

Prognosegüte alternativer Frühindikatoren für die Konjunktur in Deutschland

Joachim Benner und Carsten-Patrick Meier
Institut für Weltwirtschaft, Kiel

28. Februar 2003

Zusammenfassung

Untersuchungen zur Prognosegüte sollten nicht nur Prognosefehler, die auf der Schätzung der Parameter beruhen, berücksichtigen, sondern auch solche, die aus der stichprobenabhängigen Auswahl des Prognosemodells resultieren. Wird die Prognosefehlervarianz durch rekursive Out-of-Sample Prognosen geschätzt, so sollte dabei nicht nur die Parameterschätzung, sondern auch die Modellselktion rekursiv vorgenommen werden. Wir wenden dieses Prinzip auf die Analyse der Prognosegüte dreier wichtiger Indikatoren für die Konjunktur in Deutschland an, den vom ifo-Institut erhobenen „Geschäftserwartungen“, den vom Zentrum für Europäische Wirtschaftsforschung veröffentlichten „Konjunkturerwartungen“ und des von der „Wirtschaftswoche“ berechneten „Earlybird“-Indikators. Es zeigt sich, dass die Prognosefehler bei der realistischen rekursiven Modellauswahl größer sind als bei nicht-rekursiver Spezifikation. Die untersuchten Indikatoren liefern unter bestimmten Umständen bessere Prognosen als ein einfaches autoregressives Modell

1 Motivation

Konjunkturindikatoren stellen ein wesentliches Instrument zur Abschätzung der zukünftigen wirtschaftlichen Entwicklung dar. Für die Analyse der konjunkturellen Situation am aktuellen Rand, über die noch keine Produktionsdaten vorliegen, sowie für die Prognose der allernächsten Zukunft dürften sie sogar die wichtigste Informationsquelle des Konjunkturforschers darstellen. Konjunkturindikatoren können realwirtschaftliche Größen, wie etwa Auftragseingänge oder sektorale Produktionszahlen, monetäre Größen, wie Zinsen oder Geldmengenaggregate, oder Umfragedaten zur Einschätzung von gegenwärtiger oder zukünftiger Lage sein. Wesentlich ist, dass sie die nachfolgende Entwicklung der vorherzusagenden gesamtwirtschaftlichen Produktionsdaten mit einem möglichst weiten und stabilen Vorlauf zumindest teilweise vorwegnehmen. Gelegentlich werden verschiedene Einzelindikatoren auch zu Gesamtindikatoren aggregiert, von denen bessere Prognoseeigenschaften erwartet werden.

Bekannteste und zuverlässigste Indikatoren für Deutschland waren in der Vergangenheit die vom Münchner ifo-Institut durch monatliche Umfragen bei über 7000 Entscheidungsträgern in der gewerblichen Wirtschaft erhobenen „Konjunkturtest“-Ergebnisse, insbesondere die „Geschäftserwartungen“. Anfang der 90er Jahre haben diese Indikatoren allerdings „Konkurrenz“ bekommen. Das Zentrum für Europäische Wirtschaftsforschung (ZEW) in Mannheim befragt monatlich etwa 350 professionelle Konjunkturexperten, vor allem bei Banken, Versicherungen und Großunternehmen, nach ihrer Einschätzung der künftigen Entwicklung und fasst die Ergebnisse in einem Index der „Konjunkturerwartungen“ zusammen. Felix P. Hufner und

Michael Schröder (2002), im Folgenden H&S abgekürzt, haben jüngst den Informationsgehalt von ifo-Geschäftserwartungen und ZEW-Konjunkturerwartungen für die Prognose der deutschen Industrieproduktion miteinander verglichen.¹ Sie kommen dabei zu dem bemerkenswerten Schluss, dass zwar beide Indikatoren die zukünftige Entwicklung der Industrieproduktion besser vorherzusagen vermögen als eine autoregressive Prognose allein auf Basis der historischen Industrieproduktionsdaten, ökonometrische Prognosemodelle auf Basis der ZEW-Konjunkturerwartungen dabei aber bei allen Horizonten bis auf den einmonatigen zu signifikant niedrigeren Prognosefehlern führen als Modelle auf Basis der ifo-Geschäftserwartungen. Der deutlich kleineren Gruppe von professionellen Kapitalmarkt- und Konjunkturbeobachtern bei Banken und Versicherungen gelingt es insgesamt offenbar systematisch besser, die konjunkturelle Entwicklung vorauszusehen als der Gesamtheit der Entscheidungsträger in den Produktionsunternehmen.

