

Geschichte der quantitativen Konjunkturprognose-Evaluation in Deutschland

Von Birger Antholz*

1 Einleitung: Konjunkturprognosen ab 1922 und Prognosenkritik ab 1925

Im Februar 1928 wird vom Institut für Konjunkturforschung in Berlin folgende Konjunkturprognose bekannt gegeben:

„Der Konjunkturhythmus war in Deutschland während der letzten vier Jahre, d. h. im Zeitraum nach der Stabilisierung, gegenüber der Vorkriegszeit durch große Heftigkeit und rasche Aufeinanderfolge der einzelnen Konjunkturwellen gekennzeichnet. Manches spricht aber dafür, daß sich die Konjunkturbewegung in den nächsten Jahren ruhiger gestalten könnte“ (DIW 1928: 10).

Liest sich so die Prognose des Schwarzen Freitags an der New Yorker Börse und des Beginns der Weltwirtschaftskrise nur ein Jahr später 1929? Die große Depression wird weder in Amerika (Dominguez et al. 1988: 605) noch in Deutschland (Krengel 1984: 31) prognostiziert. Früher als gedacht scheint sich das Untersuchungsergebnis von Oskar Morgenstern zu bewahrheiten. In seiner 1928 erscheinenden Habilitation „Wirtschaftsprognose: eine Untersuchung ihrer Voraussetzungen und Möglichkeiten“ kommt er zu dem Resümee:

„Ökonomische Prognose ist, nach unseren Ergebnissen mit den Mitteln der ökonomischen Theorie und Statistik, aus sachlichen Gründen grundsätzlich unmöglich“ (108).

Die „Unmöglichkeit von Wirtschaftsprognosen“ (Tietzel 1981/82) bestimmte daraufhin über Jahrzehnte das Denken der Nationalökonomie in Deutschland.

Die ab 1922 in der Frankfurter Zeitschrift „Wirtschaftskurve“ veröffentlichten Konjunkturprognosen (als Teil der Konjunkturdiagnosen), ähnlich ab 1926 in den Berliner „Vierteljahrsheften zur Konjunkturforschung“, rufen schon ab 1925 eine lebhafte Diskussion über Pro und Contra hervor (Löwe 1925: 374, Meerwarth 1925: 500, Schumpeter 1926: 367, Delbanco 1926, Vogel 1927, Sommer 1928, Vogel 1929, Morgenstern 1930, Vogel 1930). Im Folgenden soll die Entwicklung der Konjunkturprognoseüberprüfungsforschung in Deutschland von diesen Anfängen bis heute vorgestellt werden. Rein theoretische Abhandlungen werden nur am Rande erwähnt, weil es in erster Linie darum gehen soll, alle Prognosen-Soll/Raten-Ist-Überprüfungen zu analysieren, also alle Arbeiten, die empirische Daten benutzen. Internationale Untersuchungen, die OECD-, IMF- oder Europäische Kommissions-Prognosen für mehrere Länder untersuchen und dabei auch Deutschland einbeziehen, bleiben ausgeklammert (z. B. Keereman 1999, Pons 2000, Öller

* Universität Hamburg, E-Mail: birger.antholz@public.uni-hamburg.de

und Barot 2000, Blix et al. 2001, Loungani 2001, Rabitsch 2002). Insgesamt lassen sich für Deutschland ohne die erste, leider makulierte Arbeit, von 1961 bis heute 88 Prognoseüberprüfungsuntersuchungen finden. In einem chronologischen Abriss werden die wichtigsten Ergebnisse dieser Arbeiten vorgestellt. So entsteht ein vollständiger Einblick in Methoden und Entwicklungen der deutschen Konjunkturprognose-Evaluationsforschung und man erhält einen kompletten Überblick über den Wissensstand in diesem Forschungsgebiet.

2 1925 bis 1963: Theoretische Konjunkturprognosekritik

Die ersten Konjunkturprognosen nach dem 2. Weltkrieg werden 1948 von der OEEC (Organisation für wirtschaftliche Zusammenarbeit) für das Jahr 1949 vorgelegt. Sie beziehen sich allerdings nur auf die Bizone. Die erste westdeutsche Wirtschaftsprognose erfolgt vom Bundesministerium für den Marshallplan für das Jahr 1950. Auch für die nächsten Jahre können von der Bundesregierung Wirtschaftswachstumsprognosen mit Zahlenangaben gefunden werden. Da gleichzeitig auch vom Statistischen Bundesamt in der Zeitschrift *Wirtschaft und Statistik* Wirtschaftswachstumsraten bekannt gegeben werden, wäre eine Überprüfung der Konjunkturprognosen ab 1950 möglich. Es gibt aber nur zwei später erscheinende Arbeiten, die die Prognosen der 1950er Jahre überprüfen. Die 1950er Jahre sind eher durch eine Fortsetzung der Morgensternschen theoretisch-logischen Kritik an Wirtschaftsprognosen gekennzeichnet. Bosse betont 1957 (70f), ähnlich wie Bombach 1962, noch einmal die Selbstaufhebung der Prognosen. Die Self-destroying- oder Self-fulfilling-prophecy-Theorien beschäftigten bis in die 1990er Jahre das Denken der Prognoseevaluierer. Morgenstern untermauert seine Prognosenkritik noch einmal 1950 und spieltheoretisch 1963. Erich Preiser prägt in den 1950er Jahren drei Schüler, die bis heute bei der Konjunkturprognosen-Evaluation eine wichtige Rolle spielen: Zarnowitz, Oppenländer und Helmstädter (z.B. Zarnowitz 2005). 1961 gibt es dann eine hochkarätig besetzte Tagung „Diagnose und Prognose als wirtschaftswissenschaftliche Methodenprobleme“. Im dazugehörigen Tagungsband (Giersch und Borchardt 1962) findet sich aber immer noch kein Aufsatz, der einen Soll-Ist-Prognose-Rate-Vergleich durchführt, obwohl seit 1955 mit der ersten Zahlenprognose vom DIW Berlin auch von den Wirtschaftsforschungsinstituten Ganzjahreszahlenprognosen veröffentlicht werden.

3 1967 bis 1969: Anfänge numerischer Prognosenüberprüfung

Die erste Arbeit, die Wirtschaftsprognosen zahlenmäßig überprüft, ist die Diplomarbeit von Kaltefleiter 1961. Leider kann trotz intensivster Recherche kein Exemplar dieser Arbeit mehr gefunden werden, obwohl Kaltefleiter in den 1970er Jahren mit sehr vielen Fernsehauftritten neben Noelle-Neumann zu den bekanntesten (Wahl-)Prognosenbewertern in Deutschland gehört.

Der erste Fachaufsatz, der die Wirtschaftsprognosen überprüft, wird 1967 vom späteren Bundesbankpräsidenten Schlesinger im Allgemeinen Statistischen Archiv veröffentlicht. Er überprüft die Prognosen des interministeriellen Arbeitskreises Gesamtwirtschaftliche Vorausschätzungen der Jahre 1958–1966 auf ihr Eintreffen. Neben dem anschaulichen Prognose-Realisations-Diagramm benutzt er den von Theil 1955 eingeführten und durch seine Bücher ab 1966 verbreiteten Theilschen Ungleichheitskoeffizienten. Das Jahr 1967

markiert überdies nicht nur die Take-off-Phase der Konjunkturüberprüfungsforschung, sondern es begründet auch den Auftakt für die Institutionalisierung der Konjunkturprognosen. Ausgelöst durch die von den Instituten nicht vorausgesagte Rezession 1967 (es tritt ein Rückgang von $-0,5\%$ des BSP ein, prognostiziert werden aber $2\text{--}3,5\%$ Wachstum), fragt der Bundeswirtschaftsminister bei allen Instituten nach Prognosen für das Jahr 1968. Seitdem findet man seit 1968 bis heute von allen Wirtschaftsforschungsinstituten Ganzjahresprognosen, was eine Prognosenüberprüfung leichter macht und damit diesen Forschungszweig erst ermöglicht. 1967 untersuchen Lamberts und Schüssler die Halbjahresprognosen der Gemeinschaftsdiagnose der Institute von 1962–1966 und befinden diese für besser als eine reine Trendextrapolation (296).

Mit dem Artikel von Spahn 1968 startet eine typisch deutsche Eigenschaft der Prognoseüberprüfung. Um ein relatives Prognosemaß zu schaffen, wird auf recht aufwendige Weise ein Trefferbereich konstruiert und dann wird kontrolliert, ob die Prognosen noch innerhalb dieses Trefferintervalls liegen. Vogel und Duelli betrachten 1968 die Prognosen der Gemeinschaftsdiagnose der Institute 1962–67 und untersuchen auf Treffer im Intervall $(-1; 1)$. Sie stellen eine leichte Überschätzungstendenz fest und bezeichnen das Jahr 1967 als Jahr der Fehlprognosen. Im Jahre 1969 legt Wilpert die erste Dissertation zu diesem Thema vor. Er untersucht jedes Prognosejahr von 1950–1965 einzeln und gibt dabei den meist Verbalprognosen für jedes Jahr eine Schulnote von 1–5. Die Institute bekommen von ihm nur eine 4 und er spricht sogar vom Versagen der Gemeinschaftsdiagnose. Der Sachverständigenrat könne in seinen ersten Aussagen auch nicht überzeugen (243). Insgesamt kommt Wilpert zu einer schlechten Einschätzung der Prognoseleistung. Damit zeigen die ersten drei Prognoseüberprüfungsjahre 1967, 1968 und 1969 ein typisches Bild der Evaluationsarbeiten. Die Evaluationsfront zerfällt in zwei Lager: Schlesinger (Mitglied im Arbeitskreis Gesamtwirtschaftliche Vorausschätzungen), Lamberts und Schüssler (RWI Essen), Spahn (DIW Berlin) und Vogel und Duelli (BDI) sind sämtlich Mitglieder von Prognoseorganen und haben damit quasi eine Selbstbewertung vorgenommen. Entsprechend fällt ihr Urteil zur Prognosequalität gut oder zumindest mittelmäßig aus, während die eindeutig schlechte Beurteilung von dem Externen Wilpert stammt.

