 4.2.3, wonach beide Gruppen sehr deutlich mit einer gegenläufigen Trendentwicklung gerechnet haben. Bei den meisten Gruppen ist jedoch die Korrelation genau dann wieder am stärksten, wenn man die

Prognosen mit dem Aktienkurs zum Zeitpunkt der Prognoseerstellung vergleicht. Der Korrelationskoeffizient zum tatsächlich eingetretenen Kurs ist in allen Fällen sehr niedrig und schwankt zwischen -0,4 (Gruppe 8) und 0,26 (Gruppe 4).

In den folgenden beiden Abbildungen ist die Korrelation zwischen der durchschnittlichen Prognose aller Gruppen und dem Aktienkurs abgebildet. Bei den 3-Monats-Prognosen (Abbildung 9) wäre bei guten Prognosen eine hohe Korrelation mit dem Kurs 3 Monate nach der Prognoseabgabe zu erwarten, bei den 12-Monats-Prognosen (Abbildung 10) 12 Monate später (jeweils rot markiert). In beiden Fällen zeigt sich die höchste Korrelation zu demjenigen Aktienkurs, wie er zum Zeitpunkt der Prognoseabgabe vorliegt. Eine messbar höhere Korrelation zum tatsächlich eingetretenen Kurs liegt in beiden Fällen nicht vor.

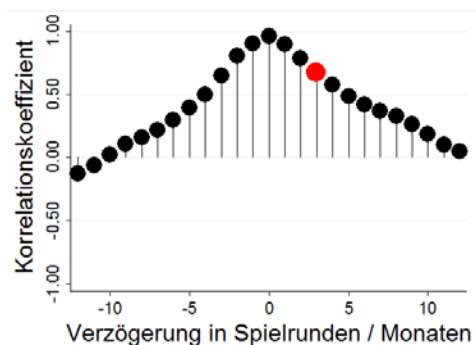


Abbildung 9: Kreuzkorrelation zwischen 3-Monats-Durchschnittsprognosen und dem Aktienkurs

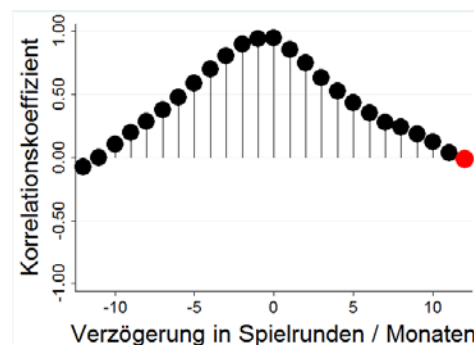


Abbildung 10: Kreuzkorrelation zwischen 12-Monats-Durchschnittsprognosen und dem Aktienkurs

Der Korrelationskoeffizient zum jeweils aktuellen Aktienkurs zum Zeitpunkt der Prognoseerstellung ist in beiden Fällen sehr hoch. Auch zu den Monaten kurz zuvor und kurz nach der Prognoseabgabe lässt sich noch eine hohe Korrelation messen. Dies ist durch die starke Autokorrelation erster Ordnung bedingt, die mittels des ADF-Tests in Abschnitt 4.1.1 festgestellt worden ist. Um zu einem noch klareren Ergebnis zu kommen, wird die Korrelation mit den in ihre ersten Differenzen transformierten Zeitreihen wiederholt (zur Bildung der ersten Differenzen siehe Abschnitt 4.1.1). Das Ergebnis ist Abbildung 11 und Abbildung 12 zu entnehmen:

Abbildung 11: Kreuzkorrelation zwischen 3-Monats-Durchschnittsprognosen und dem Aktienkurs (jeweils erste Differenzen)

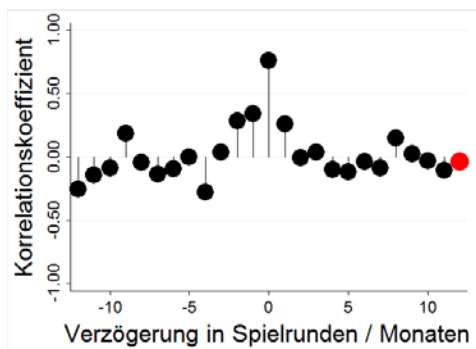


Abbildung 11: Kreuzkorrelation zwischen 3-Monats-Durchschnittsprognosen und dem Aktienkurs (jeweils erste Differenzen)

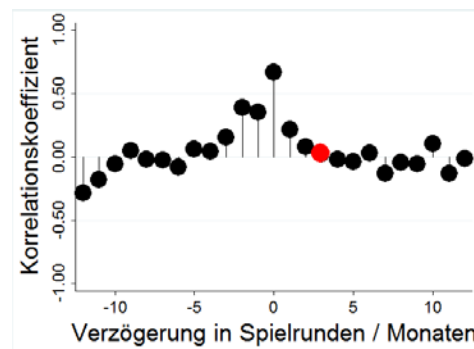


Abbildung 12: Kreuzkorrelation zwischen 12-Monats-Durchschnittsprognosen und dem Aktienkurs (jeweils erste Differenzen)

Es ist nun noch deutlicher zu erkennen, dass ein Zusammenhang zwischen den Prognosen und dem jeweils aktuellen Aktienkurs besteht. Durch die Betrachtung der ersten Differenzen lässt sich sagen, dass eine Änderung des aktuellen Kursniveaus im Vergleich zur Vorperiode mit einer Änderung der Prognose für den Kurs in 3 Perioden bzw. 12 Perioden einhergeht. Die Korrelation des Kurses zum Geltungszeitpunkt ist wieder rot markiert worden und spielt abermals keine Rolle.

4.3.2 Gegenwartsorientierte Verlaufsanpassung

Der über die Kreuzkorrelation im vorigen Abschnitt deutlich gewordene starke Zusammenhang der Prognosen zum Aktienkurs bei der Prognoseerstellung lässt sich mit weiteren Prognosegütemaßen bestätigen. Andres und Spiwoks (1999) nennen das Phänomen des starken Zusammenhangs zur Gegenwart gegenwartsorientierte Verlaufsanpassung (GOVA) und schlagen zur Messung die Bildung eines Koeffizienten aus dem Bestimmtheitsmaß (R^2) des Zusammenhangs zwischen Prognosen P und tatsächlich zum Zeitpunkt t eingetretenem Kurs A sowie den Prognosen und dem Kurs zum Zeitpunkt der Prognoseerstellung vor. Um den Kurs zum Entstehungszeitpunkt der Prognose zu erhalten, wird der Aktienkurs um den Prognosehorizont h zurückverschoben.

$$(12) \quad GOVA\text{-Koeffizient} = \frac{R^2_{\text{Prognosen zum Geltungszeitpunkt, Kurs}}}{R^2_{\text{Prognosen zum Entstehungszeitpunkt, Kurs}}}$$

Die Interpretation des GOVA-Koeffizienten lässt sich wie folgt vornehmen: Bei einem Wert kleiner als 1 ist ein stärkerer Zusammenhang zwischen der Prognosezeitreihe und dem Gegenwartskurs zu ihrem Entstehungszeitpunkt messbar als zu ihrem eigentlichen Geltungszeitpunkt. Ein Wert von größer als 1 weist auf das umgekehrte Resultat hin. Im Hinblick auf eine gute

Prognosequalität ist ein GOVA-Koeffizient wünschenswert, der deutlich größer als 1 ist. Prognosen, die mehr über die Gegenwart als über die Zukunft aussagen, sind in aller Regel wenig hilfreich.

Ein GOVA-Koeffizient größer als 1 ist jedoch nicht unbedingt ein Hinweis für gute Prognosen. Sind beide Bestimmtheitsmaße sehr niedrig, kann durchaus der Zufall über die Höhe des GOVA-Koeffizienten entscheiden. In diesem Fall weisen die Prognosen weder zur Gegenwart noch zur Zukunft einen starken Zusammenhang auf. Außerdem ist zu beachten, dass die betrachteten Zeitreihen nicht einfach einem Trend folgen sondern mehrere Hoch- und Tiefpunkte aufweisen.³⁵ Eine Prognosezeitreihe, die dem Trend in der Gegenwart folgt, würde in einem solchen Fall ein hohes Bestimmtheitsmaß für den Zusammenhang zu den zukünftigen Werten aufweisen, obwohl sich dieser möglicherweise nur zufällig ergibt. Der dem Planspiel zugrunde liegende Aktienkurs ist von ausreichender Varianz, so dass sich diese Problematik nicht ergibt.

In Tabelle 17 sind die GOVA-Koeffizienten für die Planspielgruppen sowie die zugrundeliegenden Bestimmtheitsmaße für beide Prognosehorizonte dargestellt.