Das vorliegende Papier nimmt die Ergebnisse von H&S zum Ausgangspunkt um eine Reihe von methodischen Veränderungen bei der Analyse von Konjunkturprognosemodellen zu untersuchen, von denen wir meinen, dass sie zu einer besseren Modellauswahl führen. Die Veränderungen beziehen sich zum einen direkt auf den Modellselektionsprozess, zum anderen auf die Methodik, mit der die Prognosegüte eines gefundenen Modells evaluiert wird.

Die Methodik, mit der H&S zu ihrem Ergebnis gelangen, ist bei der Untersuchung von Prognosemodellen weit verbreitet.² Dabei wird zunächst auf der Basis der gesamten zur Verfügung stehenden Daten ein Prognosemodell spezifiziert, wobei die üblichen Modellselektionskriterien (Lütkepohl 1991, Kap. 4) zum Einsatz kommen können. Anschließend wird die gewählte Spezifikation rekursiv über den Stichprobenzeitraum geschätzt und Prognosen über eine festgelegte Anzahl von Perioden berechnet. Anhand der Abweichungen zwischen der Vorhersage des Modells und den tatsächlichen Werten wird dessen Eignung zur Prognose beurteilt, zumeist indem die ermittelten Abweichungen mit denen eines Referenzmodells verglichen werden. Die rekursive Schätzung soll dabei sicherstellen, dass das Modell für die Prognose stets nur Informationen verwendet, die zum Start des Prognosezeitraums auch wirklich vorgelegen haben.

Für die Prognosegüte spielt allerdings neben der Parameterschätzung auch die Modellspezifikation eine wesentliche Rolle. Da letztere in dem oben beschriebenen Verfahren auf der Basis der vollen Stichprobe bestimmt wird, gehen an dieser Stelle grundsätzlich Informationen in die Prognosen ein, die zum Prognosezeitpunkt noch nicht verfügbar waren.³ Dies verzerrt die geschätzte Prognosefehlervarianz nur dann nicht, wenn die optimale Parametrisierung des Prognosemodells über den gesamten Stichprobenzeitraum stabil ist, also auch bei einer rekursiven Modellauswahl stets dasselbe Modell gewählt wird wie auf Basis der vollen Stichprobe. Da die volle Stichprobe aufgrund der höheren Zahl an Freiheitsgraden eine sichere Mo-

¹In den vergangenen Jahren beschäftigte sich eine Reihe von Studien mit der Güte von Konjunkturindikatoren in Deutschland, so etwa Döpke/Krämer/Langfeldt (1994), Fritsche (1999) und (2000), Fritsche/Stephan (2000), Funke (1997) und Langmantel (1999). Mit dem direkten Vergleich zwischen ifo-Geschäftserwartungen und ZEW-Konjunkturerwartungen befassten sich hingegen nur die Untersuchungen von Hüfner/Schröder (2002) und Breitung/Jagodzinski (2001).

²Etabliert hat sich diese Methodik nicht zuletzt im Zusammenhang mit der empirischen Überprüfung von theoretischen Wechselkursmodellen. So zeigten Meese und Rogoff (1983) in einer grundlegenden Arbeit, dass die Prognosegüte einer Reihe von Wechselkursmodellen nicht höher lag als die des einfachen Random-walk-Modells, dessen Prognosewerte stets mit dem zuletzt beobachteten Wert übereinstimmen.