4 1970er Jahre: Wenige Einzeluntersuchungen

1970 gibt es zwei Untersuchungen von Albeck für 1960–67 und von Kleinewefers 1964–68, die beide nur Halbjahresprognosen untersuchen. Albeck weist die ökonomischen Methoden bei Wendepunkten als überlegen aus (58) und bestimmt so das Denken vieler Prognoseersteller für Jahrzehnte. Anfang der 1970er werden von vielen Instituten wie RWI Essen, DIW Berlin, HWWA Hamburg und auch vom Sachverständigenrat und von der Gemeinschaftsdiagnose der Institute ökonomische Prognosemodelle entwickelt, deren Überlegenheit vor allem in der Anzahl ihrer mathematischen Gleichungen gesehen wird (vgl. Helmstädter 1973: 130). Vajina betrachtet 1972 den Zeitraum 1967 bis 1971, also den Übergang von der Rezession zum Boom und zurück zum gemäßigten Wachstum, und errechnet eine durchschnittliche Abweichung von $2,24\%$ für die Gemeinschaftsdiagnose und von $1,90\%$ für den Sachverständigenrat. Der Jahreswirtschaftsbericht, WSI und IW Köln prognostizierten besser, weil sie zeitlich später ihre Prognose abgaben. Brinkmann, ebenfalls ein Mitarbeiter von IW Köln (BDI), errechnet 1976 für den Prüfzeitraum 1968 bis 1975 eine durchschnittliche absolute Abweichung (MAE) von $2,3\%$ für die Institute (im Folgenden die Kurzform für die Herbst-Gemeinschaftsdiagnose (GD) der Insti-

tute), 2,1 % für den SVR und 1,8 % für den Jahreswirtschaftsbericht. Die Abweichung für die Prognosejahre 1968, 1969 und 1975 hält er für nicht mehr tolerierbar, die Treffsicherheit für die Jahre 1970 bis 1974 stuft er als zufriedenstellend ein (15). Vajina stellt noch einmal 1977 eine Untersuchung vor, die für den Zeitraum 1968 bis 1975 eine durchschnittliche absolute Abweichung (MAE) von 2,15 % für die Institute ausweist. Die Änderungen des Produktionstempos würden meist zutreffend prognostiziert (68). In den 1970er Jahren gibt es also wie in den 1960er Jahren 5 Untersuchungen, davon 3 von Institutsmitarbeitern, die mehrheitlich einen mittleren absoluten Fehler (MAE) von über 2 % ermitteln.

5 1980er Jahre: Prognoseuntersuchungsstandards werden geschaffen

In den 1980er Jahren werden die betrachteten Zeiträume immer länger, was die Maßzahlen automatisch auf eine solidere Basis stellt. Frerichs und Kübler betrachten 1980 in ihrem Lehrbuch den Zeitraum 1963 bis 1979 und kommen zu dem Ergebnis, dass die großen Konjunkturausschläge 1967 und 1975 nicht erkannt werden (253) und der Konjunkturanstieg nach einer Rezession unterschätzt werde. Für sie haben die Konjunkturprognosefehler in den 1970er Jahren zugenommen (254). Heilemann veröffentlicht im Anschluss an seine bei Helmstädter geschriebene Dissertation über die „Prognoseleistung ökonomischer Konjunkturmodelle“ (1981a) im Allgemeinen Statistischen Archiv (Heilemann 1981b) einen Artikel, in dem er das von ihm später noch häufig gebrauchte Prognosemaß RM-SPE (root-mean-square percentage error, 3,7 für den Zeitraum 1969 bis 1977) einführt. Er kommt durch Simulationsrechnungen zu dem Ergebnis, dass Modell-Simulationen mit dem Lüdecke- oder RWI-Modell, allerdings ex post gerechnet, außerhalb des Stützbereichs günstiger als die Gemeinschaftsdiagnose seien (264). Roberts, bis 1978 Mitarbeiter bei ifo München, kommt 1981 zu einer sehr schlechten Bewertung von Konjunkturprognosen. Die BSP-Prognosen des 2. Halbjahres zeugten von keinem verlässlichen prognostischen Vermögen (29). Die Vorhersagen für das 2. Halbjahr 1973, 1974, 1975 und 1977 seien totale Fehlprognosen (27), nur für das Jahr 1973 würde die konjunkturelle Entwicklung in der Jahresprognose getroffen (64). Hätte man statt der hier betrachteten Prognosen von Instituten und Sachverständigenrat nur das 2. Halbjahr des Vorjahres fortgeschrieben, wäre das nicht schlechter (75). Er stellt, wie später viele andere auch, fest, dass die Prognosen nach oben verzerrt seien (82). Mit Roberts tritt ein weiterer Typus des Prognoseevaluierers auf, der zu einer schlechten Prognosebewertung kommt (Roberts meint sogar, dass die Prognosen kläglich versagt hätten (107)). Es handelt sich hierbei um ehemalige Mitarbeiter der Konjunkturprognosen erstellenden Wirtschaftsforschungsinstitute.

Weichhardt legt als Institutsexterner 1982 eine Habilitationsschrift „Zur Beurteilung von Konjunkturprognosen“ vor, die eine Reihe von Erkenntnissen offenbart, die später noch häufiger auftauchen. Die Institute würden eine zu schwache konjunkturelle Reagibilität aufweisen (192). Aufschwünge würden unter- und Abschwünge überzeichnet (196). Die Genauigkeit wäre invers zur Schwankungsintensität. Prognosen könnten deshalb nur beschränkt Grundlage von Konjunkturpolitik sein (202). Weichhardt hält – er betrachtet die Jahre 1963 bis 1975 – die Gemeinschaftsdiagnose und den Sachverständigenrat für recht übereinstimmend (198).

1982 führt Kirchgässner, der ähnlich wie Heilemann bis heute die Prognoseevaluationsdiskussion bestimmt, in seiner Prognosenbetrachtung zwei neue Aspekte ein. Erstens

greift er das von Muth 1961 in Amerika eingeführte Konzept einer Rationalitätsprüfung von Prognosen auf. Die Rationalitätsprüfung erfreut sich wegen ihrer mathematischen Abstraktheit besonders bei Ökonometrikern einiger Beliebtheit, hat aber bis heute gerade in Deutschland viele Kritiker (z.B. Tichy 1994: 247, 254, Helmstädter 1991c: 162, Heilemann 2003: 14). Kirchgässner, der im Verlaufe der nächsten Jahrzehnte regelmäßig zur Qualität von Wirtschaftsprognosen publiziert, ändert im Verlaufe der Zeit auch seine Meinung über die Rationalität von Wirtschaftsprognosen. Zuerst hält er sie z.B. in einer Veröffentlichung von 1984 für streng rational (297 f.), um dann in den 1990er Jahren davon abzurücken und nur noch eine schwache Rationalität zu konstatieren (z.B. 1993: 557). In den 2000ern rückt er auch davon ab und findet Belege, dass die Prognosen noch nicht einmal schwach rational seien (2002: 6). Kirchgässner löst 1982 auch den ersten Streit um Prognosemaßzahlen aus. Neumann und Buscher werfen ihm 1985 mehrfach vor, dass sein Test auf Unverzerrtheit unangebracht erscheine, sie raten sogar von dem Rationalitätstestprüfverfahren von Kirchgässner ab (1985b: 325). Im Gegensatz zu Kirchgässner halten sie die Prognosen auch für nicht rational (1985a: 197). Kirchgässner offenbart in seiner Veröffentlichung 1982 (weitere Aufsätze 1984 und 1985) schon ein vierfaches Dilemma der Rationalitätstest. Rationalitätstest weisen je nach Testverfahren (es gibt viele) unterschiedliche Signifikanzwerte aus. Die Signifikanzgrenzen sind individuell wählbar. Bei den meist nur geringen Stichprobenlängen sind sehr große Abweichungen von der Nullhypothese erforderlich, um signifikante Ergebnisse festzustellen. Für die Prüfung auf strenge Rationalität sind beliebig viele Variablen denkbar. So kommt jeder Rationalitätstest zu einem anderen Ergebnis (was natürlich viele Veröffentlichungsmöglichkeiten schafft). Um es mit Tichy auszudrücken: Diese

„Vorgehensweise [...] erscheint [...] in einer fast verantwortungslosen Weise ad hoc“ (1994: 242).

Zarnowitz zeigt anhand vieler Beispiele die Widersprüchlichkeit der Ergebnisse bei amerikanischen Rationalitätstestsuntersuchungen auf (1992a: 490 f.), womöglich wäre die Angabe nur der marginalen Ablehnungswahrscheinlichkeiten eine Alternative, und lehnt an anderer Stelle Tests auf die Signifikanz von Maßzahlunterschieden ganz ab (1992b: 133).

Neben den umstrittenen Rationalitätstests führt Kirchgässner aber noch einen zweiten Aspekt in die deutsche Prognoseevaluationsforschung ein. 1982 überprüft er auch die Prognosen vom Meinungsforschungsinstitut Allensbach und stellt deren Überlegenheit gegenüber den Instituten oder dem Sachverständigenrat heraus (233). Die Beobachtung, dass die Stimmungsprozente von Allensbach, also der Anteil derjenigen, der bei der seit 1949 im Dezember jährlich durchgeführten Bevölkerungsbefragung (N = 2000) mit Hoffnung in die Zukunft blickt, die konjunkturelle Entwicklung besser als die Wirtschaftsforschungsinstitute anzeigt, wird zuerst 1978 von Steinbuch festgestellt (Noelle-Neumann 1980: 2).

1983 wird von den damaligen IfW Kiel Mitarbeitern Spinanger und Walter das erste Mal eine englischsprachige Evaluation deutscher Wirtschaftsprognosen vorgelegt. Sie betrachten verschiedene Zeiträume, und stellen einen später mehrfach bestätigten Befund fest: Die OECD-Prognosen seien die schlechtesten (18), der heutige Deutsche Bank-Chefvolkswirt Walter errechnet für die Jahre 1969–77 eine durchschnittliche absolute Abweichung von 2,03 %.

1985 legt Ahn eine Dissertation „Konjunkturprognosen im Vergleich“ vor. Darin wird auf 20 Seiten (48–67) ein eigenes, kaum verständliches Prüfmaß entwickelt, so dass man die Bewertung der Institute mit Ergebniswerten aus Normwerten kaum nachvollziehen kann, sogar Korrelationen bezieht er auf Normwerte. Berechnungen gängiger Fehlermaße fehlen. Dafür kommt er auf der Basis einer längeren Zeitreihe, 1963–1979, zu dem Ergebnis, dass abgesehen von der wegen des späten Erscheinungstermins automatisch bevorteilten Frühjahrsschätzung der Gemeinschaftsdiagnose ifo München die besten Prognosen abgäbe (369–371). In seiner Dissertation offenbaren sich aber noch andere Mängel, die für die Prognoseüberprüfung typisch sind. Während ansonsten die Nachvollziehbarkeit und damit der Quellenbeleg zum Standardrepertoire wissenschaftlicher Arbeiten gehören, ist es in der Prognoseevaluationsforschung bis heute Standard, keinerlei Seitenzahlen-Fundorte für überprüfte Prognosen und Raten anzugeben. Sehr oft fehlen die verwendeten Prognosen oder Raten sogar selber. So sind die Berechnungen fast sämtlicher Aufsätze nicht mit dem von den Autoren verwendeten Daten nachrechenbar. Für die Zukunft wäre zu hoffen, dass die Autoren die verwendeten Datensätze der Prognosen und Raten wenigstens ins Internet stellen. Ein weiterer typischer Missstand der Prognosevaliditätsforschung findet sich auch bei Ahn. Er zieht von den Instituten auch Prognoseschätzungen aus dem laufenden Jahr heran. Erst in den 1990er Jahren wird es zum Standard, nur echte Vorjahresprognosen zu überprüfen.