Tabelle 17: Bestimmtheitsmaße und GOVA-Koeffizienten

	PH	1	2	3	4	5	6	7	8	9	10	Ø
R ² Prognosen; Zukunft		0,51	0,39	0,57	0,47	0,51	0,40	0,42	0,31	0,49	0,31	0,44
R ² Prognosen; Gegenw.	3	0,93	0,73	0,91	0,80	0,88	0,90	0,86	0,84	0,89	0,83	0,86
GOVA-Koeffizient		0,55	0,54	0,63	0,59	0,58	0,44	0,49	0,37	0,55	0,37	0,51
R ² Prognosen; Zukunft		0,02	0,15	0,00	0,07	0,02	0,14	0,03	0,16	0,04	0,08	0,07
R ² Prognosen; Gegenw.	12	0,87	0,36	0,77	0,70	0,86	0,69	0,84	0,05	0,82	0,60	0,66
GOVA-Koeffizient		0,02	0,41	0,00	0,10	0,03	0,21	0,04	3,07	0,04	0,14	0,41

PH = Prognosehorizont in Monaten / Spielrunden; Ø = arithmetischer Mittelwert der 10 Gruppen ohne Gewichtung nach der Anzahl Beobachtungen; R² = Bestimmtheitsmaß

Bei den 3-Monats-Prognosen sind die Bestimmtheitsmaße für den Zusammenhang der Prognosen zur Gegenwart etwa doppelt so hoch wie zur Zukunft. Die GOVA-Koeffizienten liegen demnach deutlich unter 1 (Mittelwert 0,51). Bei den 12-Monats-Prognosen ist das Bild etwas differenzierter. Einen deutlichen Zusammenhang zur Zukunft weist keine der untersuchten Prognosezeitreihen auf. Das höchste Bestimmtheitsmaß wurde für Gruppe 8 mit 0,16 gemessen. Diese Gruppe weist mit 3,07 einen sehr hohen GOVA-Koeffizienten auf, was daran liegt, dass die Gruppe zu den Gegenwartswerten einen sehr niedrigen Zusammenhang aufweist ($R^2 =$

³⁵ Andres und Spiwoks (1999), S. 533-534.

0,05). Für alle anderen Gruppen ist der Zusammenhang zur Gegenwart bei den 12-Monats-Prognosen deutlich stärker ausgeprägt und die Koeffizienten liegen in allen Fällen sehr deutlich unter 1.

4.3.3 Erfassung des Ausmaßes der Änderungen

Gubaydullina u. a. (2011) sehen eine mögliche Ursache für die schlechte Qualität bei Kapitalmarktprognosen in der mangelnden Fähigkeit der Prognostiker, das Ausmaß der zukünftigen Änderungen zu erfassen. Ob die Planspielgruppen im Durchschnitt ebenso große Änderungen prognostiziert haben, wie sie dann tatsächlich eingetreten sind, wird mit folgendem Koeffizienten ermittelt:

$$(13) \quad \text{Ausmaß-Koeffizient} = \frac{1}{T} \sum_{t=1}^T \frac{|P_t - A_{t-h}|}{|A_t - A_{t-h}|}$$

Es erfolgt eine Division der absoluten Abstände zwischen den Prognosen P zum Geltungszeitpunkt t und dem tatsächlichen Wert A zum Zeitpunkt der Prognoseerstellung $t-h$ im Zähler sowie der durchschnittlichen tatsächlichen Änderung im Nenner. Weichen die Prognosen im Mittel stärker von der Gegenwart zum Zeitpunkt der Prognoseerstellung ab als die durchschnittliche Variation der zugrundeliegenden Kurszeitreihe, so ergibt der Koeffizient einen Wert von größer als 1. Unterschätzen die Planspielteilnehmer die zukünftigen Änderungen, so ergibt sich ein Wert von kleiner als 1.

Im Abstand von 3 Spielrunden bzw. Monaten hat sich der empirische Aktienkurs im Durchschnitt um 3,06 € und innerhalb von 12 Monaten um 6,85 € geändert. In Tabelle 18 werden die Ergebnisse der Berechnung des Koeffizienten dargestellt.

Tabelle 18: Berechnung des Ausmaß-Koeffizienten

	PH	1	2	3	4	5	6	7	8	9	10	Ø
Ausmaß-Koeffizient	3	0,48	0,85	0,49	1,02	0,60	0,62	0,82	0,56	0,99	0,94	0,74
	12	0,69	1,03	0,56	1,35	0,82	0,85	0,75	0,79	1,11	1,12	0,91

PH = Prognosehorizont in Monaten / Spielrunden; Ø = arithmetischer Mittelwert der 10 Gruppen ohne Gewichtung nach der Anzahl Beobachtungen

Bei den 3-Monats-Prognosen haben die Gruppen 4 und 9 das korrekte Ausmaß der zukünftigen Änderungen erfasst. Alle anderen Gruppen haben im Durchschnitt geringere Änderungen als die empirischen 3,06 € prognostiziert. Bei den 12-Monats-Prognosen haben 4 Gruppen das Ausmaß der Änderungen überschätzt (Koeffizient > 1) und 6 Gruppen unterschätzt (Koeffizient

<1). In der Tendenz lag also eine Unterschätzung vor, besonders deutlich und aussagekräftig ist dieses Ergebnis jedoch keinesfalls.

4.3.4 Vergleich zu empirischen Studien

In diesem Abschnitt erfolgt ein Vergleich des Gegenwartsbezugs der Prognosen der Studierenden mit dem Gegenwartsbezug der Kapitalmarktprognosen von Experten. Wie bereits in Abschnitt 4.1.3 beim Vergleich mit den naiven Prognosen empirischer Studien thematisiert, ist ein solcher Vergleich nicht unproblematisch, da es erhebliche Unterschiede bei den maßgeblichen Faktoren für das Zustandekommen der Resultate gibt (Prognostiker, Prognosegegenstand, Zeitraum, regionale Unterschiede). Mit der Forderung an ausreichend lange Zeitreihen mit Hoch- und Tiefpunkten sowie der Zufälligkeit des Zustandekommens bei sehr kleinen Bestimmtheitsmaßen weist der GOVA-Koeffizient darüber hinaus auch mögliche Ungenauigkeiten auf, die bei der Interpretation der Ergebnisse ebenfalls berücksichtigt werden müssen.

Die Recherche beschränkt sich auf Kapitalmarktprognosen und spart bspw. Konjunktur- und Inflationsprognosen aus, die ebenfalls Gegenstand einer Überprüfung auf das Vorliegen einer GOVA geworden sind.³⁶ Die Ergebnisse entsprechender Studien sind Tabelle 19 zu entnehmen.

Tabelle 19: Übersicht zu empirischen Studien mit Vergleichsmöglichkeit des GOVA-Koeffizienten

Autoren / Jahr	Gegenstand	# Zeitreihen (Länge)	PH in Mon.	GOVA-K. (% < 1)
Spiwoks (2002a)	Zinsniveau Deutschland	gleicher Datensatz, aber aktueller: Spiwoks / Bedke / Hein 2010		
Bofinger und Schmidt (2003)	Devisenkurs: EUR/USD	8 (48-50)	1/3/6 /12/24	0,40 (100%)
Spiwoks (2004a)	Aktienindizes: DJ, FTSE, DAX,...	6 (99)	6	0,73 (100%)
Scheier und Spiwoks (2006)	Zinsniveau Großbritannien	73 (63-180)	3/12	0,72 (100%)
Spiwoks u. a. (2008)	Zinsniveau USA	132 (50-171)	3/12	0,56 (100%)
Spiwoks u. a. (2009)	Zinsniveau Schweiz	40 (ca. 80)	3/12	0,42 (97,5%)
Spiwoks u. a. (2010)	Zinsniveau Deutschland	112 (39-203)	3/12	0,77 (100%)

PH = Prognosehorizont; GOVA-K. = GOVA-Koeffizient

³⁶ Vgl. Spiwoks u. a. (2011). Ein Vergleich ist wegen des völlig unterschiedlichen Prognosetypus nicht möglich. Bei Konjunkturprognosen handelt es sich um sogenannte fixed-event-Prognosen, bei denen ein gesamter Zeitraum (bspw. Jahr) bei sich stetig verkürzendem Prognosehorizont vorhergesagt wird. Bei Kapitalmarktprognosen handelt es sich um fixed-horizon-Prognosen, bei denen Prognosen für einen bestimmten Zeitpunkt bei konstantem Prognosehorizont abgegeben werden.

Von 371 betrachteten Prognosezeitreihen weist lediglich eine einzige Zeitreihe einen GOVA-Koeffizienten > 1 auf. Das Phänomen tritt bei Kapitalmarktprognosen völlig unerheblich vom Prognosegegenstand und –horizont auf. Auch die betrachtete Volkswirtschaft spielt keine Rolle. Im Mittel stellen die Studien fest, dass die Prognosen in etwa einen doppelt so hohen Zusammenhang zum Gegenwartswert wie zur Zukunft aufweisen. In Tabelle 20 erfolgt ein Vergleich mit dem Mittelwert der Studierendengruppen.

Tabelle 20: Gegenüberstellung der GOVA-Koeffizienten empirischer Studien und der Planspielteilnehmer

Wert	Empirische Studien					Planspiel
	Ø ungew.	Ø gew. n. # Zeitreihen	Median	Min.	Max.	Ø
GOVA-K.	0,60	0,64	0,64	0,40	0,77	0,46
% < 1	99,6%	99,7%	100%	97,5%	100%	95%

Ø = arithmetischer Mittelwert; gew. = gewichtet; # = Anzahl; GOVA-K. = GOVA-Koeffizient

Beim Planspiel ist ebenfalls nur für eine einzige Zeitreihe ein Wert > 1 festgestellt worden. Die Mittelwerte fallen dabei niedriger als bei den empirischen Studien aus. Wegen der Ungenauigkeiten des GOVA-Koeffizienten bei sehr niedrigen Bestimmtheitsmaßen und der geringeren Anzahl an Beobachtungen (20 Zeitreihen) ist das Ergebnis wohl noch im Bereich einer zufälligen Abweichung einzuordnen. Ein weiteres Indiz hierfür ist die Tatsache, dass zwei Studien einen noch geringeren Mittelwert des GOVA-Koeffizienten aufweisen.