³Ein weiteres Problem, das in diesem Zusammenhang relevant ist, betrifft die Art der verwendeten Daten. Handelt es sich dabei, wie beispielsweise bei der Industrieproduktion, um Daten, die nach der erstmaligen Veröffentlichung noch Revisionen unterworfen sind, so müsste prinzipiell auch berücksichtigt werden, dass zum Zeitpunkt der Prognose nur Daten zu dem jeweiligen Revisionsstand (z.B. Erstveröffentlichung, erste Revision, usw.) vorgelegen haben. In Ermangelung von Echtzeit-Daten zur Industrieproduktion wird dieser Punkt im folgenden vernachlässigt.

dellselektion erlaubt als eine Teilstichprobe, dürfte dies selbst bei grundsätzlicher Parameterkonstanz des datenerzeugenden Prozesses in endlichen Stichproben nicht der Fall sein. Insofern ist davon auszugehen, dass das obige Verfahren mit einmaliger Modellauswahl zu einer systematischen Unterschätzung der Prognosefehlervarianz führt.

In der vorliegenden Arbeit wird die Prognosegüte der Indikatoren für die deutsche Konjunktur für den realistischeren Fall untersucht, dass nicht nur die Parameterschätzung, sondern auch die Spezifikation der Prognosemodelle rekursiv vorgenommen wird. D. h. an jedem Zeitpunkt in dem untersuchten Stichprobenzeitraum, zu dem eine Prognose abgegeben wird, basiert die Modellauswahl allein auf den bis dahin bekannten Daten. Als Modellauswahlkriterium wird wie bei H&S das Bayesianische Informationskriterium (BIC) herangezogen. Dabei wird zu prüfen sein, ob sich die von H&S festgestellte Überlegenheit der indikatorgestützten Prognosen gegenüber den rein autoregressiven Prognosen unter diesen Bedingungen noch nachweisen lässt und welcher Indikator ggf. die bessere Prognose liefert. Im Gegensatz zu H&S finden Jörg Breitung und Doris Jagodzinsky (2001) für einen etwas anders abgegrenzten Untersuchungszeitraum, dass die Prognosefehlervarianz der beiden umfragebasierten Indikatoren sehr ähnlich ist, wobei die ifo-Geschäftserwartungen sogar noch etwas besser abschneiden als die ZEW-Konjunkturerwartungen.⁴

Eine Möglichkeit besteht darin, dass das von H&S gewählte Modellauswahlverfahren die ifo-Geschäftserwartungen benachteiligt. So spezifizieren H&S ihre Prognosemodelle für die Industrieproduktion indem sie zunächst ein optimales autoregressives Modell für die Industrieproduktion gestützt auf das BIC ermitteln und diese autoregressive Struktur wird dann für das bivariate Modell, dass die Industrieproduktion mit Hilfe des jeweiligen Indikators erklärt, übernommen. Die optimale Autoregressionstruktur kann im multivariaten Fall jedoch eine andere sein als im univariaten. Für den Fall, dass die für das univariate Modell gefundene autoregressive Struktur für das Modell auf der Basis der ZEW-Konjunkturerwartungen weniger suboptimal ist als für das Modell auf der Basis der ifo-Geschäftserwartungen, verzerrt das Modellauswahlverfahren den Vergleich der Prognosegüte zugunsten der ZEW-Geschäftserwartungen. Um diese Möglichkeit auszuschließen, wird in der vorliegenden Untersuchung die Spezifikation der bivariaten Modelle unabhängig von der Spezifikation der autoregressiven Modelle vorgenommen.

Außerdem berücksichtigt das hier verwendete Modellauswahlverfahren, dass sich aufgrund des approximativen Charakters der unterstellten Modelle Beobachtungen oder Perioden ergeben können, die nur ungenügend durch das jeweilige Modell beschrieben werden (vgl. Franses und Lucas 1998). Bei mangelnder Berücksichtigung können derartige „Ausreißer“ zu einer suboptimalen Modellauswahl sowie zu verzerrten Parameterschätzern führen und so die Prognosegüte des Modells beeinträchtigen. Sofern Prognosemodelle auf der Basis der ifo-Geschäftserwartungen stärker als jene auf Basis der ZEW-Konjunkturerwartungen von Ausreißern betroffen sind, könnte deren Vernachlässigung im Modellauswahlprozess bei H&S ebenfalls die relative Prognosegüte der Indikatoren beeinflusst haben.