1985 untersuchen die IfW Kiel-Mitarbeiter Langfeldt und Walter die Prognosen von Instituten, Sachverständigenrat, Allensbach, IfW Kiel und Jahreswirtschaftsbericht und kommen zu dem später noch häufiger propagierten Ergebnis: Die „Prognosen sind besser als ihr Ruf“ (70). Sie prägen auch ein bis heute geltendes Standardforschungsergebnis. Das wichtigste Prognosefehlermaß, die durchschnittliche absolute Abweichung, liege nach ihren Ergebnissen bei etwa einem Prozentpunkt und sei damit als gering einzuschätzen (71). Sie kommen zu ihrem Ergebnis aber durch einen, gerade von Institutsmitarbeitern, insbesondere vom IfW Kiel, noch oft praktizierten Kunstgriff. Sie klammern die Fehlprognosen 1975 und 1967–1969 einfach aus, indem sie den Untersuchungszeitraum 1976 beginnen lassen (hier 1976–1984). Langfeldt legt 1986 (zusammen mit Trapp) und 1988 (ebenfalls mit Trapp) zwei Untersuchungen vor, die die gleichen Prognoseersteller umfassen, und ähnlich die Zeit der Fehlprognosen vor 1976 ausklammern. Entsprechend kommen sie zu Resümees, wie

„die professionellen Konjunkturprognosen haben sich dabei als relativ zuverlässige Informationsquelle bewährt“ (1986: 33)

oder die mittlere Abweichung (ME) hardly shows „any systematic bias“ (1988: 433). Sie ermitteln die gute Qualität der Konjunkturprognosen aber vor allem im Vergleich zur naiven Prognose. Dies ist jedoch nicht besonders redlich, weil die naive Fortschreibung direkt nach einer starken konjunkturellen Richtungsänderung zu starken absoluten Prognosefehlern führt. In diesem Untersuchungszeitraum erscheint eine einfache naive Fortschreibung des Rezessionsjahres 1975 (–3,6 %) als Prognose für 1976 auch unangebracht, weil sich schon seit November 1975 der massive Aufschwung zeigt. Ebenso ist schon im 2. Halbjahr 1980 eine Rezession gut sichtbar, so dass man für 1981 nicht die 1,8 % Wachstum aus 1980 fortgeschrieben hätte.

Pflaumer (1986) beziehungsweise Pflaumer und Swart (1987) prägen in ihrer Untersuchung das sich lange Zeit in den Köpfen der Evaluierer festsetzende Bild, dass der Sachverständigenrat am besten prognostiziere. Dabei hätte man aber berücksichtigen müssen, dass dieses Ergebnis nur durch den Vergleich mit einer naiven (1986: 381) und der Institutsprognose (1987: 357) zustande kommt. Cornelius festigt 1989 dieses Bild, dass der Sachverständigenrat am besten prognostiziere. Dies ist allerdings nicht verwunderlich, denn Cornelius ist Mitarbeiter des Sachverständigenrats. Zu seinem Ergebnis kommt er wiederum nur durch den Vergleich mit noch schwächeren Prognostizierern, wie der Gemeinschaftsdiagnose, dem IMF und der OECD. Cornelius stellt für den Zeitraum 1973–1987 im Übrigen für die Gemeinschaftsdiagnose eine mittlere absolute Abweichung von 1,47 % fest. Besonders große Abweichungen träten bei den großen konjunkturellen Schwankungen 1973/74 und 1981/82 auf. Seit 1983, also seit den 1980er Jahren, gäbe es eine höhere Treffergenauigkeit als in den 1970er Jahren (46).

1989 legt Stege an der Universität Hamburg eine Dissertation vor, die wie eigentlich alle von Institutsexternen geschriebenen Dissertationen (Heilemann war zum Zeitpunkt seiner Dissertation schon Mitarbeiter vom RWI Essen) zu einer schlechten Prognosebeurteilung kommt. Die Institute prognostizierten antizyklisch und intentional, so sein Ergebnis. Es gäbe viele Wendepunktfehler. Stege zieht recht umfangreiches Datenmaterial heran. So wertet er die Gemeinschaftsdiagnose der Institute, den Sachverständigenrat, DIW Berlin, RWI Essen, HWWA Hamburg, IfW Kiel und ifo München aus und betrachtet den Zeitraum 1966 bis 1985. Bei ihm liegt wie auch schon bei der vorherigen Dissertation von Ahn ifo München als bester Prognostizierer vorne.

Ende 1989 veröffentlicht Helmstädter das erste Mal eine Überprüfung der Prognosen des Sachverständigenrates, dessen Mitglied er von 1983–1988 war, und der Gemeinschaftsdiagnose der Institute für den Zeitraum 1978–1987. Im Gegensatz zu Kirchgässner kommt er durch graphische Analyse zu dem Ergebnis, dass die Prognosen autokorreliert seien, sogar streng in Bezug auf das Vorzeichen (163).

6 1990er Jahre: Viele solide Überprüfungsarbeiten

Stege schafft es als einer der ganz wenigen Prognoseevaluierer durch einen Artikel in einer Publikumszeitschrift, nämlich dem Manager Magazin, seine Prognoseüberprüfungsergebnisse einem breiten Publikum zu präsentieren. Dabei wird auf Basis der Daten seiner Dissertation aus 50 % Theilscher Ungleichheitskoeffizient und 50 % Wendepunkterkennung eine Schulnote für jedes Institut berechnet. Die GD der Institute bekommt eine 4+, der SVR eine 4 und die anderen Institute Noten von 3+ bis 4+. Neben den altbekannten Feststellungen, dass Aufschwungphasen unter- und Abschwungphasen überschätzt werden (184), wird noch eine Begründung für die schlechte Prognosequalität angeführt. Das Prognosekartell/Meinungsmonopol bewirke, dass die Prognosen untereinander weniger als von der Realität abweichen (180). Stege befindet im Manager Magazin-Artikel Allensbach für das bestprognostizierendste Organ, ohne es allerdings voll in die Auswertung einzubeziehen (187).

1991 betrachtet der damalige Generalsekretär des Sachverständigenrats Heise die Prognosequalität für die Jahre 1986 bis 1989 und kommt zu dem Ergebnis, dass der Rat mehr Erklärungsgehalt gegenüber einer No-change-Prognose besitze (33). Harvey stellt im glei-

chen Jahr beim Vergleich von Institutsprognosen und einem der wichtigsten Konjunkturindikatoren (langfristiger – kurzfristiger Zinssatz) für den Zeitraum 1983–1989 fest: „The term structure forecasts are much more accurate than the forecasts being used” (711).

1991 untersucht Helmstädter die Jahre 1979–90 und konstatiert, dass der Allensbacher Stimmung Indikator den Verlauf der Konjunktur besser prognostiziert als der Sachverständigenrat und die Institute. Der Allensbacher Stimmung Indikator wäre besser als die Prognostikerprofession. Dies fordert eine Replik vom damaligen Generalsekretär des Sachverständigenrats (SVR), van Suntum, heraus. Van Suntum untersucht die gleichen Daten wie Helmstädter, behauptet aber, obwohl er mit 0,98 % gegenüber 1,04 % durchschnittlicher absoluter Abweichung eine geringere MAE für Allensbach berechnet und auch die von ihm angegebene Verzerrung von Allensbach kleiner ist, dass Allensbach keine Vorteile gegenüber dem Sachverständigenrat besäße. Man dürfe im Übrigen den Stützzeitraum zur Umrechnung der Allensbacher Stimmungsprozente in konkrete Prognosewerte nur bis zum Vorjahr wählen. Man könne schon alleine deswegen kein Prognoseversagen konstatieren, weil 99 % der absoluten Höhe des BSP richtig vorhergesagt werde. Eine Replik erntet Helmstädter auch von Kirchgässner. Unter Heranziehung der Jahre 1966–90 findet Kirchgässner 1991 weiterhin keine positive Autokorrelation der Prognosen, weil der Durbin-Watson-Test weit von der Signifikanzgrenze entfernt sei (156). Helmstädter erwidert 1991, dass schon alleine die Methode des genauen Hinsehens bei der graphischen Analyse offenbare, dass sich die Autokorrelation in den letzten 3 Jahren noch verschlimmert hätte. Eindeutig zeige sich in den letzten Jahren, dass die Prognosefehler systematisch im Zusammenhang mit der Konjunkturphase stünden (= Autokorrelation): bei Abschwüngen entstünden Überschätzungsfehler und bei Aufschwungphasen Unterschätzungsfehler (1991b: 161). In einer weiteren Veröffentlichung 1992 bekräftigt Helmstädter seine Kritik an dem Prognoseverhalten der Prognoseersteller. In jeder Phase würden immer wieder die gleichen Fehler gemacht. Charakteristisch sei die Konjunkturphasen verzögerte Reaktion der Prognostiker. Die Prognostiker seien nicht konjunkturbeusst (122). Allensbach prognostiziert besser (125). Helmstädter führt, um der Kritik an fehlender Signifikanz etwas entgegenzusetzen, die Berechnung mit Dummy-Variablen ein (1991b: 160 f.). Da die Wahl der Dummies frei ist, erscheint dieser Ansatz etwas angreifbar und wird später auch fast nicht mehr von anderen aufgegriffen. Der Helmstädter-Kirchgässner-Disput über die Autokorrelation von Konjunkturprognosen wäre aber hin-fällig, würde man das Prüfinstrument in Frage stellen. Der Durbin-Watson-Test ist zwar der gängige Test auf Autokorrelation, aber seine Teststärke (power) ist einfach zu klein (d.h. es wird ein zu hoher β -Fehler in Kauf genommen). Eventuell ist der Breusch-Godfrey-Test eine Alternative. Kirchgässner findet 1993 im Gegenzug auch ein anderes Verfahren, um Konjunkturprognosen zu überprüfen. Er definiert die gängigen Prognosen der Institute als Zwei-Schritt-Prognosen. Die Prognostiker hätten zuerst den Ausgang des aktuellen Jahres vorauszusagen (erster Schritt), um aufbauend auf dieser Unsicherheit erst im zweiten Schritt das nächste Jahr prognostizieren zu können. Mit dieser Unterscheidung in Ein- und Zwei-Schritt-Prognosen kommt Kirchgässner zu überraschenden Ergebnissen. Im Gegensatz zu seinen bisherigen Arbeiten sieht er Anzeichen für Autokorrelation und rückt von seinem Rationalitätsaxiom ab:

„This could indicate that the first-order autocorrelation in the two-step prediction errors is higher than would be consistent with rational expectations” (557).

Im Jahre 1993 erscheinen auch noch drei andere Prognoseüberprüfungen, die nachvollziehbarere Prognosemaße berechnen. Die in bisherigen Untersuchungen meist als schlechtestester Prognostizierer eingestufte OECD präsentiert eine Untersuchung der Jahre 1987–92, in der sie zwar rechnerisch beim mittleren Fehler (Verzerrung) wie auch beim RSME schlechtere Werte für die OECD ausweist, allerdings verbal zu dem Schluss kommt:

„OECD´s projection record [...] has been as good as and perhaps even slightly better than those made by the IMF or national authorities“ (53).