Im vorigen Abschnitt wurde mittels des Ausmaß-Koeffizienten (13) gezeigt, dass die Studierenden bei den 3-Monats-Prognosen im Mittel das Ausmaß der zukünftigen Änderungen des Aktienkurses nur zu $\frac{3}{4}$ erfasst haben. Bei den 12-Monats-Prognosen war dieser Effekt nicht eindeutig nachweisbar. Gubaydullina u. a. (2011) stellen die entsprechenden Abweichungen vom Gegenwartswert bei einer umfangreichen Analyse von Konsensprognosen für das Zinsniveau verschiedener Volkswirtschaften dar. Bei Ihnen liegen die Mittelwerte mit 0,48 (3-Monats-Prognosen) und 0,51 (12-Monats-Prognosen) deutlich niedriger. Zumindest bei Zinsprognosen im Feld wird das Ausmaß weitaus erheblicher unterschätzt, Unterschiede bei der GOVA werden jedoch nicht sichtbar.

4.3.5 Zusammenfassung

Die in Abschnitt 2.3 hergeleitete Hypothese lautet:

H3: Die Prognosen weisen gegenwartsorientierte Verlaufsanpassung auf.

Schon die Betrachtung der Kreuzkorrelationen zwischen den Prognosen und den tatsächlichen Aktienkursen zu verschiedenen Zeitpunkten zeigt, dass zum jeweiligen Kurs zum Zeitpunkt der Prognoseerstellung der höchste Zusammenhang nachweisbar ist. Die nähere Untersuchung mittels des GOVA-Koeffizienten zeigt die stark ausgeprägte GOVA der Prognosen in aller Deutlichkeit. Die Hypothese wird also vorläufig bestätigt.

Anschließend wurde der Frage nachgegangen, ob die Gruppen das Ausmaß der künftigen Änderungen unterschätzt haben. Dies ist bei den 3-Monats-Prognosen der Fall, wenngleich deutlich weniger ausgeprägt als bei Zinsprognosen von Experten im Feld. Die GOVA ist bei den 12-Monats-Prognosen etwas stärker ausgeprägt, obwohl das Ausmaß der Änderungen einigermaßen richtig eingeschätzt wird. Die GOVA ist bei den Experten sehr ähnlich ausgeprägt, weshalb die mangelnde Einschätzung des Ausmaßes der künftigen Änderungen wohl nicht unbedingt ursächlich hierfür ist.

Insgesamt lässt sich festhalten, dass Erklärungsvarianten, die sich auf das institutionelle Umfeld von Finanzanalysten beziehen, angesichts der Ergebnisse des Planspiels als weniger evident erscheinen.

4.4 Prognoseunsicherheit

Die im Abschnitt 2.4 hergeleitete Hypothese lautet:

H4: Die subjektiv empfundene Unsicherheit der Gruppen stimmt mit der anhand des Kriteriums Prognosedivergenz objektiv messbaren Unsicherheit überein.

Im Unterschied zu Evaluationen von Kapitalmarktprognosen im Feld kann die Unsicherheit der Prognostiker im Planspiel unkompliziert abgefragt werden. Da die Prognosen in Gruppen entstehen ist auch die Angabe der Sicherheit Folge einer in den Gruppen getroffenen Entscheidung. Wie sich diese subjektive Sicherheit im Planspielverlauf entwickelt hat wird in Abschnitt 4.4.1 erläutert. Die Frage, wie sich demgegenüber die objektiv anhand des Kriteriums der Prognosedivergenz messbare Unsicherheit im Planspielverlauf dargestellt hat, wird in Abschnitt 4.4.2 beantwortet. In Abschnitt 4.4.3 schließlich kommt es zu einem Vergleich der beiden Unsicherheiten und somit zur Überprüfung der Hypothese. Die Zusammenfassung dieses Kapitels erfolgt in Abschnitt 4.4.4.

4.4.1 Subjektive Sicherheit

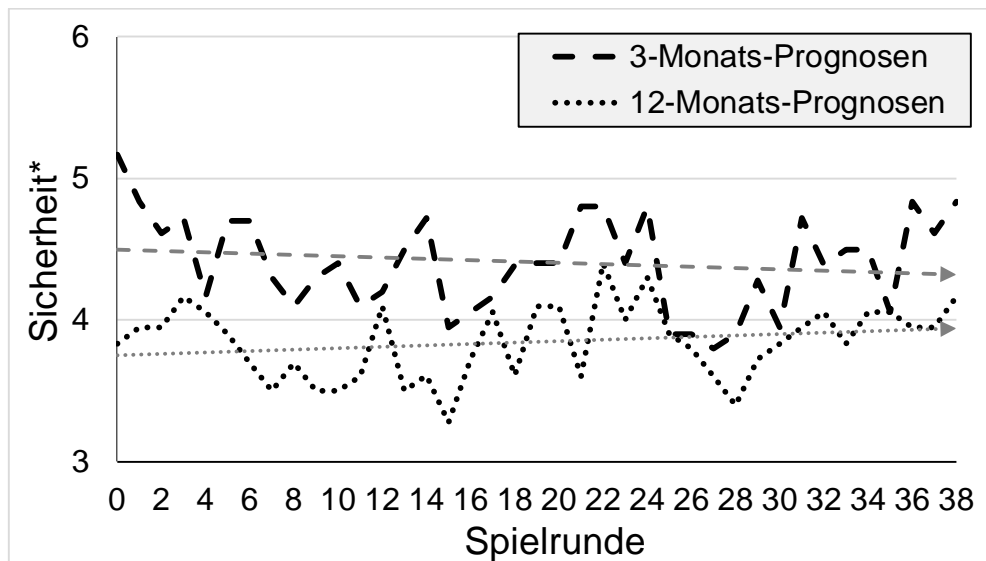
Neben den 3- und 12-Monats-Prognosen ist in jeder Spielrunde die Sicherheit der Gruppe über die Prognoseabgabe abgefragt worden. Die Gruppen sollten den Grad der Sicherheit auf einer Likert-Skala von 1 (sehr unsicher) bis 6 (sehr sicher) beantworten. Die durchschnittliche Unsicherheit der einzelnen Gruppen und die jeweilige Standardabweichung ist Tabelle 21 zu entnehmen.

Tabelle 21: Unsicherheit der einzelnen Gruppen bei der Prognoseabgabe

Horizont	Maß / Gruppe	1	2	3	4	5	6	7	8	9	10	Ø	σ
3-Monate	Sicherheit	4,28	3,56	3,46	3,85	4,59	3,74	3,97	4,67	3,74	3,39	3,93	0,43
	σ	0,88	1,01	0,81	1,03	0,93	0,84	1,40	0,69	0,71	0,54	0,88	
12-Monate	Sicherheit	3,05	2,79	3,43	2,90	3,82	3,15	3,79	4,25	3,49	3,11	3,38	0,44
	σ	1,11	0,85	0,60	1,01	0,55	0,80	1,22	0,78	0,67	0,31	0,79	

Ø = arithmetischer Mittelwert der 10 Gruppen ohne Gewichtung nach der Anzahl Beobachtungen; σ = Standardabweichung

Bei den 3-Monats-Prognosen ist im Durchschnitt eine Sicherheit von 3,93 angegeben worden. Bei den 12-Monats-Prognosen war diese niedriger (3,38). Während Gruppe 7 während des Planspielverlaufs häufig unterschiedliche Angaben gemacht hat (Standardabweichung 1,4 / 1,22) hat Gruppe 10 deutlich weniger variiert (Standardabweichung von 0,54 / 0,31).



* Angabe auf einer Likert-Skala von 1 bis 6 mit 1=sehr unsicher / 6=sehr sicher

Abbildung 13: Unsicherheit der Gruppen im Planspielverlauf

In Abbildung 13 wird die durchschnittliche Sicherheit der einzelnen Gruppen im Planspielverlauf gemeinsam mit den jeweiligen Trendlinien dargestellt. Der Eindruck aus Tabelle 21 hinsichtlich der niedrigeren Sicherheit bei den 12-Monats-Prognosen bestätigt sich. Im Durchschnitt wird die Sicherheit immer niedriger als bei den 3-Monats-Prognosen angegeben. Die Betrachtung der Trendlinie weist auf eine leichte Steigerung der Sicherheit bei den 12-Monats-Prognosen im Planspielverlauf hin. Bei den 3-Monats-Prognosen zeigt sich hier ein gegenteiliges Ergebnis.

Die grafische Darstellung der durchschnittlich angegebenen Sicherheit der Gruppen lässt einen Zusammenhang zwischen beiden Prognosehorizonten erahnen. In Tabelle 22 sind die Ergebnisse der Berechnung des Rangkorrelationskoeffizienten nach Spearman mit der entsprechenden Irrtumswahrscheinlichkeit der fälschlichen Unterstellung eines Zusammenhangs nach Gruppen dargestellt.