2 Auswahl der Prognosemodelle

Die vorliegende Untersuchung verwendet als Ausgangspunkt den Datensatz von H&S (2002). Hierbei handelt es sich um die ifo-Geschäftserwartungen, die ZEW-Konjunkturerwartungen und die saisonbereinigte Industrieproduktion über den Zeitraum von Dezember 1991 bis September 2000. Die Prognosemodelle von H&S

⁴Auch H&S selbst können in ihrer Studie in der zweiten Hälfte ihres Untersuchungszeitraum keine signifikanten Unterschiede zwischen den beiden Indikatoren feststellen.

basieren auf Vektorautoregressionen für die Veränderungsrate der Industrieproduktion gegenüber dem Vorjahr, $\Delta_{12}IP_t$, und die jeweilige Indikatorvariable X (ifo-Geschäftserwartungen, ZEW-Konjunkturerwartungen). Wir modellieren dagegen die Veränderung der Industrieproduktion gegenüber dem Vormonat, ΔIP_t , da bei einer Spezifikation allein in zwölfsten Differenzen grundsätzlich nicht auszuschließen, dass die Residuen autokorreliert sind. Außerdem spezifizieren wir die Gleichung für die Indikatorvariable in Fehler-Korrektur Form. Da bei dieser Modellspezifikation die Verzögerungen in Veränderungsraten in die Gleichung eingehen, sollte das Multikollinearitätsproblem geringer und die damit Modellselektion besser sein. Insgesamt lautet unser Prognosemodell damit:

$$\Delta IP_t = a_1 + X_{t-1} + \sum_{i=1}^{P_1} \beta_{1i} \Delta IP_{t-i} + \sum_{j=1}^{P_2} \delta_{1j} \Delta X_{t-j} + \varepsilon_{1t} \quad (1)$$

$$\Delta X_t = a_2 + X_{t-1} + \sum_{k=1}^{P_3} \beta_{2k} \Delta X_{t-k} + \sum_{n=1}^{P_4} \delta_{2n} \Delta IP_{t-n} + \varepsilon_{2t} \quad (2)$$

Die Prognosegüte vektorautoregressiver Modelle hängt entscheidend von der Spezifikation der Verzögerungsstruktur ab. H&S wählen diese aus, indem sie in einem ersten Schritt für jede der drei Variablen ein autoregressives Modell auf Basis des Bayesianischen Informationskriteriums (BIC) bestimmen und auf Freiheit von Autokorrelation testen. Im zweiten Schritt halten sie diese autoregressive Verzögerungsstruktur konstant und fügen, wiederum unter Verwendung des BIC, Verzögerungen der jeweils anderen Systemvariable (also z.B. $\sum_{j=1}^{P_2} \delta_{1j} X_{t-j}$ in die Gleichung für $\Delta_{12}IP_t$) ein. Dabei ergibt sich in allen Fällen für die hinzugefügten Variablen eine optimale Verzögerung von eins.