Die Inkonsistenz von berechneten Ergebnissen und verbaler Interpretation wird allerdings nicht zu einem typischen Merkmal der Prognoseevaluationsforschung, obwohl sie vorkommt. Der RWI Essen-Mitarbeiter Filusch untersucht 1993 die Prognosejahre 1964–1990 der Gemeinschaftsdiagnose mit einem einleuchtenden Prognosemaß. In den betrachteten 27 Prognosejahren sei die Abweichung 10-mal 1–2 %, 3-mal 2–3 %, 1-mal 4–5 % und 1-mal 5–6 % gewesen (13). Büttner (DIW-Praktikant) und Horn (DIW-Mitarbeiter) kommen 1993 bei Betrachtung der Prognosen 1970–92 der Institute und des DIW Berlin zu dem Schluss, dass die Prognosen von GD und DIW bis zu einem halben Jahr relativ verlässlich seien (16). Tichy widmet in seinem weit verbreiteten Lehrbuch der Treffsicherheit von Konjunkturprognosen ein ganzes Kapitel (235–259). Er findet, die SVR-Prognosen 1968–1992 graphisch analysierend, die Prognosereihe für glatter verlaufend als die der Realisationen, kommt aber nach einer Wendepunkt Betrachtung zu einer positiven Gesamtbilanz (1994: 256 f.). Erst 1995 erscheint wieder eine kritische Arbeit. Wörz kommt in ihrer Diplomarbeit an der Universität Innsbruck zu dem Schluss, dass man nicht behaupten könne, dass das BIP-Wachstum zufriedenstellend erfasst worden wäre. Die Prognosefehler betragen über 70 % der eingetretenen Werte (161). Wörz untersucht ifo München und OECD für den Zeitraum 1974–1991 und befindet ifo für weitaus besser.

Vom IfW Kiel erscheint 1995 ein später oft zitiertes Diskussionspapier, in dem Döpke und Langfeldt die Institute, den Sachverständigenrat und IfW Kiel für den Zeitraum 1976–1994 untersuchen und 15 Prognosemaße berechnen. Der MAE liege bei 1,02 % bei SVR und bei 0,99 % beim IfW Kiel (4). Der Prognosefehler der naiven Prognose wäre doppelt so hoch. Es bestünde Bedarf, Prognosen an den konjunkturellen Wendepunkten zu verbessern (8). Eine weitere solide Arbeit wird 1996 von Hinze in den mittlerweile eingestellten Hamburger Jahrbüchern veröffentlicht. Für den Zeitraum 1970–1995 findet er zwischen den Instituten, SVR, Jahreswirtschaftsbericht und HWWA Hamburg, dessen Mitarbeiter er ist, keine gravierenden Unterschiede (70). Bei der Berechnung des MAE stellt er für die 1990er leicht höhere Abweichungen verglichen mit den 1980er Jahren fest. Die Begründung wären vermehrt auftretende exogene Faktoren, auch wegen der deutschen Einheit, und damit wäre die Prognoseerstellung schwieriger geworden. (64, 81). Oppenländer sieht, die Institute 1964–1990 betrachtend, für die Jahre 1967, 1975, 1982 und 1993 Wendepunktfehler (1996: 574). Langfeldt (IfW Kiel) verlängert seinen 1980er Untersuchungsdatensatz und kommt 1996, weil er wiederum die Zeit vor 1976 ausklammert, auf das bekannte Resümee, dass die Prognosen besser als ihr Ruf seien. In anderen Ländern wären die Prognosefehler ähnlich hoch (181). 1996 führt Helmstädter in einer Untersuchung als Begründung für den von ihm wiederum festgestellten Befund, dass die Konjunkturausschläge von den SVR-Prognosen nur gedämpft wiedergegeben werden, das Gruppenverhalten an (165f). Winkler widmet in seinem Lehrbuch 1997 der Bewertung der Prognosegüte ein ganzes Unterkapitel (241–246) und gibt Zahlen aus anderer Quelle wieder, die für

den Zeitraum 1976–1994 beim Sachverständigenrat 1,02 % und bei den Instituten 1,18 % als durchschnittliche absolute Abweichung errechnet haben (244).

Kirchgässner und Hagen weisen 1997 mit einem ähnlichen Datensatz wie bei vergangenen Untersuchungen das Ergebnis von Harvey (1991) zurück. Sie halten die Instituts- und SVR-Prognosen einer zinssatzbasierten Prognose für überlegen (5). Ihre weiteren Ergebnisse lauten:

„There are dramatic differences between the results of within-sample and out-of-sample tests. While both procedures provide evidence for inefficiency, this evidence seems to be much stronger in within-sample than out-of-sample tests“ (6).

Rinne bespricht die Jahre 1975–1994 unter Benutzung von 17 Prognosefehlermaßen 1997. Sein MAE für den untersuchten Sachverständigenrat liegt bei 1,21 % und er konstatiert, dass der SVR besser als die naive Prognose prognostiziert (332). Gleichwohl sieht er auch eine Fehlleistung bei der Prognose der Konjunktur (333).

1998 tritt Heilemann, bis 2004 noch Mitarbeiter beim RWI Essen, wieder als Prognose-evaluierer auf. Für den von ihm betrachteten Zeitraum 1966–95 kommt er auf folgende mittlere absolute Abweichungen: Institute 1,6 %, SVR 1,4 %, Jahreswirtschaftsbericht 1,3 %, OECD 1,4 %, RWI Essen 1,1 %. Für RWI Essen zieht er allerdings nur Daten von 1980–1995 heran. Aus dieser Konstellation heraus kann sein Fazit über die RWI-Prognosen nicht überraschen: „So mag die Prognosegenauigkeit, verglichen z.B. mit der von GD oder Sachverständigenrat, wenig Anlass zu besonderer Kritik zu geben“ (1998a: 92). In einer weiteren Veröffentlichung 1998 mit ungefähr dem gleichen Datensatz kommt er zu dem Resultat, dass die Prognosen verglichen mit einer naiven No-change-Prognose vergleichsweise gut abschnitten. Er hält alle Prognosen für frei von Verzerrungen (1998b: 90, 95). 1998 leitet Smolny den Rückzug vom Glauben an die ökonomischen Prognosemodelle ein, indem er, den Zeitraum 1970–96 betrachtend, die iterativ erstellten Vorhersagen von SVR und Gemeinschaftsdiagnose für den ökonomischen Methoden überlegen einstuft (17). Der ifo München Mitarbeiter Nierhaus bespricht 1998 die Prognosejahre 1991–1997 einzeln und sieht am Ende keine schlechte Bilanz. Zusammenfassend sei festzuhalten, dass die Prognosen den Wirtschaftsverlauf nicht schlecht abgebildet hätten (R 9). Heilemann examiniert 1999 die ostdeutschen Konjunkturprognosen 1991–1998 und kommt zum Ergebnis einer hohen relativen Treffgenauigkeit (76).

7 2000er Jahre: Vollständigere Datensätze, verfeinerte Methoden, deutlichere Prognosenkritik

Döpke (IfW Kiel) untersucht 2000 die politische Verzerrung von Prognosen und präsentiert ein schwer zu interpretierendes Ergebnis: Nach nichtparametrischen Tests wären keine politischen Einflüsse feststellbar (607), nach parametrischen Tests gäbe es Anzeichen für politische Intentionen (609). Antholz analysiert 2000 in seiner Hamburger Magisterarbeit 248 Wirtschaftsprognosen, die auf der Rückseite der IW-Trends für den Zeitraum 1974–1999 ausgewiesen werden. Nach seiner Auswertung der Abweichungsverteilung seien die Prognosen signifikant zu positiv (63 % Überschätzungsprozentpunkten stünden 37 % Unterschätzungsprozentpunkte gegenüber). Würde man wie oft in den Wirtschafts- und Sozialwissenschaften ein 95 %-Konfidenzintervall zu einem Prognosewert angeben,

hätte es eine Spannweite von $\pm 4,3\%$ (64). Kirchgässner und Savoiz (2001), gegenüber 1997 jetzt mit 1975–1989 einen sehr viel eingeschränkteren Prüfzeitraum benutzend, kommen zu dem Ergebnis, dass Institutsprognosen verglichen mit Prognosen, die auf der Zinsspanne basieren, effizient sind (360), nicht jedoch verglichen mit solchen, die entweder auf kurzfristigen Zinssätzen oder auf der realen Geldmenge M1 basieren (361). Grömling, IW Köln-Mitarbeiter, untersucht in zwei Veröffentlichungen 2002 die Jahre 1995–2001 und befindet für die Jahre 1997, 1998, 1999 eine gute Trefferqualität. In den übrigen Jahren 1995, 1996, 2001 wäre merklich abgewichen worden (2002a: 20). Außerdem stellt er fest, die Prognosen „show optimistic tendencies“ and „were smoother than reality“ (2002b: 247). Vom IW Köln gibt es ebenfalls 2002 eine Veröffentlichung mit dem gleichen Datensatz, in dem das IW Köln für genauso gut wie die Institute oder der SVR befunden werden. Ein Mitarbeiter eines anderen Wirtschaftsforschungsinstituts, Radowski vom ZEW in Mannheim, untersucht 2002 den Zeitraum 1980–2000 für eine ganze Reihe von Prognoseerstellern. Sein Gesamtergebnis ist, dass die Prognosen deutlich geringere Fehler als die naive Prognose und keine systematischen Fehler aufwiesen (12). Mehrmals wären Wendepunkte zutreffend angekündigt worden (11). Genau gegenteilig meinen Rine und Specht für die Jahre 1975–99, dass „die Wendepunkte durch den SVR denkbar schlecht prognostiziert wurden“ (2002: 134). Kirchgässner und Müller finden 2002 leichte Beweise, „that forecasters are reluctant to revise their predictions“ (14). Im Jahre 2002 kommt es in der Zeitschrift „Wirtschaftsdienst“ zu einem Disput über die Verbesserung der Qualität von Wirtschaftsprognosen im Zeitablauf zwischen den dies verneinenden IfW Kiel Mitarbeitern Dicke und Glismann und dem Generalsekretär des Sachverständigenrats Weidmann. Heilemann und Stekler schlagen für diese Fragestellung den COSUM-Test vor (2003: 16). Antzoulatos und Wilfing kommen 2003 für die OECD-Prognosen von 1967–2000 zu dem Ergebnis: „GDP-growth forecasts are less efficient than the consumption-growth ones, but the inefficiency is weak“ (11). Da sie für ihre Untersuchung einen eigenkonstruierten SURGE-Maßstab verwenden, kann man ihre Ergebnisfindung kaum verstehen.

Besser verständlich sind fünf Veröffentlichungen von Heilemann, der, immer auf seinen bekannten Prognosesatz zurückgreifend, gegen Ende seiner Zeit beim RWI Essen (ab 2004 Professor in Leipzig) die Prognoseleistung zunehmend kritischer bewertet. So konstatiert er eine „strong tendency to turning point errors“ (2003: 12). Die Prognoseersteller „have seen none of the recession“ (12). Mit dem späten Erkennen der Abschwünge blieben die Prognosen hinter den Erfordernissen einer antizyklischen Politik zurück (2004b: 190). In der oben erwähnten die Prognosen eher positiv bewertenden Arbeit mit Stekler resümiert er 2003: Der MAE „was 40% of the mean absolute change“ (8). In den Jahrbüchern für Nationalökonomie und Statistik stellt er 2004 aufgrund modelltheoretischer Überlegungen eine Obergrenze der Treffsicherheit von Prognosen fest, die bei 0,7 vH- bis 1,0 vH-Punkten liege (58). In einer weiteren Veröffentlichung 2004 zusammen mit seiner Mitarbeiterin Blaschzik setzt er den MAE ins Verhältnis zur durchschnittlichen Ratenhöhe und kommt zum Ergebnis, dass sich die Prognosen von den 1970er bis zu den 1990er-Prognosen verschlechtert hätten (9). Ein 50 %-Vertrauensintervall betrüge ($\pm 0,9\%$).