Tabelle 22: Zusammenhang zwischen der Sicherheit bei 3- und 12-Monats-Prognosen

Maß / Gruppe	1	2	3	4	5	6	7	8	9	10	Alle
Korrelation	0,50	0,14	0,26	0,54	0,75	0,50	0,58	0,50	0,62	-0,09	0,53
p-Wert	0,001	0,401	0,128	0	0	0,001	0	0,013	0	0,598	0
# Beobachtungen	39	39	35	39	39	39	39	24	39	38	370

Korrelation = Rangkorrelationskoeffizient nach Spearman (-1 = perfekte negative Korrelation, 0 = keine Korrelation, 1 = perfekte positive Korrelation). p-Wert = Irrtumswahrscheinlichkeit für die unberechtigte Ablehnung der Hypothese, dass keine Korrelation zwischen den beiden Zeitreihen besteht. # = Anzahl.

Demnach ist bei 7 der 10 Gruppen ein signifikanter positiver Zusammenhang zwischen beiden Prognosehorizonten messbar. In der Regel hängt eine höhere (niedrigere) empfundene Sicherheit bei der Abgabe der 3-Monats-Prognose in einer Periode mit einer zeitgleich empfundenen höheren (niedrigeren) Sicherheit bei der Abgabe der 12-Monats-Prognose zusammen.

4.4.2 Prognosedivergenz

Das in der Literatur diskutierte Maß für die Sicherheit der Prognosen berechnet sich aus der Standardabweichung der abgegebenen Prognosen. Dieses Maß ist für die Prognosezeitreihen dieser Studie ungeeignet, da sie, wie in Abschnitt 4.1.1 dargestellt, nicht stationär sind. Die Standardabweichung würde ebenfalls dem zugrunde liegenden Trend folgen, was eine unbegründete Gewichtung der Ergebnisse und somit deren Verfälschung zur Folge hätte. Eine Normierung ist daher unumgänglich. Hierzu wird für jede Spielrunde und beide Prognosehorizonte der Variationskoeffizient v der abgegebenen Prognosen berechnet. Es erfolgt jeweils eine Division der Standardabweichung σ durch den arithmetischen Mittelwert \bar{x} der abgegebenen Prognosen. Die Interpretation erfolgt üblicherweise in Prozent.

$$(14) \quad v = \frac{\sigma}{\bar{x}}$$

In Abbildung 14 sind die Variationskoeffizienten sowohl für die 3- als auch für die 12-Monats-Prognosen abgebildet.

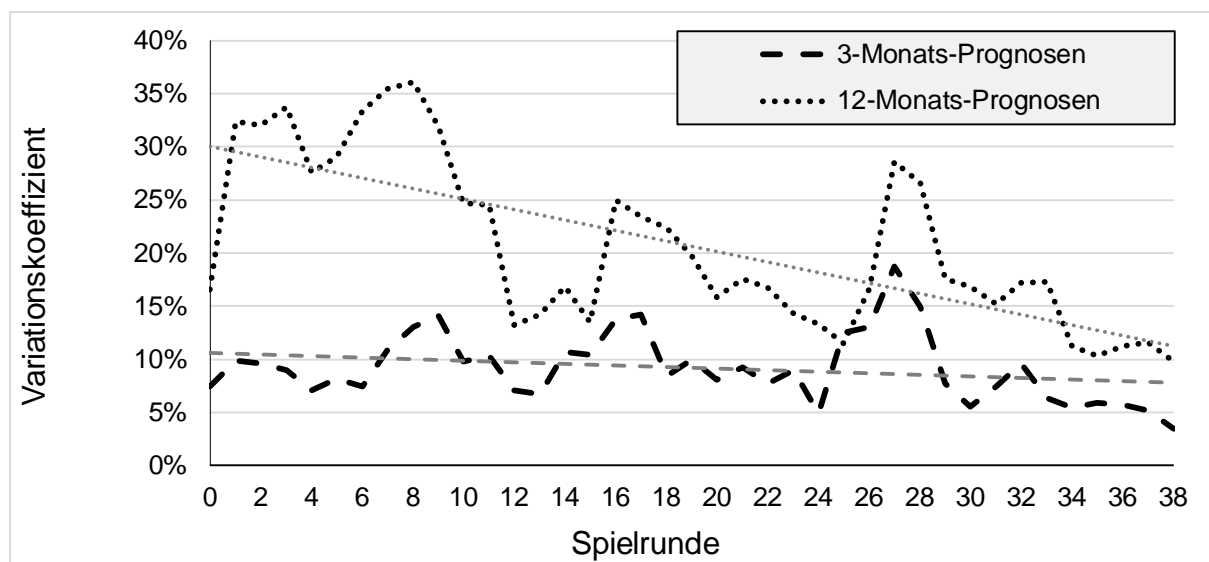


Abbildung 14: Variationskoeffizienten der 3- und 12-Monats-Prognosen im Spielverlauf

Bei der Bestimmung der Einigkeit mittels der Variationskoeffizienten der abgegebenen Prognosen zeigt sich ein deutlicher Unterschied bei den Prognosehorizonten. Bis auf eine Spielrunde

ist der Variationskoeffizient der 12-Monats-Prognosen deutlich höher als der Koeffizient der 3-Monatsprognosen. Im Mittel liegt er doppelt so hoch (20,5% zu 9,2%). Unter starken Schwankungen verringert sich die Prognosedivergenz im Planspielverlauf bei den 12-Monats-Prognosen deutlich und bei den 3-Monats-Prognosen leicht (Steigung der Trendgeraden -0,0213 und -0,0007). Bei einem Korrelationskoeffizienten von 0,52 zeigt sich ein starker und signifikanter Zusammenhang zwischen den Zeitreihen der 3- und der 12-Monats-Prognosen (Testgröße von 3,67; p-Wert 0,0008).

4.4.3 Vergleich der Unsicherheitsmessungen

Anschließend steht die Frage im Vordergrund, ob die in den Fragebögen angegebene Sicherheit mit der Abweichung der Prognosen zwischen den Gruppen zusammenhängt. In Tabelle 23 werden die Korrelationen zwischen den Variationskoeffizienten der Prognosen und der von den Gruppen auf einer Likert-Skala angegebenen mittleren Sicherheit dargestellt. Dabei wird sowohl der übliche Korrelationskoeffizient nach Pearson als auch der Rangkorrelationskoeffizient nach Spearman berechnet.

Tabelle 23: Korrelation zwischen subjektiver Sicherheit und Variationskoeffizient der abgegebenen Prognosen

	3 Monate		12 Monate	
	Pearson	Spearman	Pearson	Spearman
# Beobachtungen	39	39	39	39
Korrelation	-0,57	-0,51	-0,31	-0,36
p-Wert	0,00***	0,00***	0,06*	0,02**

= Anzahl; p-Werte: Irrtumswahrscheinlichkeit für die Gültigkeit der Hypothese, dass ein Zusammenhang besteht (*p<0,1; **p<0,05; ***p<0,01).

In allen Fällen sind die Korrelationskoeffizienten negativ und signifikant. Je unsicherer sich die Gruppen bei der Prognose fühlen, desto stärker weichen die Prognosen der Gruppen voneinander ab. Die in Abschnitt 2.4 aus der Literatur hergeleitete Hypothese („H4: Die subjektiv empfundene Unsicherheit der Gruppen stimmt mit der anhand des Kriteriums Prognosedivergenz objektiv messbaren Unsicherheit überein.“) wird für die Prognosen des Planspiels vorläufig bestätigt.

4.4.4 Zusammenfassung

In jeder Spielrunde des Planspiels geben die Gruppen die bei der Prognoseabgabe empfundene Sicherheit an. Bei den 12-Monats-Prognosen ist diese niedriger als bei den 3-Monats-Prognosen und bei der Mehrzahl der Gruppen besteht ein Zusammenhang zwischen der Sicherheit bei beiden Prognosehorizonten. Im Planspielverlauf steigt die Sicherheit bei den 12-Monats-Prognosen und sinkt bei den 3-Monats-Prognosen. Die in der Literatur vorgeschlagene indirekte Messung der Sicherheit mittels der Prognosedivergenz zeigt sich im Planspiel als signifikant zusammenhängend mit der von den Gruppen subjektiv empfundenen Sicherheit. Die Hypothese („Die subjektiv empfundene Unsicherheit der Gruppen stimmt mit der anhand des Kriteriums Prognosedivergenz objektiv messbaren Unsicherheit überein“) bestätigt sich also zunächst. Je höher die Sicherheit der Gruppen, desto niedriger die Prognosedivergenz und vice versa.

Dies ist bemerkenswert, da die Gruppen ihre Prognosen nicht gegenseitig beobachten. Sie fühlen sich unabhängig voneinander sicherer bzw. unsicherer. Es ist also nicht so, dass die Abweichung von den Prognosen der jeweils anderen Gruppen zur Unsicherheit führt. Dies wäre ein einfaches Erklärungsmodell: Abweichungen von der allgemeinen Einschätzung würde sowohl zu einer höheren Unsicherheit als auch zu einer größeren Prognosedivergenz führen.

Ein anderer Erklärungsansatz ist die Informationsbasis, auf der die Gruppen ihre Einschätzungen aufbauen. So korrespondiert die zunächst deutlich verminderte Sicherheit ab Runde 25 (Abbildung 13) mit einem starken Kursrückgang des Aktienkurses der Alpha-Aktie (Abbildung 5 und Abbildung 6). Die Gruppen fühlten sich dadurch möglicherweise verunsichert und hatten scheinbar unterschiedliche Ansichten, wie sich der starke Kursrückgang auf die neuen Prognosen auswirken soll. Auch die Informationen zur allgemeinen Wirtschaftslage und der Situation des Unternehmens kommen als Ursache in Frage.

4.5 Zusammenfassung

Die im 2. Kapitel aufgestellten Hypothesen und die Ergebnisse sind Tabelle 24 zu entnehmen.