Döpke (vorher IfW Kiel, jetzt Deutsche Bundesbank) und Fritsche (DIW Berlin) legen 2004 ein Diskussionspapier vor, das auf der bisher größten Datengrundlage basiert. Für 14 Prognoseersteller und die Jahre 1970–2003 berechnen sie 13 Prognosemaße. Trotz einer

Reihe von Unstimmigkeiten in den von ihnen lobenswerterweise angegebenen Raten und Prognosen lassen sich die von ihnen berechneten MAE, ME und RMSE recht gut reproduzieren. Als Ergebnisse werden von ihnen genannt: Prognosen seien unverzerrt und schwach rational. Die Hypothese starker Rationalität müsse aber zurückgewiesen werden (22). Weitere Ergebnisse werden weiter unten präsentiert. Dicke (IfW Kiel) und Glismann (ehemals IfW Kiel) vergleichen 2004 den Durchschnitt der Prognosen der Wirtschaftsforschungsinstitute mit dem Durchschnitt der Prognosen von fünf Banken für den Zeitraum 1989–2002. Sie stellen fest, dass die staatlichen den privaten Prognosen unterlegen seien (8) und stellen schlussfolgernd die Sinnfrage staatlicher Konjunkturforschung in Deutschland (9). Betz, der allerdings nur eine reine Sekundäranalyse von einigen Evaluationsarbeiten durchführt, findet 2004 genügend Belege für die Beobachtung, dass Konjunkturprognosen kaum besser als Keine-Änderungs-Prognosen abschnitten (187). Ähnlich wie Dicke und Glismann ist sein Abschlussstatement:

„Und wenn ganz einfache Regeln genauso gut abschneiden wie professionelle, hoch komplizierte und kostenreiche Prognoseverfahren, so gibt es keinen Grund, an letzteren festzuhalten“ (188).

Döpke veröffentlicht in *Wirtschaftswissenschaftliches Studium (WiSt)* einen Lehrbeitrag zur Messung der Qualität von Konjunkturprognosen. Bedauerlicherweise leitet er einen Vorzeichenwechsel beim Verzerrungsmaß ME ein: Überschätzungen, jahrzehntelang mit einem positiven Vorzeichen ausgewiesen, erhalten nun ein negatives. (2004: 8). Er schlägt den Wilcoxon-Rangsummen-Test zur Prüfung von Autokorrelation vor, der aber ähnlich wie der Durbin-Watson-Test keine Autokorrelation der Konjunkturprognosen erkennen lasse (10). Eine Kombination von Prognosen könne sinnvoll sein (12). Zur Prognosequalitätsmessung fordert er den verstärkten Einsatz von Testverfahren (u. a. Forecast-encompassing-Test, Diebold-Mariano-Test (13)).

Grömling (IW Köln) ergänzt 2005 seine vorherigen Aufsätze um die neuen Prognosen und betont die Datenrevisionen als Ursache von Prognosefehlern (11f). Hinze (HWWA Hamburg, 2005), ebenfalls wie Grömling die Jahre 1995–2004 berechnend und das Revisionsproblem anführend, findet, dass der mittlere absolute Fehler mit 0,6 % nicht weit von der mittleren absoluten Ratenrevision mit 0,4 % entfernt sei (119). In seiner von Helmstädter betreuten Dissertation kommt Anholz 2005 unter Auswertung aller Zahlenprognosen von 1950–2004 und der Einbeziehung von 13 Prognoseerstellern (N = 486 Prognosen) zu folgenden Ergebnissen: MAE 1,26 % und ME +0,10 % Überschätzung. Umfragebasierte Prognosemethoden lieferten die besten Ergebnisse, während ökonomische und aufwendige iterative gleich, aber schlechter abschnitten. Die ostdeutschen BIP-Konjunkturprognosen 1992–2004 seien ungenau (MAE 1,83 %) und massive Überschätzungen (ME +1,23 %). In jedem Konjunkturzyklus würden seit 1950 immer wieder die gleichen Prognosefehler gemacht: Der Rezessionsbeginn werde nicht erkannt, es werde fälschlich ein Sofortaufschwung vorhergesagt und die Rezessionsdauer und -tiefe würden unterschätzt. Der erste Wachstumsschub werde für niedriger prognostiziert als er sei, die Tiefe der dann folgenden Wachstumsdelle werde nicht vorhergesagt und die Höhe des zweiten Teil des Aufschwungs unterschätzt. Kirchgässner führt 2005 unter Benützung seines speziellen Überprüfungsmethodensets einen Vergleich der Prognosegüte von Allensbach, Instituten und Sachverständigenrat für den Zeitraum 1969–2003 durch. Dabei stellt er nicht nur für Allensbach den geringsten MAE fest (124), sondern Allensbach bestünde als einziger

Prognostizierer den Test auf starke Rationalität und sei somit effizienter als die anderen beiden (127–130). Döpke und Fritsche führen 2005 ihren 2004er Ansatz weiter und wenden ihr ökonometrisches Testrepertoire auf ein anderes Gebiet, nämlich auf den Zusammenhang zwischen makroökonomischen Variablen und Prognosefehlern an. Um den von ihnen vermuteten nichtlinearen Zusammenhang mit Rationalitätstest nachzuweisen, konstruieren sie eine Reihe von Modellen. Am Ende stellt sich nur die Geldpolitik beziehungsweise geldpolitische Störungen als relevanter Faktor für große Prognosenunter- und -überschätzungen heraus (25). Vring, SVR und Institute 1993–2003 untersuchend, stellt fest, „dass diese jährlichen Wachstumsprognosen der Experten äußerst unzuverlässig sind“ (2005: 52).

Genau gegenteilig, was sowohl die Untersuchungsmethoden als auch das Fazit angeht, äußert sich Fritsche 2005 (er benutzt den Datensatz von Döpke und Fritsche 2004): Prognosen können „zur Orientierung dienen“ (368). Das 95 %-Konfidenzintervall betrage derzeit ± 2 %. Die Prognosegenauigkeitsunterschiede zwischen den Instituten ließen sich durch den Erstellungszeitpunkt erklären (367). Heilemann und Klinger benutzen 2006 den aus vorherigen Veröffentlichungen bekannten Datensatz der Jahre 1967–2003, erhalten wieder MAEs von 1,1 % (Jahreswirtschaftsbericht) bis 1,4 % (Institute), betonen erneut die Stagnation der Prognosequalität von den 1970er Jahren bis heute und befinden die Prognosefehler ähnlich wie Helmstädter für autokorreliert (231). Außerdem vergleichen mit Hilfe des OECD-Prognosensatzes die Vorhersagegenauigkeit einiger Länder miteinander. Deutschland liege 1967–2001 dabei mit 1,4 % MAE eher im hinteren Feld, es gäbe mit Frankreich (0,9 %), Großbritannien (1,1 %) und USA (1,1 %) durchaus besser prognostizierende Länder (233). Heilemann und Klinger bezweifeln, dass mit mehr Wettbewerb sich automatisch die Prognosequalität verbessern würde, weil kommerzielle Prognostiker nicht unbedingt bessere Prognosen erstellten (248). Nierhaus, Mitarbeiter von ifo München, berechnet 2006 für die Jahre 1992–2005 für ifo München einen mittleren absoluten Prognosefehler von 0,7 %, für den Sachverständigenrat von 0,9 % und für die Institute von 1,0 % (41). In neun von 14 Jahren sei die Abweichung kleiner als 0,5 % gewesen (40). Teile ihres Discussion Papers von 2004 veröffentlichen Döpke und Fritsche 2006 im *International Journal of Forecasting*. Ihr Untersuchungsgegenstand sind die Unterschiede zwischen den Prognoseerstellern. Zunächst wird ein Maß zur Messung der Streuung zwischen den wieder betrachteten 14 Prognoseorganen (1970–2004) gefunden. Sie stellen fest, dass die Prognosen besonders vor Rezessionen weit auseinander lägen. Über die meisten Jahre dominiere das Bild der Konsens-Prognose. Danach entwickeln sie Prüfstests, um den Zusammenhang zwischen der Schwankung makroökonomischer Größen und der Streuung der Prognosen festzustellen. Als Ergebnis halten sie fest:

„A positive correlation between forecast dispersion and the volatility of macroeconomic variables indicates that the forecasters disagree on the relative importance of and the dynamic responses to shocks“ (134).

Kappler führt in seiner Untersuchung 2006, die Jahre 1980–2004 auswertend, eine Reihe von Kritiken (Spannweite nicht objektiv bestimmbar, schwer kommunizierbar) an einer möglichen Angabe von Prognoseintervallen an (13).

8 Schlussbewertung

Heilemann fasst das Ergebnis von über vier Jahrzehnten Konjunkturprognosen zusammen:

„In den letzten 40 Jahren hat sich die Treffsicherheit [...] praktisch nicht verbessert (2005: 30).

Gilt Ähnliches etwa auch für die Konjunkturprognose-Evaluationsforschung? Wenn man sich die letzten 39 Jahre mit ihren 88 Prognoseüberprüfungsarbeiten anschaut, kann man das nicht sagen. Früher wurden mit selbstkonstruierten und damit oft zu komplizierten Prüfmaßen (Spahn 1968, Ahn 1983) nur recht knappe Prognosendatensätze ausgewertet. Meist wurde nur der Sachverständigenrat und die Gemeinschaftsdiagnose der Institute überprüft. Die herangezogenen Zeiträume waren kurz und schöpften den eigentlich möglichen Zeitraum seit 1950 nicht aus. Heute gibt es Untersuchungen mit standardisierten Prognosemaßen, die seit Mitte der 00er Jahre auch auf langen Zeitreihen und vielen Prognosestellern basieren, also auf einer großen Datengrundlage. Ob man aber die seit Ende der 1990er Jahre um sich greifende Adaptation amerikanisch-ökonomischer Prognoseverfahren (vgl. Clements 2005) und die Englischsprachigkeit als reinen Fortschritt einstufen soll, sei dahingestellt. In Deutschland gibt es aus gutem Grund eine jahrzehntelange Tradition, einfache und damit nachvollziehbare Prognosemaße anzuwenden und auch einmal kritische Worte zur Prognosenqualität zu äußern, die nur auf deutsch geschrieben einen ausreichend großen Rezipientenkreis erreichen.

Die qualitative Verbesserung der Prognoseevaluationsforschung im Zeitablauf spiegelt sich auch quantitativ wider. Von den 1960er Jahren mit fünf Arbeiten, den 1970er mit fünf Arbeiten, den 1980er mit 19, den 1990er mit 27 bis zu den vermutlich über 50 Arbeiten in den 2000er Jahren (2000–2006 sind es jetzt schon 32) ist eine beeindruckende Zunahme von Prognoseüberprüfungsarbeiten erkennbar.

Die hier durchgeführte Zusammenstellung aller 88 Prognoseüberprüfungsarbeiten würde umfangreiche Sekundäranalysen erlauben. Es sei abschließend nur eine präsentiert. Für die folgende Tabelle werden alle 88 Prognoseüberprüfungsarbeiten erstens im Hinblick auf die Tendenz der Grundaussage über die Qualität der Konjunkturprognosen eingestuft (meist anhand verbaler Aussagen im Text). Zweitens wird die Herkunft des Autors (Insti-

Tabelle 1

Zusammenhang zwischen Prognostikerherkunft und Prognoseüberprüfungsergebnis 1967–2006 (N = 88 Arbeiten)

Anzahl der Prognoseüberprüfungsarbeiten Prognose	Bewertung mittel, mittel bis gut, gut	Prognose-Bewertung mittel bis schlecht, schlecht
Institutsmitarbeiter (Eigenevaluation)	42	4
Institutsexterne (meist universitätsnah) (Fremdevaluation)	15	27

Chi-Quadrat 29,73 > Chi-Quadrat-Prüfgröße (df 1; 99,9 %) 10,83 ==> *** signifikant

Phi-Koeffizient = $(\text{Chi-Quadrat}/N)^{1/2} = 0,58$ ==> enger Zusammenhang

Quelle: Eigene Zusammenstellung; Seitenzahlenbeleg im Anhang Tabelle A 1.

tutsmitarbeiter oder Nicht-Institutsmitarbeiter) aufgezeichnet. Beides wird im Anhang dokumentiert.

Institutsmitarbeiter, die mit 46 mehr als die Hälfte der 88 Prognoseevaluationsarbeiten veröffentlichen, kommen zu 91 % zu dem Ergebnis, dass die Prognoseleistung mittel bis gut sei. Prüft ein Externer, also jemand, der nicht beruflich die Prognosen erstellt, in der Regel ein Universitätsangehöriger, ist die Wahrscheinlichkeit immerhin 64 %, dass seine Prognoseeinschätzung schlecht ist. Da viele der mittel bis guten Prognosebeurteilungen der Externen von Kirchgässner, also nur von einem Autor stammen, erhält man unter Ausklammerung dieses Autors ein ziemlich frappierendes Ergebnis: ein institutsinterner Evaluierer kommt zu einem positiven Prüferesultat, ein Externer zu einem negativen. Dieser einem eigentlich zu Denken gebende Cuius-regio-, Eius-religio-Zusammenhang ist statistisch hoch signifikant (Irrtumswahrscheinlichkeit kleiner 0,01 %). Die besonders kritischen Prognoseüberprüfungen stammen im Übrigen erstens von den Promovierenden und zweitens von ehemaligen Prognoseerstellern kurz nach Beendigung ihrer Institutstätigkeit wie Helmstädter, Roberts, Glismann oder Heilemann.

Wenn man sich die fast 80 Jahre Konjunkturprognosenüberprüfungsarbeiten anschaut, scheint sich der von Morgenstern 1928 mit einem lauten Nein zur Konjunkturprognose („Sie sollen die Prognose aufgeben“; (121)) gestartete Prognoseevaluationskreis eher zu schließen. In den letzten beiden Jahren mehren sich immer mehr die Stimmen, die aufgrund der ernüchternden Ergebnisse von 39 Jahren deutscher Prognoseevaluation die mit 3,5 Mill. Euro (Heilemann und Blaschik 2004: 8) jährlich geförderte staatliche Prognosenerstellung durch die Wirtschaftsforschungsinstitute komplett in Frage stellen (Betz 2004: 188, Dicke und Glismann 2004: 9, Antholz 2005: 134, Heilemann und Klinger 2005: 32), am deutlichsten Vring:

„Der Politik ist deshalb ein konsequenter Verzicht auf Wachstumsprognosen zu empfehlen“ (2005: 56).

Literaturverzeichnis

- Ahn, H.-A. (1985): Konjunkturprognosen im Vergleich. Eine vergleichende Untersuchung der Jahresprognosen von Verbänden, Wirtschaftsforschungsinstituten und des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung im Zeitraum von 1963–1979, Dissertation. Technische Hochschule Aachen, Frankfurt a.M.
- Albeck, H. (1970): Ein Vergleich der Treffsicherheit deutscher und niederländischer Konjunkturprognosen. *Jahrbücher für Nationalökonomie und Statistik*, 184, 31–59
- Antholz, B. (2000): Validität deutscher Wirtschaftswachstumsprognosen 1974–1999. Magisterarbeit an der Universität Hamburg, Hamburg.
- Antholz, B. (2005): Zur Treffsicherheit von Wachstumsprognosen. Prognostizierte und tatsächliche Wachstumsraten des Bruttoinlandsprodukts Deutschlands 1950–2004. Dissertation. Universität Lüneburg 2005. Münster.
- Betz, Gregor (2004): Apriorische und empirische Grenzen von Wirtschaftsprognose: Oskar Morgenstern nach 70 Jahren Prognoseerfahrung. In: Ulrich Frank (Hrsg.): *Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik. Theoriebildung und -bewertung, Ontologien, Wissensmanagement*. Wiesbaden, 171–190.

- Blix, Märten, Joachim Wadefjord, Ulricke Wienecke und Martin Ådahl (2001): How Good is the Forecasting Performance of Major Institutions? *Sveriges Riksbank Economic Review*, 3, 38–68.
- Bombach, G. (1962): Über die Möglichkeiten wirtschaftlicher Voraussagen. *Kyklos*, 15, 29–67.
- Bosse, Lothar (1957): Über die Möglichkeit und den Nutzen von kurzfristigen Wirtschaftsprognosen. *Weltwirtschaftliches Archiv*, LXXIX, 65–83.
- Brinkmann, Theodor (1976): Zielprojektionen und Wirtschaftspolitik 1968/75. *Beiträge zur Wirtschafts- und Sozialpolitik des Instituts der deutschen Wirtschaft* Nr. 31, 4/1976.
- Büttner, Thiess und Gustav A. Horn (1993): *Wie gut sind Prognosen? Eine Überprüfung der Vorhersagen der Gemeinschaftsdiagnose und des DIW am Beispiel des BSP*. DIW Diskussionspapier Nr. 73. Berlin.
- Clements, Michael P. (2005): *Evaluation Econometric Forecasts of Economic and Financial Variables Hampshire*. New York.
- Cornelius, Peter (1989): Wie treffsicher sind Konjunkturprognosen? *Wirtschaftsdienst*, 1, 42–48.
- Delbanco, G.A. (1925): Konjunktur, Konjunkturlosigkeit und Konjunkturprognose. *Jahrbücher für Nationalökonomie und Statistik* (damaliger Nebentitel: Conrads Jahrbücher, III Folge, Bd. 68), 123, 808–815.
- Dicke, Hugo und Hans H. Glismann (2002a): Konjunkturprognosen und wissenschaftlich-technischer Fortschritt. *Wirtschaftsdienst*, 3, 167–169.
- Dicke, Hugo und Hans H. Glismann (2002b): Haben sich die Konjunkturprognosen des Sachverständigenrates verbessert? *Wirtschaftsdienst*, 12, 736–740.
- Dicke, Hugo und Hans H. Glismann (2002c): Eine Antwort auf Jens Weidmann. *Wirtschaftsdienst*, 12, 747 f.
- Dicke, Hugo und Hans H. Glismann (2004): *Vom Elend der Konjunkturprognosen*. Deutsche Bank Research Notes Working Paper Series Nr. 15, 03.11.2004.
- DIW (1928): Vierteljahrshefte zur Konjunkturforschung, 2 (4), abgeschlossen 25.02.1928.
- Dominguez, K.M., R.C. Fair und M.D. Shapiro (1988): Forecasting the Depression. *American Economic Review*, 78 (9), 596–612.
- Döpke, J. (2000): Haben Konjunkturprognosen in Deutschland einen politischen Bias? *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaft (ZWS)*, 120, 587–620.
- Döpke, J. (2004): Zur Qualität von Konjunkturprognosen. *Wirtschaftswissenschaftliches Studium (WiSt)*, 1, 8–13.
- Döpke, J. und U. Fritsche (2004): *Growth and Inflation Forecasts in Germany – An Assessment of Accuracy and Dispersion*. DIW Diskussionspapier Nr. 399. Berlin.
- Döpke, J. und U. Fritsche (2005): *Forecast Errors and the Macroeconomy – A Non-Linear Relationship?* DIW Diskussionspapier Nr. 498. Berlin.
- Döpke, J. und U. Fritsche (2006): When do Forecasters Disagree? An Assessment of German Growth and Inflation Forecast Dispersion. *International Journal of Forecasting*, 22, 125–135.
- Döpke, Jörg und Enno Langfeldt (1995): *Zur Qualität der Konjunkturprognosen für Westdeutschland 1976–1994*. Kieler Diskussionsbeiträge, Institut für Weltwirtschaft IfW Kiel. März.
- Filusch, Bernhard (1993): Die Konjunkturprognosen der Forschungsinstitute. Anspruch und Wirklichkeit. In: Willi Leibfritz (1993): *Die Konjunkturprognose der Forschungsinstitute. Anspruch und Wirklichkeit aus der Sicht der Prognostiker und der wirt-*