Tabelle 24: Ergebniszusammenfassung

Hypothese	Abschnitt	Ergebnis
<i>H1: Die Prognosen der Teilnehmer schneiden schlechter als empirische Kapitalmarktprognosen von Experten ab.</i>	4.1	Abgelehnt
<i>H2: Die Prognosefehler weisen systematische Komponenten auf und entsprechen somit nicht den Kriterien der rationalen Erwartungsbildung.</i>	4.2	Bestätigt
<i>H3: Die Prognosen weisen gegenwartsorientierte Verlaufsanpassung auf.</i>	4.3	Bestätigt
<i>H4: Die subjektiv empfundene Unsicherheit der Gruppen stimmt mit der anhand des Kriteriums Prognosedivergenz objektiv messbaren Unsicherheit überein.</i>	4.4	Bestätigt

Dass es sich bei den Planspielteilnehmern um Anfänger ohne Prognoseerfahrung handelt, hat die Prognosequalität vermutlich nicht zusätzlich negativ beeinflusst. Die Prognosen sind von schlechter Qualität, jedoch nicht schlechter als die der Experten. Dies führt zur Ablehnung der ersten Hypothese.

Die anderen drei Hypothesen können (vorläufig) bestätigt werden. Die Prognosen weisen systematische Fehler auf, werden sehr nah am jeweiligen Kurs zum Zeitpunkt der Prognoseerstellung abgegeben und die von den Teilnehmern angegebene Sicherheit ihrer Prognosen hängt in Zusammenhang mit der messbaren Prognosedivergenz aller Gruppen.

5 Fazit

Ganz grundsätzlich ist die Prognose von Aktienkursen scheinbar nicht möglich. Nach der Theorie der Informationseffizienz von Fama (1970) wären alle relevanten Informationen immer Bestandteil des jeweils aktuellen Aktienkurses. In diesem Fall wäre die naive Prognose die bestmögliche Strategie. Es herrschen inzwischen jedoch große Zweifel an der effizienten Informationsbeurteilung, da das hierfür notwendige rationale Verhalten der Marktteilnehmer in der Regel ausbleibt.³⁷ Die alte These der Unmöglichkeit von Wirtschaftsprognosen von Morgenstern (1928) bietet weitere theoretische Erklärungsansätze.³⁸ Besonders ernst zu nehmen ist die These der ausbleibenden stabilen Zusammenhänge.³⁹ So könnte sich die Entwicklung in einer endlosen Abfolge kleiner und großer exogener Schocks vollziehen, was eine erfolgreiche Prognose ausschließt.

Kapitalmarktprognosen von Experten weisen demnach wenig überraschend in der Regel eine schlechte Qualität auf. Bei einer differenzierten Betrachtung zeigen sich darüber hinaus bestimmte wiederkehrende Phänomene. Diese Studie zeigt, dass sich die Ergebnisse der Prognoseevaluationen replizieren lassen, auch wenn Anfänger in einem anderen institutionellen Umfeld die Prognosen abgeben.

In der Analyse von Ergebnissen eines Planspiels mit Studierenden wird deutlich, dass die Qualität der Aktienkursprognosen im Planspiel genauso schlecht wie die Prognosen von Experten in verschiedenen Bereichen (Zinsen, Wechselkurse, Aktien) abschneiden. In beiden Fällen sind die Prognosen in der Regel schlechter als die naiven Prognosen. Rationale Erwartungsbildung liegt ebenfalls in beiden Fällen nicht vor, da die Prognosefehler systematische Komponenten aufweisen und nicht zufällig verteilt sind. Einfachere Formen der Erwartungsbildung liegen im Planspiel vor, wobei sich hier ein sehr gemischtes Bild zeigt. Die verschiedenen Gruppen scheitern mit unterschiedlichen Ausprägungen der Erwartungsbildung an der Prognoseaufgabe.

Den größten Zusammenhang weisen die Prognosen in der Regel zum jeweiligen Gegenwartswert zum Zeitpunkt der Prognoseerstellung auf. Diese sogenannte gegenwartsorientierte Verlaufsanpassung haben die Prognosen der Planspielteilnehmer ebenfalls mit den Experten aus

³⁷ Für einen Überblick über die Diskussion zur Hypothese informationseffizienter Märkte siehe Spiwoks (2002b). Fama (1991) zweifelt selbst daran, ob die Hypothese falsifizierbar ist.

³⁸ Zu einer Diskussion der These unter Berücksichtigung der seither getätigten Forschungsanstrengungen vgl. Betz (2004).

³⁹ Vgl. Morgenstern (1928).

dem Feld gemein. Dort ist dieses Phänomen bei entsprechenden Evaluationen ebenfalls mit aller Regelmäßigkeit festgestellt worden. Eine in der Literatur diskutierte Ursache stellt diese Studie in Frage: Das Ausmaß der zukünftigen Änderungen wird von den Planspielteilnehmern recht gut erfasst.

Ein weiteres Ergebnis ist Folge der Frage nach der Sicherheit bei der Prognoseabgabe. Wenn sich die Teilnehmer weniger sicher waren, dann ist auch eine größere Divergenz der Prognosen messbar. Dies ist bemerkenswert, da die Gruppen ihre Prognosen nicht gegenseitig beobachten.

Die Besonderheit des Planspiels liegt in der langen Laufzeit (39 Spielrunden, ca. 13 Stunden über 2½ Monate) und in der umfangreichen Informationsauswertung von anonymisierten Realdaten durch die Teilnehmer. Dies grenzt diese Studie von anderen Forschungsanstrengungen ab und erklärt möglicherweise die realitätsnahen Ergebnisse. In weiteren Forschungsanstrengungen sollte der Frage nachgegangen werden, ob eine etwaige faktische Unmöglichkeit, die Prognoseaufgabe zu lösen, diese Art und Weise des Scheiterns der Prognostiker wahrscheinlich werden lässt.

Literatur

Allen, H. / Taylor, M. P. (1990): „Charts, Noise and Fundamentals in the London Foreign Exchange Market“, *The Economic Journal*, Bd. 100, Nr. 400, S. 49–59.

Andres, P. / Spiwoks, M. (1999): „Prognosequalitätsmatrix - Ein methodologischer Beitrag zur Beurteilung der Güte von Kapitalmarktprognosen“, *Jahrbücher für Nationalökonomie und Statistik*, Bd. 219, Nr. 5, S. 513–542.

Becker, O. / Bolle, F. (1996): „Expectations in economics: rational or not?“, *Jahrbuch Ökonomie und Gesellschaft*, Bd. 13, S. 88–119.

Beckman, S. R. (1992): „The sources of forecast errors: Experimental evidence“, *Journal of Economic Behavior & Organization*, Bd. 19, Nr. 2, S. 237–244.

Bedke, N. (2008): Eine empirische Evaluation von Zinsprognosen sowie experimentelle Evidenz des Reputational Herding, in: Meyer-Bullerdiek, F. / Spiwoks, M. (Herausgeber), *Reihe Bank- und Finanzwirtschaft*, Bd. 5, Frankfurt am Main.

Benke, H. (2006): „Was leisten Kapitalmarktprognosen? Die Sicht eines Stiftungsmanagers, in: Zeitschrift für das gesamte Kreditwesen,“, Zeitschrift für das gesamte Kreditwesen, Nr. 17, S. 40–49.

Betz, G. (2004): „Apriorische und empirische Grenzen von Wirtschaftsprognosen: Oskar Morgenstern nach 70 Jahren Prognoseerfahrung“. In Frank, U. (herausgeber) Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik: Theoriebildung und -bewertung, Ontologien, Wissensmanagement, Wiesbaden, S. 171–190.

Bhargava, A. (1989): „Missing Observations and the Use of the Durbin-Watson Statistic“, Biometrika, Bd. 76, Nr. 4, S. 828–831.

Bofinger, P. / Schmidt, R. (2003): „On the Reliability of Professional Exchange Rate Forecasts: An Empirical Analysis for the €US-\$ Rate“, Financial Markets and Portfolio Management, Bd. 17, Nr. 4, S. 437–449.

Bomberger, W. A. (1996): „Disagreement as a Measure of Uncertainty“, Journal of Money, Credit and Banking, Bd. 28, Nr. 3, S. 381–392.

Brennscheidt, G. (1993): „Predictive Behavior - An Experimental Study“. In Predictive Behavior, Springer Berlin Heidelberg, S. 197–200.

Breusch, T. S. / Pagan, A. R. (1979): „A Simple Test for Heteroscedasticity and Random Coefficient Variation“, Econometrica, Bd. 47, Nr. 5, S. 1287–1294.

Campbell, S. D. / Sharpe, S. A. (2007): „Anchoring Bias in Consensus Forecasts and its Effect on Market Prices“, Finance and Economics Discussion Series, Divisions of Research and Statistics, Federal Reserve Board, Nr. 12.

Carpenter, J. P. / Harrison, G. W. / John A. List (2005): „Field Experiments in Economics: An Introduction“, Research in Experimental Economics, Bd. 10, S. 1–15.

Cheung, Y.-W. / Chinn, M. D. (1998): „Integration, cointegration and the forecast consistency of structural exchange rate models“, Journal of International Money and Finance, Bd. 17, Nr. 5, S. 813–830.

Croushore, D. (1997): „The Livingston Survey: Still useful after all these years“, Federal Reserve Bank of Philadelphia - Business Review, Nr. 2.