- schaftspolitischen Praxis*. Symposium zur Verabschiedung von Dr. Arthur Krumper am 22. Mai 1992. München, 7–18.
- Frerichs, W. und K. Kübler (1980): *Gesamtwirtschaftliche Prognoseverfahren*. München.
- Fritsche, Ulrich (2005): Warum Konjunkturprognosen? *Wochenbericht des DIW Berlin*, Nr. 22, 361–369.
- Giersch, Herbert und Knut Borchardt (1962): *Diagnose und Prognose als wirtschaftswissenschaftliche Methodenprobleme*. Verhandlungen auf der ersten Tagung in Garmisch-Partenkirchen, 25.–28. September 1961 (Neue Folgen in den Schriften des Vereins für Socialpolitik. Bd. 25. Berlin).
- Grömling, Michael (2002a): Konjunkturprognosen: Methoden, Risiken und Treffsicherheiten. *IW-Trends*, 2, 18–26.
- Grömling, Michael (2002b): Evaluation and Accuracy of Economic Forecasts. *Historical Social Research*, 27 (4), 242–255.
- Grömling, Michael (2005): Konjunkturprognosen – Verfahren, Erfolgskontrolle und Prognosefehler. Diskussionsbeiträge Nr. 123 des Volkswirtschaftlichen Seminars an der Universität Göttingen, Februar.
- Hagen, H.M. und G. Kirchgässner (1997): *Interest Rate Based Forecasts vs. Professional Forecasts for German Economic Growth: A Comparison*. Discussion Paper Nr. 9714. Universität St. Gallen.
- Harvey, Campbell R. (1991): Interest Rate Based Forecasts of German Economic Growth. *Weltwirtschaftliches Archiv*, 127, 701–718.
- Heilemann, U. (1981a): *Zur Prognoseleistung ökonometrischer Konjunkturmodelle für die Bundesrepublik Deutschland*. Berlin.
- Heilemann, U. (1981b): Zur Prognosegenauigkeit ökonometrischer Konjunkturmodelle für die Bundesrepublik Deutschland. *Allgemeines Statistisches Archiv*, 65 (3), 242–272.
- Heilemann, U. (1998a): Erfahrungen mit dem RWI-Konjunkturmodell 1974 bis 1994. In: U. Heilemann und J. Wolters (Hrsg.): *Gesamtwirtschaftliche Modelle in der Bundesrepublik Deutschland: Erfahrungen und Perspektiven*. Berlin, 61–93.
- Heilemann, U. (1998b): Paradigm lost – Zu den Projektionen des Jahreswirtschaftsberichtes der Bundesregierung. In: U. Heilemann, D. Kath und N. Klotten (Hrsg.): *Entgrenzung als Erkenntnis- und Gestaltungsaufgabe*. Festschrift für Reimut Jochimsen zum 65. Geburtstag. Berlin, 79–100.
- Heilemann, U. (1999): Kleine „Mühen beim Umsteigen“ – Zur Treffsicherheit der Prognosen für Ostdeutschland 1991 bis 1997. In: M. Erlei, M. Leschke, D. Sauerland und E. Schulz (Hrsg.): *Beiträge zur angewandten Wirtschaftstheorie*. Manfred Borchert zum 60. Geburtstag. Regensburg, 67–78.
- Heilemann, U. (2003): *A Bridge to Far? 20 Years of ex ante Forecasts with the RWI Business Cycle Model*. Papier vorgestellt auf dem Jahrestreffen des Ausschusses für Empirische Wirtschaftsforschung des Vereins für Socialpolitik am 6. März 2003 in Schloss Rauischholzhausen.
- Heilemann, U. (2004a): Besser geht's nicht – Genauigkeitsgrenzen von Konjunkturprognosen. *Jahrbücher für Nationalökonomie und Statistik*, 224 (1), 51–64.
- Heilemann, U. (2004b): Das RWI-Konjunkturmodell – Ein Überblick. In: W. Gaab, U. Heilemann und J. Wolters (Hrsg.): *Arbeiten mit ökonometrischen Modellen* (Studies in Contemporary Economics). Heidelberg, 161–212.
- Heilemann, U. und H.O. Stekler (2003): Has the Accuracy of German Makroeconomic Forecasts Improved? Technical Report 31/2003 im Sonderforschungsbereich 475 Komplexitätsreduktion in multivariaten Datenstrukturen an der Universität Dortmund.

- Heilemann, Ulrich und Annika Blaschzik (2004): Que sera, sera? Zur Genauigkeit von Konjunkturprognosen. *Universitätsjournal – Mitteilungen und Berichte für die Angehörigen und Freunde der Universität Leipzig*, 7, 8–9.
- Heilemann, Ulrich und Sabine Klinger (2006): Zu wenig Wettbewerb? Zu Stand und Entwicklung der Genauigkeit makroökonomischer Prognosen (Vortrag im Rahmen der Jahrestagung 2005 des Wirtschaftspolitischen Ausschusses des Vereins für Socialpolitik, Marburg, 15.–17. März 2005. Entstanden im SFB 475 Komplexitätsreduktion in multivariaten Datenstrukturen an der Universität Dortmund). In: Wolf Schäfer (Hrsg.): *Wirtschaftspolitik im Systemwettbewerb*. Schriften des Vereins für Socialpolitik, N.F., Bd. 309. Berlin, 225–257.
- Heise, Michael (1991): Verfahren und Probleme der Konjunkturvorhersage – dargestellt am Beispiel der Sachverständigenratsprognosen. In: Gholamreza Nakhaeizadeh und Karl-Heinz Vollmer (1991): *Anwendungsaspekte von Prognoseverfahren*. Beiträge zum 2. Karlsruher Ökonometrie-Workshop. Heidelberg, 18–37.
- Helmstädter, Ernst (1973): Was leisten Prognosemodelle? *Zeitschrift für allgemeine und textile Marktwirtschaft*, 4, 118–152.
- Helmstädter, Ernst (1989): Zur Autokorrelation der Konjunkturprognosen. *ifo Studien*, 35, 158–164.
- Helmstädter, Ernst (1991a): Die Konjunkturprognosen und die Stimmungsprozente. *Wirtschaftsdienst*, 7, 360–365.
- Helmstädter, Ernst (1991b): Entgegnung auf van Suntum. *Wirtschaftsdienst*, 7, 368–369.
- Helmstädter, Ernst (1991c): Hinsehen kann nicht schaden. Replik auf Kirchgässner. *ifo Studien*, 37, 159–162.
- Helmstädter, Ernst (1992): Die Wirtschaftsprognose – ein untauglicher Versuch? *List-Forum für Wirtschafts- und Finanzpolitik*, 18 (2), 112–126.
- Helmstädter, Ernst (1996): Zur wirtschaftlichen Nützlichkeit von Konjunkturprognosen. Prognosen und ihre wirtschaftlichen Anwendungsgebiete. *Zeitschrift für Wirtschaftspolitik*, 45 (2), 162–170.
- Hinze, J. (1996): Konjunkturprognosen unter veränderten Rahmenbedingungen: Auswirkungen der deutschen und europäischen Einigung. *Hamburger Jahrbuch für Wirtschafts- und Gesellschaftspolitik*, 41, 59–84.
- Hinze, Jörg (2005): Konjunkturprognosen: Falsche Erwartungen an die Treffgenauigkeit. *Wirtschaftsdienst*, 2, 117–123.
- IW/Institut der Deutschen Wirtschaft (2002): Konjunkturprognosen. Viele Stolpersteine. *Informationsdienst des Instituts der deutschen Wirtschaft/iwd*, Nr. 17. 25. April 2002, 6–7.
- Kaltefleiter, W. (1961): Gegenwärtige Methoden der Konjunkturprognose und ihre Anwendung in Deutschland. Unveröffentlichte Diplomarbeit am Institut für Wirtschaftspolitik an der Universität zu Köln, Wintersemester 1960/61.
- Kappler, Marcus (2006): *Wie genau sind die Konjunkturprognosen der Institute für Deutschland?* ZEW Discussion Paper No. 06-004. Mannheim.
- Keereman, Filip (1999): *The Track Record of the Commission Forecasts*. Economic Papers No. 137. Europäische Kommission.
- Kirchgässner, G. (1982): Sind die Erwartungen der Wirtschaftssubjekte rational? Eine empirische Untersuchung für die Bundesrepublik Deutschland. *Weltwirtschaftliches Archiv*, 118, 215–240.

- Kirchgässner, G. (1984): Wie gut sind Prognosen der Arbeitsgemeinschaft wirtschaftswissenschaftlicher Forschungsinstitute in der Bundesrepublik Deutschland. *Weltwirtschaftliches Archiv*, 120, 279–300.
- Kirchgässner, G. (1985): Sind die Prognosen der Arbeitsgemeinschaft deutscher wirtschaftswissenschaftlicher Forschungsinstitute schwach rational? – Eine Antwort auf Neumann und Buscher. *Weltwirtschaftliches Archiv*, 121, 331–345.
- Kirchgässner, G. (1991): Sind die Prognosefehler autokorreliert? *ifo-Studien*, 37, 151–158.
- Kirchgässner, G. (1993): Testing Weak Rationality of Forecasts with Different Time Horizons. *Journal of Forecasting*, 12, 541–558.
- Kirchgässner, G. (2005): On the Rationality of the General Public. *Applied Economics Quarterly*, 51 (2), 120–131.
- Kirchgässner, G. und U. Müller (2002): Are Forecasters Reluctant to Revise Their Predictions? Some German Evidence. Vortrag auf dem 22. Internationalen Symposium on Forecasting in Dublin. Juli.
- Kirchgässner, G. und M. Savoiz (2001): Monetary Policy and Forecasts for Real GDP Growth: An Empirical Investigation for the Federal Republic of Germany. *German Economic Review*, 2 (4), 339–365.
- Kleinewefers, H. (1970): Die Konjunkturprognosen des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung. *Zeitschrift für die gesamte Staatswissenschaft*, 126, 249–289.
- Koch, H. und Ch. Stege (1990): Vorsicht, Prognosen. *Manager Magazin*, 5, 180–187.
- Krengel, Rolf (1984): Die wirtschaftliche Entwicklung im Deutschen Reich in den 20er und 30er Jahren aus der Sicht des Instituts für Konjunkturforschung. *Vierteljahrshäfte zur Wirtschaftsforschung*, 1, 30–37.
- Lamberts, W. und L. Schüssler (1967): Zur Treffsicherheit von Konjunkturprognosen der Wirtschaftsinstitute. *Mitteilungen des Rheinisch-Westfälischen Instituts für Wirtschaftsforschung Essen* (später RWI-Mitteilungen), 18 (4), 269–296.
- Langfeldt, E. und N. Walter (1985): Wirtschaftsprognosen. Die Macht der Fakten. *Wirtschaftswoche*, Nr. 10, 70–74.
- Langfeldt, Enno (1996): Konjunkturprognosen – unvollkommen aber unverzichtbar. *Zeitschrift für Wirtschaftspolitik*, 45 (2), 180–186.
- Langfeldt, Enno und Peter Trapp (1986): Zur Problematik und Treffsicherheit von Konjunkturprognosen für die Bundesrepublik Deutschland. *Weltwirtschaft* (IfW Kiel), 1, 26–37.
- Langfeldt, Enno und Peter Trapp (1988): Experiences in Macroeconomic Forecasting in the Federal Republic of Germany 1976–1987. *Jahrbücher für Nationalökonomie und Statistik*, Bd. 205/5, 427–442.
- Löwe, Alfred (1925): Der gegenwärtige Stand der Konjunkturforschung in Deutschland. In: M.J. Bonn und M. Palyi (Hrsg.): *Die Wirtschaftswissenschaft nach dem Kriege*. München, 329–377.
- Loungani, P. (2001): *How Accurate are Private Sector Forecasts? Cross-Country Evidence from Consensus Forecasts of Output Growth?* IMF Working Paper. Washington, D.C.
- Meerwarth, R. (1925): *Nationalökonomie und Statistik*. Berlin.
- Morgenstern, Oskar (1928): *Wirtschaftsprognose. Eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*. Wien.
- Morgenstern, Oskar (1930): Nachwort zur Wirtschaftsprognose. Eine Entgegnung. *Allgemeines Statistisches Archiv*, 20, 273–277.