- Dickey, A. D. / Fuller, W. A. (1979): „Distribution of the Estimators for Autoregressive Time Series with a Unit Root“, *Journal of the American Statistical Association*, Bd. 74, S. 427–431.
- Diebold, F. X. / Mariano, R. S. (1995): „Comparing Predictive Accuracy“, *Journal of Business & Economic Statistics*, Bd. 13, Nr. 3, S. 253–263.
- Dokko, Y. / Edelstein, R. H. (1989): „How Well Do Economists Forecast Stock Market Prices? A Study of the Livingston Surveys“, *The American Economic Review*, Bd. 79, Nr. 4, S. 865–871.
- Fama, E. (1970): „Efficient Capital Markets, A Review of Theory and Empirical Work“, *Journal of Finance*, Bd. 25, S. 383–417.
- Fama, E. (1991): „Efficient Capital Markets: II“, *The Journal of Finance*, Bd. 46, Nr. 5, S. 1575–1617.
- Figlewski, S. / Wachtel, P. (1981): „The Formation of Inflationary Expectations“, *The review of economics and statistics*, Bd. 63, Nr. 1, S. 1–10.
- Garner, C. A. (1982): „Experimental evidence on the rationality of intuitive forecasters“, *Research in Experimental Economics*, Bd. 2, S. 113–128.
- Giordani, P. / Söderlind, P. (2003): „Inflation forecast uncertainty“, *European Economic Review*, Bd. 47, Nr. 6, S. 1037–1059.
- Granger, C. W. J. / Newbold, P. (1973): „Some comments on the evaluation of economic forecasts.“, *Applied Economics*, Bd. 5, Nr. 1, S. 35.
- Gubaydullina, Z. / Hein, O. / Spiwoks, M. (2011): „The Status Quo Bias of Bond Market Analysts“, *Journal of Applied Finance & Banking*, Bd. 1, Nr. 1, S. 31–51.
- Harrison, G. W. / List, J. A. (2004): „Field Experiments.“, *Journal of Economic Literature*, Bd. 42, Nr. 4, S. 1009–1055.
- Hausman, J. A. (1978): „Specification Tests in Econometrics“, *Econometrica*, Bd. 46, Nr. 6, S. 1251–1271.
- Holden, K. / Peel, D. A. (1990): „On testing for unbiasedness and efficiency of forecasts“, *The Manchester School of Economic and Social Studies*, Bd. 58, Nr. 2, S. 120–127.

Kahneman, D. / Tversky, A. (1979): „Prospect Theory: An Analysis of Decision under Risk“, *Econometrica*, Bd. 47, Nr. 2, S. 263–292.

Keane, M. P. / Runkle, D. E. (1990): „Testing the Rationality of Price Forecasts: New Evidence from Panel Data“, *The American Economic Review*, Bd. 80, Nr. 4, S. 714–735.

Keynes, J. M. (1936): *The General Theory of Employment, Interest and Money*, London.

Lahiri, K. / Sheng, X. (2010): „Measuring Forecast Uncertainty by Disagreement: The Missing Link“, *Journal of applied econometrics*, Bd. 25, Nr. 4, S. 514–538.

Lakonishok, J. (1980): „Stock Market Return Expectations: Some General Properties“, *Journal of Finance*, Bd. 35, Nr. 4, S. 921–931.

Laplace, P.-S. (1814): *Essai philosophique sur les probabilités*, Paris, Corcier.

Leitner, J. / Leopold-Wildburger, U. (2011): „Experiments on forecasting behavior with several sources of information – A review of the literature“, *European Journal of Operational Research*, Bd. 213, Nr. 3, S. 459–469.

Leitner, J. / Schmidt, R. (2006): „A systematic comparison of professional exchange rate forecasts with the judgemental forecasts of novices“, *Central European Journal of Operations Research*, Bd. 14, Nr. 1, S. 87–102.

Levitt, S. D. / List, J. A. (2009): „Field experiments in economics: The past, the present, and the future“, *European Economic Review*, Bd. 53, Nr. 1, S. 1–18.

List, J. A. (2011): „Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off“, *The Journal of Economic Perspectives*, Bd. 25, Nr. 3, S. 3–15.

Lucas, R. E. (1972): „Expectations and the neutrality of money“, *Journal of Economic Theory*, Bd. 4, Nr. 2, S. 103–124.

Manzur, M. (1988): „How Much Are Exchange Rate Forecasts Worth?“, *Australian Journal of Management*, Bd. 13, Nr. 1, S. 93–113.

Mincer, J. / Zarnowitz, V. (1969): „The Evaluation of Economic Forecasts“. In Mincer, J. (herausgeber) *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, New York, S. 14–20.

Morgenstern, O. (1928): *Wirtschaftsprognose. Eine Untersuchung ihrer Möglichkeiten und Voraussetzungen*, Wien.

Mukherji, S. (2011): „Are stock returns still mean-reverting?“, *Review of Financial Economics*, Bd. 20, Nr. 1, S. 22–27.

Muth, J. F. (1961): „Rational Expectations and the Theory of Price Movements“, *Econometrica*, Bd. 29, Nr. 3, S. 315–335.

Nordhaus, W. D. (1987): „Forecasting Efficiency: Concepts and Applications“, *The review of economics and statistics*, Bd. 69, Nr. 4, S. 667–674.

Pearce, D. K. (1984): „An Empirical Analysis of Expected Stock Price Movements“, *Journal of Money, Credit and Banking*, Bd. 16, Nr. 3, S. 317–327.

Samuelson, W. / Zeckhauser, R. (1988): „Status Quo Bias in Decision Making“, *Journal of Risk and Uncertainty*, Bd. 1, Nr. 1, S. 7–59.

Scheier, J. / Spiwoks, M. (2006): *Aktives Portfoliomanagement am britischen Anleihenmarkt*, in: Meyer-Bullerdiek, F. / Spiwoks, M. (Herausgeber), *Reihe Bank- und Finanzwirtschaft*, Bd. 4, Frankfurt am Main.

Schwarzbach, C. / Kunze, F. / Rudschuck, N. / Windels, T. (2012): „Asset management in the German insurance industry: the quality of interest rate forecasts“, *Zeitschrift für die gesamte Versicherungswirtschaft*, Bd. 101, Nr. 5, S. 693–703.

Selten, R. / Abbink, K. / Buchta, J. / Sadrieh, A. (2003): „How to play (3×3)-games.: A strategy method experiment“, *First World Congress of the Game Theory Society*, Bd. 45, Nr. 1, S. 19–37.

Söderlind, P. (2010): „Predicting stock price movements: regressions versus economists“, *Applied Economics Letters*, Bd. 17, Nr. 9, S. 869–874.

Spierdijk, L. / Bikker, J. A. / van den Hoek, P. (2012): „Mean reversion in international stock markets: An empirical analysis of the 20th century.“, *Journal of International Money & Finance*, Bd. 31, Nr. 2, S. 228–249.

Spiwoks, M. (2002a): *Vermögensverwaltung und Kapitalmarktprognose*, in: Meyer-Bullerdiek, F. / Spiwoks, M. (Herausgeber), *Reihe Bank- und Finanzwirtschaft*, Bd. 1.

Spiwoks, M. (2002b): „Ansätze zur Überprüfung der Hypothese Informationseffizienter Kapitalmärkte - Ein Literaturüberblick“, Sofia - Studien zur Institutionenanalyse, Bd. 5, Nr. 2.

Spiwoks, M. (2004a): „Die Verwendbarkeit der ZEW-Aktienindex-Prognosen für aktive Portfoliomanagement-Strategien“, Jahrbücher für Nationalökonomie und Statistik, Bd. 224, Nr. 5, S. 557–578.

Spiwoks, M. (2004b): „External Triggered Herding bei Rentenmarkt-Analysten“, Financial Markets and Portfolio Management, Bd. 18, Nr. 1, S. 58–83.

Spiwoks, M. / Bedke, N. / Hein, O. (2008): „Forecasting the past - The case of U.S. interest rate forecasts“, Financial Markets and Portfolio Management, Bd. 22, S. 357–379.

Spiwoks, M. / Bedke, N. / Hein, O. (2009): „The Pessimism of Swiss Bond Market Analysts and the Limits of the Sign Accuracy Test – An Empirical Investigation of Their Forecasting Success Between 1998 and 2007“, International Bulletin of Business Administration, Bd. 4, S. 6–19.

Spiwoks, M. / Bedke, N. / Hein, O. (2010): „Topically Orientated Trend Adjustment and Autocorrelation of the Residuals - An Empirical Investigation of the Forecasting Behavior of Bond Market Analysts in Germany“, Journal of Money, Investment and Banking, Nr. 14, S. 16–35.

Spiwoks, M. / Hein, O. (2007): „Die Währungs-, Anleihen- und Aktienmarktprognosen des Zentrums für Europäische Wirtschaftsforschung“, AStA Wirtschafts- und Sozialstatistisches Archiv, Bd. 1, Nr. 1, S. 43–52.

Spiwoks, M. / Scheier, J. / Hein, O. (2011): „Zur Beurteilung von Konjunkturprognosen - Eine Auswertung von Prognosen zur Entwicklung des BIP, der Industrieproduktion und der privaten Konsumausgaben in zwölf Industrienationen“, Sofia - Diskussionsbeiträge zur Institutionenanalyse, Bd. 11, Nr. 1.

Svindland, E. (1983): „Konjunkturtheoretische Implikationen der Hypothese rationaler Erwartungen“, Kredit und Kapital, Bd. 16, Nr. 3, S. 331–350.

Theil, H. (1961): Economic Forecasts and Policy, Amsterdam.

Theil, H. (1971): Applied Economic Forecasting, in: Theil, H., Reihe Studies in Mathematical and Managerial Economics, Bd. 4, Amsterdam, North-Holland Publishing Company.

Tversky, A. / Kahneman, D. (1974): „Judgement under Uncertainty: Heuristics and Biases“, Science, Bd. 185, S. 1124–1131.

Webb, R. H. (1987): „The Irrelevance of Tests for Bias in Series of Macroeconomic Forecasts“, Economic Review, Federal Reserve Bank of Richmond, Bd. 73, Nr. 6.

Zarnowitz, V. / Lambros, L. A. (1987): „Consensus and uncertainty in economic prediction“, The journal of political economy, Bd. 95, Nr. 3, S. 591–621.

Zhang, X. F. (2006): „Information Uncertainty and Stock Returns“, Journal of Finance, Bd. 61, Nr. 1, S. 105–136.

Anhang A – Modifizierter Test auf Unverzerrtheit

In Abschnitt 4.2.2 erfolgt eine Durchführung des Tests auf Unverzerrtheit ohne Berücksichtigung der mangelnden Stationarität der Zeitreihen, wie es zuvor in Abschnitt 4.1.1 festgestellt wurde. Unter anderem Webb (1987) weist auf die Problematik der Anwendung dieses Testverfahrens bei nicht-stationären Zeitreihen hin. Daher wird an dieser Stelle für eine höhere Robustheit der Ergebnisse das Testverfahren nicht auf den eigentlich prognostizierten Aktienkurs, sondern auf die relativen Werte nach Bildung des natürlichen Logarithmus⁴⁰ angewendet. t ist dabei immer der Kurs zum Geltungszeitpunkt der Prognose. Im Nenner wird der Kurs A zum Geltungszeitpunkt t um den Prognosehorizont h zurückverschoben. Es ist dann also der Kurs zum Entstehungszeitpunkt der Prognose.

$$(15) \quad \Phi P_t = \ln \frac{P_t}{A_{t-h}} \quad \text{und} \quad \Phi A_t = \ln \frac{A_t}{A_{t-h}}$$

Die Mincer-Zarnowitz-Regression aus (8) wird somit in folgender Variation geschätzt:

$$(16) \quad \Phi A_t = \alpha + \beta \Phi P_t + u_t$$

In Tabelle 25 werden die Ergebnisse des modifizierten Tests auf Unverzerrtheit dargestellt.

⁴⁰ Eine einfache Transformation der Zeitreihen in ihre ersten Differenzen wäre möglich, wobei Theil (1971) dabei möglich Probleme durch Asymmetrie sieht. Daher wird der natürliche Logarithmus auf die relativen Werte angewendet.

Tabelle 25: Ergebnisse des Tests auf Unverzerrtheit der einzelnen Gruppen

Gruppe	3-Monats-Prognosen					12-Monats-Prognosen			
	# Beob.	α	β	D-W-Stat.	p-Wert H: $\alpha=0$ und $\beta=1$	α	β	D-W-Stat.	p-Wert H: $\alpha=0$ und $\beta=1$
1	39	0,05 (0,01)	-0,24 (1,09)	0,77	0,00***	0,18 (0,03)	-0,13 (0,39)	0,76	0,00***
2	39	0,06 (0,01)	-0,45 (0,30)	0,74	0,00***	0,18 (0,03)	-0,04 (0,07)	0,80	0,00***
3	35	0,05 (0,01)	-0,52 (1,06)	0,75	0,01***	0,20 (0,03)	-0,18 (0,23)	0,81	0,00***
4	39	0,05 (0,01)	-0,19 (0,26)	0,71	0,00***	0,22 (0,02)	-0,25 (0,08)	0,81	0,00***
5	39	0,06 (0,02)	-1,29 (0,89)	0,72	0,00***	0,19 (0,03)	-0,24 (0,24)	0,80	0,00***
6	39	0,05 (0,01)	-0,46 (0,49)	0,73	0,00***	0,16 (0,03)	0,14 (0,16)	0,77	0,00***
7	39	0,06 (0,02)	-0,70 (0,43)	0,79	0,00***	0,21 (0,03)	-0,53 (0,28)	0,83	0,00***
8	24	0,07 (0,02)	-0,45 (0,44)	0,79	0,00***	0,13 (0,02)	-0,02 (0,11)	1,08	0,00***
9	39	0,05 (0,02)	0,02 (0,48)	0,77	0,05**	0,20 (0,03)	-0,23 (0,21)	0,80	0,00***
10	38	0,05 (0,01)	-0,24 (0,27)	0,78	0,00***	0,18 (0,03)	-0,03 (0,10)	0,73	0,00***

p-Werte: Irrtumswahrscheinlichkeit für die Gültigkeit der Hypothese, dass die Prognosefehler systematische Komponenten aufweisen (* $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$). Standardfehler in Klammern. Durbin-Watson-Statistik: 0 = perfekt positive Autokorrelation, 2 = keine Autokorrelation, 4 = perfekt negative Autokorrelation. Der kritische Wert (5%-Signifikanzniveau) liegt je nach Zahl der Beobachtungen zwischen 1,45 und 1,54.

Alle untersuchten Prognosezeitreihen weisen auch mit dem modifizierten Testverfahren systematische Prognosefehler im Sinne des Tests auf Unverzerrtheit von Mincer und Zarnowitz (1969) auf. Die Durbin-Watson-Statistik zeigt für alle Prognosezeitreihen eine deutliche positive Autokorrelation erster Ordnung auf. Es sind also ebenfalls systematische Bestandteile in den Residuen messbar. Die Gemeinschaftshypothese einer Konstanten von 0 und eines Koeffizienten von 1 wird auch hier für alle Prognosezeitreihen sicher abgelehnt. Die Koeffizienten β der 3-Monats-Prognosen sind nun auch nahe Null. Die mit dem Standardverfahren ermittelten höheren Koeffizienten scheinen demnach eine Folge der Nicht-Stationarität der zugrundeliegenden Zeitreihen zu sein.

Auch die Panelberechnung wird mit dem modifizierten Verfahren durchgeführt. Die in (16) vorgestellte Schätzung wird dabei um die Gruppendimension i erweitert.

$$(17) \quad \theta A_{t,i} = \alpha + \beta \theta P_{t,i} + u_{t,i}$$

Um die Robustheit der Schätzergebnisse zu überprüfen, wurden verschiedene Konsistenztests durchgeführt. Demnach ist eine Schätzung mit der Standardmethode der kleinsten Quadrate zulässig. Die Ergebnisse dieser Schätzung werden in Tabelle 26 dargestellt.

Tabelle 26: Ergebnisse des Tests auf Unverzerrtheit bei Berechnung auf Panelbasis

Runden	3-Monats-Prognosen					12-Monats-Prognosen			
	# Beob.	α	β	R ²	p-Wert H: $\alpha=0$ und $\beta=1$	α	β	R ²	p-Wert H: $\alpha=0$ und $\beta=1$
Alle (1-39)	370	0,06 (0,00)	-0,31 (0,12)	0,02	0,00***	0,18 (0,04)	-0,09 (0,04)	0,01	0,00***
1-20	190	0,05 (0,00)	-0,30 (0,13)	0,03	0,00***	0,16 (0,01)	-0,06 (0,05)	0,01	0,00***
21-39	180	-0,64 (0,01)	-0,36 (0,19)	0,02	0,00***	0,21 (0,01)	-0,14 (0,05)	0,04	0,00***

R²: Bestimmtheitsmaß der Schätzgleichung; p-Werte: Irrtumswahrscheinlichkeit für die Gültigkeit der Hypothese, dass die Prognosen systematische Verzerrungen aufweisen (*p<0,1; **p<0,05; ***p<0,01). Standardfehler in Klammern.

Die geschätzten Koeffizienten β der Prognosezeitreihen sind in allen Fällen negativ. Auch die Gemeinschaftshypothese $\alpha=0$ und $\beta=1$ kann für alle Schätzungen sicher abgelehnt werden, womit die Prognosen auch in der Panelbetrachtung systematische Fehler aufweisen. Dies gilt für beide Planspielhälften gleichermaßen.

Anhang B – Kursinformationen und Fragebögen

Zu Beginn jeder Spielrunde hat jede Gruppe einen solchen Fragebogen mit dem Kursverlauf der letzten 6 Monate / Spielrunden erhalten. Am Ende der Runde wurde der ausgefüllte Bogen von den Spielleitern eingesammelt.