- Morgenstern, Oskar (1950): *On the Accuracy of Economic Observations*. Princeton, NJ, Princeton University Press.
- Morgenstern, Oskar (1963): Vollkommene Voraussicht und wirtschaftliches Gleichgewicht. In: Oskar Morgenstern: *Spieltheorie und Wirtschaftswissenschaft*. Wien, München, 43–70.
- Muth, John F. (1961): Rational Expectations and the Theory of Price Movements. *Econometrica*, 29 (3), 315–335.
- Neumann, Manfred J.M. und Herbert S. Buscher (1985a): Wirtschaftsprognosen im Vergleich: Eine Untersuchung anhand von Rationalitätstests. *ifo Studien*, 13, 183–201.
- Neumann, Manfred J.M. und Herbert S. Buscher (1985b): Wie gut sind die Prognosen der Arbeitsgemeinschaft wirtschaftswissenschaftlicher Forschungsinstitute in der Bundesrepublik Deutschland? – Ein Kommentar. *Weltwirtschaftliches Archiv*, 121, 321–330.
- Nierhaus, Wolfgang (1998): Ifo Konjunkturprognosen und Wirklichkeit – Eine Ergebnisanalyse für die Bundesrepublik im Zeitraum 1991 bis 1997. *ifo Wirtschaftskonjunktur*, 50 (7), R1–R11.
- Nierhaus, Wolfgang (2006): Wirtschaftskonjunktur 2005: Prognose und Wirklichkeit. *ifo Schnelldienst*, 59 (2), 37–43.
- Noelle-Neumann, Elisabeth (1980): Aussichten für 1981. Über den Zusammenhang zwischen Neujahrstimmung und Wirtschaftswachstum im folgenden Jahr. *Allensbacher Berichte*, Nr. 21, Dezember, 1–7.
- OECD (1993): How Accurate Are Economic Outlook Projections? *OECD Economic Outlook*, Nr. 53, June, 49–54.
- Öller, Lars-Erik und Bharat Barot (2000): The Accuracy of European Growth and Inflation Forecasts. *International Journal of Forecasting*, 16, 293–315.
- Oppenländer, Karl Hinrich (1996): Die Konjunkturforschung auf den Punkt gebracht. *ifo Studien*, 42, 563–577.
- Pflaumer, Peter (1986): Messung und statistische Analyse der Fehler in den Prognosen des Sachverständigenrates. *Allgemeines Statistisches Archiv*, 70, 368–387.
- Pflaumer, Peter und Enno Swart (1987): Konjunkturvorhersagen im Vergleich. Expertenprognosen versus Zeitreihenprognosen. *Allgemeines Statistisches Archiv*, 71, 353–363.
- Pons, J. (2000): The Accuracy of IMF and OECD Forecasts for G7 Countries. *Journal of Forecasting*, 19, 53–63.
- Rabitsch, Franz (2002): *Die österreichischen Wachstumsprognosen 1978 bis 1999*. Working Papers 5/2002 des Bundesministerium für Finanzen in Österreich.
- Radowski, Daniel (2002): Wie treffsicher sind Prognosen? *ZEW Konjunkturreport*, März, 10–12.
- Rinne, H. und Katja Specht (2002): *Zeitreihen. Statistische Modellierung, Schätzung und Prognose*. München.
- Rinne, Horst (1997): Bemerkungen zur Qualität der Konjunkturprognosen des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung. In: Peter Lippe, Norbert von der Rehm, Heinrich Strecker und Rolf Wiegert (Hrsg.): *Wirtschafts- und Sozialstatistik heute. Theorie und Praxis*. Festschrift für Walter Krug. Sternenfels, 319–340.
- Roberts, C.C. (1981): *Konjunkturprognosen und Wirtschaftspolitik*. Köln.
- Schlesinger, H. (1967): Wert und Grenzen gesamtwirtschaftlicher Vorausschätzungen. *Allgemeines Statistisches Archiv*, 51, 401–416.
- Schumpeter, J. (1926): *Theorie der wirtschaftlichen Entwicklung*. 2. Aufl. München.

- Smolny, W. (1998): Die Wirtschaftsprognosen des Sachverständigenrates zur Begutachtung der gesamtwirtschaftlichen Entwicklung und der Arbeitsgemeinschaft wirtschaftswissenschaftlicher Forschungsinstitute – Ein Versuch einer Evaluierung (Aufsatz). Konstanz.
- Sommer, Albrecht (1929): Die Konjunkturprognose – Zur Möglichkeit der Wirtschaftsvoraussage. *Allgemeines Statistisches Archiv*, 18, 260–278.
- Spahn, B. (1968): Zur Treffsicherheit von Konjunkturprognosen der wirtschaftswissenschaftlichen Forschungsinstitute. Deutsches Institut für Wirtschaftsforschung (Institut für Konjunkturforschung). *Vierteljahrshefte zur Wirtschaftsforschung*, 2, 194–213.
- Spinanger, Dean und Norbert Walter (1983): *The Reliability of Macro-Economic Forecasts*. Arbeitspapier Nr. 181 des Instituts für Weltwirtschaft Kiel, Juli.
- Stege, Christoph (1989): *Zur Beurteilung der Prognosetätigkeit der „führenden wirtschaftswissenschaftlichen Forschungsinstitute“ in der Bundesrepublik Deutschland. Ist Politik via Prognosen möglich?* Eine empirische Analyse. Dissertation. Universität Hamburg, Hamburg.
- Suntum, Ulrich van (1991): Konjunkturprognosen: Besser als ihr Ruf. *Wirtschaftsdienst*, H.7, 365–368.
- Theil, Henri (1955): Who Forecasts Best? *International Economic Papers*, 5, 333–353.
- Theil, Henri (1966): *Applied Economic Forecasting* (Volume 4 in Studies in Mathematical and Managerial Economics). Amsterdam.
- Tichy, G. (1994): *Konjunktur – Stilisierte Fakten, Theorie, Prognose*. 2. völlig neu bearb. Aufl. Berlin u.a.
- Tietzel, Manfred (1981/82): Über die Unmöglichkeit von Wirtschaftsprognosen. *List Forum*, 11, 31–44.
- Vajina, Thomas (1972): Der Konjunkturzyklus 1967/71 im Lichte der Prognosen und Zielprojektionen. *Berichte des Deutschen Industrieinstituts zur Wirtschaftspolitik*, 6 (5).
- Vajina, Thomas (1977): *Prognosen für die Politik*. Köln.
- Vogel, E.H. (1927): *Wirtschaftsbarometer und Konjunkturprognose. Die Wege ihrer Verwirklichung in Mitteleuropa*. Teplitz-Schönau.
- Vogel, E.H. (1929): Wirtschaftsprognose. *Allgemeines Statistisches Archiv*, 19, 251–268.
- Vogel, E.H. (1930): Replik. *Allgemeines Statistisches Archiv*, 20, 277–281.
- Vogel, O. und L. Duelli (1968): Prognosewerte und Istdaten. *Beiträge des deutschen Industrieinstituts (Köln)*, 6 (2/3), 3–51.
- Vring, Thomas von der (2005): Zum Wert von Wachstumsprognosen. In: D. Ehrig und U. Staroske (Hrsg.): *Dimensionen angewandter Wirtschaftsforschung: Methoden, Regionen, Sektoren*. Festschrift für Heinz Schäfer zum 65. Geburtstag. Hamburg, 47–57.
- Weichhardt, R. (1982): *Zur Beurteilung von Konjunkturprognosen*. Tübingen.
- Weidmann, Jens (2002): Hat sich die Prognosetreffsicherheit des Sachverständigenrates systematisch verändert? *Wirtschaftsdienst*, 12, 741–747.
- Winkler, Peter (1997): *Empirische Wirtschaftsforschung*. Berlin u.a.
- Wörz, Julia (1995): Evaluierung der Leistungsfähigkeit von Prognosemethoden anhand ausgewählter Variablen. *Wirtschaftspolitische Blätter*, 2, 157–162.
- Zarnowitz, Victor (1992a): *Business Cycles: Theory, History, Indicators, and Forecasting*. Chicago und London.
- Zarnowitz, Victor (1992b): Has Macro-Forecasting Failed? *Cato Journal*, 12, 129–160.
- Zarnowitz, Victor (2005): *On Learning from History and Rules, Forecasts and Errors*. Vortrag auf der International Conference „The Future of Economic Forecasting“. Leipzig, 16.–17. Dezember 2005.

Anhang

Tabelle A1

Autor(en) Veröffentlichungsjahr: Prognosenbeurteilung findet sich auf Seite(n) ...	Eigenbeurteilung (Institutsmitarbeiter)	Fremdbeurteilung (meist universitätsnah)
Gut	Lamberts/Schüssler 1967: 296; Spahn 1968: 213; Vajina 1976: 68; Langfeldt/Walter 1985: 70 f.; Langfeldt/Trapp 1986: 33; Langfeldt/Trapp 1988: 433 f.; Heise 1991: 33; van Suntum 1991: 368; OECD 1993: 53; Büttner/Horn 1993: 16; Langfeldt 1996: 181; Nierhaus 1998: R 9; Heilemann 1998b: 90, 95; Heilemann 1999: 72, 76; Institut der deutschen Wirtschaft 2002: 7; Weidmann 2002: 746; Radowski 2002: 11 f.; Heilemann/Stekler 2003: 8, 11 f.; Heilemann 2004a: 63; Döpke 2004: 13; Hinze 2005: 117, 123; Fritsche 2005: 369; Nierhaus 2006: 42	Kleinewefers 1970: 255, 259; Kirchgässner 1982: 233; Kirchgässner 1984: 284, 297 f.; Kirchgässner 1985: 343; Pflaumer/Swart 1987: 361; Kirchgässner 1991: 155 f.; Smolny 1998: 17
Mittel bis gut	Schlesinger 1967: 410; Hinze 1996: 70 f., 81; Oppenländer 1996: 574; Heilemann 1998a: 91; Grömling 2002: 20; Heilemann 2004b: 190; Grömling 2005: 4, 6; Kappler 2006: 12	Pflaumer 1986: 381; Tichy 1994: 257; Antzoulatos/Wilfing 2003: 14
Mittel	Vogel/Duelli 1968: 19–21; Brinkmann 1976: 15; Heilemann 1981b: 264; Spinanger/Walter 1983: 12–21; Filusch 1993: 18; Döpke/Langfeldt 1995: 7; Döpke 2000: 598; Heilemann 2003: 15, 18; Döpke/Fritsche 2004: 6; Döpke/Fritsche 2005: 25; Döpke/Fritsche 2006: 134	Albeck 1970: 58; Ahn 1985; Winkler 1997; Hagen/Kirchgässner 1997: 5 f.; Kirchgässner/Savoiz 2001: 360 f.
Mittel bis schlecht	Cornelius 1989: 46; Dicke/Glismann 2002a: 169	Weichhardt 1982: 200, 202; Neumann/Buscher 1985a: 197; Neumann/Buscher 1985b: 329; Kirchgässner 1993: 557; Rinne 1997: 333; Heilemann/Blaschzik 2004: 9
Schlecht	Vajina 1972: 5, 15; Dicke/Glismann 2002: 740	Wilpert 1969: 243; Frerichs/Kübler 1980: 253f; Roberts 1981: 27–29, 64; Stege 1989: 258; Helmstädter 1989: 163; Koch/Stege 1990: 180; Harvey 1991: 711; Helmstädter 1991a: 364; Helmstädter 1991c: 160 f.; Helmstädter 1992: 122; Wörz 1995: 161; Helmstädter 1996: 165 f.; Antholz 2000: 85–87; Rinne/Speccht 2002: 134; Kirchgässner/Müller 2002: 14; Dicke/Glismann 2004: 8; Betz 2004: 186 f.; Antholz 2005: 129 f.; Kirchgässner 2005: 121; Vring 2005: 52; Heilemann/Klinger 2006: 248 f.

Quelle: Eigene Zusammenstellung anhand der Originaltexte.