Anmerkung: Im Rahmen dieser Dissertation wird aus Platzgründen ein verkürzter Anhang präsentiert. Der vollständige Anhang mit allen Fragebögen kann auf der Webseite der Sonderforschungsgruppe Institutionenanalyse heruntergeladen werden:

<http://www.sofia-darmstadt.de/studien20130.0.html>

Übersicht der Zeitachse

Runde/Monat	0	1	2	3	4	5	6	7	8	9	10	11	12
Quartal	0	I	I	I	II	II	II	III	III	III	IV	IV	IV
Monat real	9	10	11	12	1	2	3	4	5	6	7	8	9
Jahr real	85	85	85	85	86	86	86	86	86	86	86	86	86
Letzter Kurs	1.10.	1.11.	1.12.	1.1.	1.2.	1.3.	1.4.	1.5.	1.6.	1.7.	1.8.	1.9.	1.10.
Seite	<i>Download: http://www.sofia-darmstadt.de/studien20130.0.html</i>												

1985-1986

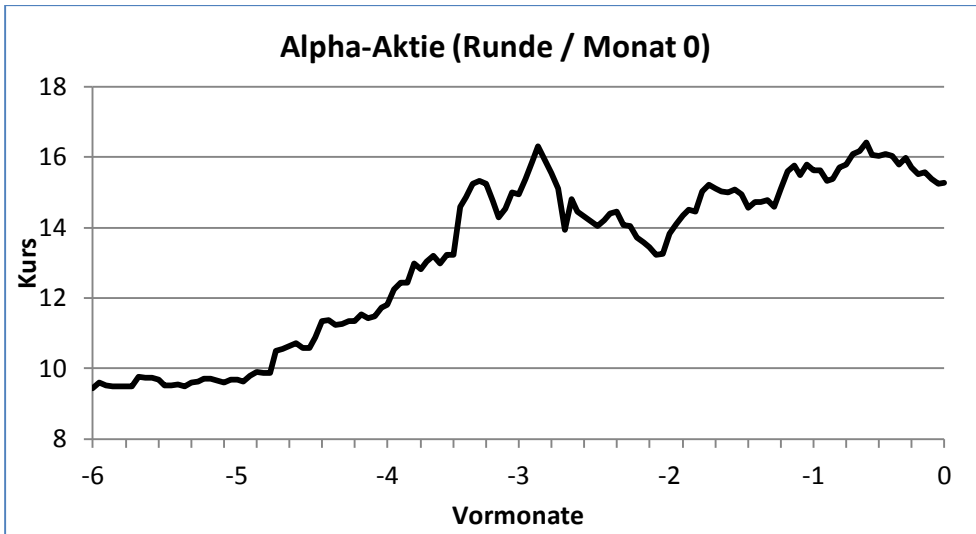
Runde/Monat	13	14	15	16	17	18	19	20	21	22	23	24	25
Quartal	V	V	V	VI	VI	VI	VII	VII	VII	VIII	VIII	VIII	IX
Monat real	10	11	12	1	2	3	4	5	6	7	8	9	10
Jahr real	86	86	86	87	87	87	87	87	87	87	87	87	87
Letzter Kurs	1.11.	1.12.	1.1.	1.2.	1.3.	1.4.	1.5.	1.6.	1.7.	1.8.	1.9.	1.10.	1.11.
Seite	<i>Download: http://www.sofia-darmstadt.de/studien20130.0.html</i>												

1986-1987

Runde/Monat	26	27	28	29	30	31	32	33	34	35	36	37	38
Quartal	IX	IX	X	X	X	XI	XI	XI	XII	XII	XII	XIII	XIII
Monat real	11	12	1	2	3	4	5	6	7	8	9	10	11
Jahr real	87	87	88	88	88	88	88	88	88	88	88	88	88
Letzter Kurs	1.12.	1.1.	1.2.	1.3.	1.4.	1.5.	1.6.	1.7.	1.8.	1.9.	1.10.	1.11.	1.12.
Seite	<i>Download: http://www.sofia-darmstadt.de/studien20130.0.html</i>												

1987-1988

Prognoseformular von Gruppe ____



3-Monats-Prognose

In 3 Spielrunden / Monaten steht der Kurs der Alpha-Aktie bei

Sehr unsicher 1 2 3 4 5 6 Sehr sicher

12-Monats-Prognose

In 12 Spielrunden / Monaten steht der Kurs der Alpha-Aktie bei

Sehr unsicher 1 2 3 4 5 6 Sehr sicher

Anhang C – Pressespiegel und Unternehmenskennzahlen

Zu Beginn jeder Spielrunde hat jede Gruppe zwei Sätze der für diese Runde vorgesehenen Pressespiegel erhalten. Die Informationen durften die Gruppen behalten.

Übersicht der Zeitachse mit Seitenzahlen

Runde/Monat	0	1	2	3	4	5	6	7	8	9	10	11	12
Quartal	0	I	I	I	II	II	II	III	III	III	IV	IV	IV
Monat real	9	10	11	12	1	2	3	4	5	6	7	8	9
Jahr real	85	85	85	85	86	86	86	86	86	86	86	86	86
Letzter Kurs	1.10.	1.11.	1.12.	1.1.	1.2.	1.3.	1.4.	1.5.	1.6.	1.7.	1.8.	1.9.	1.10.
Seite	<i>Download: http://www.sofia-darmstadt.de/studien20130.0.html</i>												

1985-1986

Runde/Monat	13	14	15	16	17	18	19	20	21	22	23	24	25
Quartal	V	V	V	VI	VI	VI	VII	VII	VII	VIII	VIII	VIII	IX
Monat real	10	11	12	1	2	3	4	5	6	7	8	9	10
Jahr real	86	86	86	87	87	87	87	87	87	87	87	87	87
Letzter Kurs	1.11.	1.12.	1.1.	1.2.	1.3.	1.4.	1.5.	1.6.	1.7.	1.8.	1.9.	1.10.	1.11.
Seite	<i>Download: http://www.sofia-darmstadt.de/studien20130.0.html</i>												

1986-1987

Runde/Monat	26	27	28	29	30	31	32	33	34	35	36	37	38
Quartal	IX	IX	X	X	X	XI	XI	XI	XII	XII	XII	XIII	XIII
Monat real	11	12	1	2	3	4	5	6	7	8	9	10	11
Jahr real	87	87	88	88	88	88	88	88	88	88	88	88	88
Letzter Kurs	1.12.	1.1.	1.2.	1.3.	1.4.	1.5.	1.6.	1.7.	1.8.	1.9.	1.10.	1.11.	1.12.
Seite	<i>Download: http://www.sofia-darmstadt.de/studien20130.0.html</i>												

1987-1988

Monat 0 / Quartal 0

Unternehmen Alpha

ALPHA plant große Investitionen.

Mit einem jährlichen Investitionsvolumen von mehr als 50 Millionen € in den beiden kommenden Jahren wird eine Produktionsstätte von Alpha derzeit erweitert und auf eine größere Fertigungstiefe ausgelegt. Eine Vielzahl von Zulieferbetrieben profitiert von diesem großen Arbeitgeber. Die Heimatstadt hat ein Logistikkonzept, das auf eine weitreichende Partnerschaft mit den Lieferanten ausgerichtet ist.

ALPHA in AMERIKA. Zulieferwerk wird geschlossen.

Die Alpha in Amerika will die Produktion von Produkten in einem Werk das 870 Personen beschäftigt in den kommenden zwei Jahren schrittweise einstellen. Man habe auch durch die Auftragsfertigung für andere Hersteller die Kapazitäten nicht befriedigend auslasten können und wolle jetzt versuchen, einen Käufer für das Werk zu finden. Die Fertigung im anderen Werk, das die Produktion einiger bisher in dem zu schließenden Werk gefertigter Teile übernehmen wird, bleibe von der Produktionseinstellung unberührt. Die Tagesproduktion verläuft gegenwärtig im Einschicht-Betrieb. Ungeachtet dessen stieg der Verkauf in den USA in den ersten acht Monaten um 10%.

ALPHA AG / Weitere Samstagsschichten sind geplant. Innerhalb von eineinhalb Jahren mehr als 10 000 Neueinstellungen.

Alles spricht dafür, dass die gegenwärtige Hochkonjunktur für Alpha bis Monat 10 anhält, erklärte der stellvertretende Vorsitzende des Gesamtbetriebsrates der Alpha AG auf einer Betriebsversammlung. Da trotz der Samstagsschichten die Nachfrage gegenwärtig nicht voll befriedigt werden kann, werde der Betriebsrat das vom Vorstand vorgeschlagene Maßnahmenpaket zur Produktionssteigerung ernsthaft prüfen. Im Einzelnen sei neben weiteren Samstagsschichten in diesem Jahr und bis Monat 10 eine Steigerung der Tagesproduktion durch organisatorische Maßnahmen geplant. Allerdings sei auch eine spürbare Personalaufstockung in den Instandhaltungs- und Wartungsbereichen erforderlich. Mit mehr als 100 000 Beschäftigten im Inland hat ALPHA gegenwärtig den höchsten Personalstand seit 12 Jahren, erklärte der Produktionschef. Außer den bisherigen rund 5000 Neueinstellungen in diesem Jahr seien bis zum Q. II weitere rund 2000 Neueinstellungen geplant, davon allein rund 1700 im Stammwerk. Insgesamt betrage die Zahl der Neueinstellungen seit ca. einem Jahr damit rund 10 800. Da es Maschinenengpässe in einigen Stätten gebe, wolle man rund 60 Mio. € investieren, um die Kapazitäten zu erhöhen und die Mehrarbeit abzubauen.

ALPHA AG / Wettbewerbsfähigkeit verbessert. Auf dem europäischen Markt nun an der Spitze.

...

Anmerkung: Im Rahmen dieser Dissertation wird aus Platzgründen ein verkürzter Anhang präsentiert. Der vollständige Anhang mit allen Pressespiegeln und Unternehmenskennzahlen kann auf der Webseite der Sonderforschungsgruppe Institutionenanalyse heruntergeladen werden:

<http://www.sofia-darmstadt.de/studien20130.0.html>