Unser Modellauswahlprozess basiert ebenfalls auf der Minimierung des BIC-Kriteriums, allerdings ist die autoregressive Verzögerungsstruktur auch bei den Gleichungen für die Industrieproduktion variabel. Um den Rechenaufwand bei der Modellselektion zu verringern, stützen wir uns dabei auf ein „General-to-Specific“-Reduktionsverfahren, das auf der sukzessiven Eliminierung von Koeffizienten mit niedrigen empirischen t -Werten beruht. Brüggemann und Lütkepohl (2001) zeigen, dass bei geeigneter Wahl der kritischen t -Werte, dieses Verfahren zu der gleichen Verzögerungsstruktur führt wie eine Suche nach dem geringsten BIC-Wert über alle möglichen Kombinationen von Koeffizienten. Als Nebenbedingung gilt dabei – ähnlich wie bei H&S –, dass das auf diese Weise gefundene Modell autokorrelationsfreie Residuen aufweisen muss.⁵ Überdies berücksichtigt unser Modellauswahlprozess allerdings auch das Vorliegen von schlecht modellierten Beobachtungen. Derartige „Ausreißer“ können zu verzerrten Parameterschätzern und führen und die Prognoseleistung des Modells verringern, ihre Wirkungen sollten daher nach Möglichkeit eliminiert werden.⁶ Um eine Beobachtung als abweichend vom Rest der Stichprobe zu identifizieren, verwenden wir das sogenannte „studentisierten Residuum“, also die empirische t -Statistik einer Dummy-Variable, die den Wert 1 bei der betreffenden Beobachtung annimmt und 0 sonst. Diese studentisierten Residuen werden über den gesamten Stützzeitraum berechnet und das Maximum mit einem kritischen Wert verglichen, den wir in Anlehnung an Chen und Lui (1993) auf 2,8 festsetzen. Im Fall der Ablehnung der Nullhypothese, dass kein Ausreißer vorliegt, wird an der Stelle des absolut größten Residuums eine 0/1-Dummy eingeführt und die Suche erneut aufgenommen bis keine Ausreißer mehr gefunden werden. Anschließend wird eine

⁵Kilian (1999) stellt in einer Simulation fest, dass das BIC Modelle auswählt, die zu gering parametrisiert sind, um die Dynamik der unterstellten Prozesse korrekt abzubilden. Er empfiehlt für kleinere Stichproben die Verwendung des AIC. Er berücksichtigt allerdings nicht die Verwendung von Tests auf Freiheit von Autokorrelation als weiteres Modellsselektionskriterium.

⁶Vgl. Madalla (1992), Franses(1998).

Tabelle 1: Parametrisierung 1991:12-2000:09

	Konstante	X_{t-1}	ΔIP_{t-i}	ΔX_{t-j}	Dummy
Parametrisierung in dieser Untersuchung					
$\Delta IP_{Autoregress.}$	ja	-	1, 17, 20	-	01.95
ΔIP_{ifo}	ja	ja	1, 2	-	07.93, 01.95
ΔX_{ifo}	ja	ja	-	1, 6, 18	04.95, 09.98
ΔIP_{ZEW}	ja	ja	1, 2, 17	-	01.95, 01.97
ΔX_{ZEW}	ja	ja	1	1, 8	07.93, 03.95,01.96, 09.98, 09.2000

erneute Spezifikation der Verzögerungsstruktur des Modells auf der Basis des BIC vorgenommen.

Die Ergebnisse unseres Modellauswahlprozesses auf der Basis des von H&S zugrundegelegten Stützzeitraums von 1991:12 bis 2000:9 sind in Tabelle 1 dargestellt. Außerdem angegeben ist die ebenfalls mit dem beschriebenen Verfahren gefundene Spezifikation für das autoregressive Modell für die Industrieproduktion, das im folgenden ähnlich der Vorgehensweise von H&S als Vergleichsbasis für die Güte unseres Prognosemodells dienen soll. Es zeigt sich, dass das Modellauswahlverfahren in allen Gleichungen Ausreißer entdeckt. Die resultierenden Spezifikationen sind grundsätzlich sehr sparsam parametrisiert; insbesondere für Veränderung der Industrieproduktion ergeben sich Modelle, die mit weniger Koeffizienten auskommen, als die entsprechenden Gleichungen bei H&S. Dies kann zum einen an der Berücksichtigung von Ausreißern liegen, zum anderen aber auch auf den Verzicht auf die Spezifikation in zwölfsten Differenzen, der möglicherweise eine kompliziertere Lag-Struktur notwendig macht, um Freiheit von Autokorrelation zu erreichen.

3 Vergleich der Güte der Modellprognosen

Im folgenden wird die Prognosegüte der Modelle verglichen. Dabei werden rekursive Out-of-Sample-Prognosen für die Veränderungsrate der Industrieproduktion im Vorjahresvergleich über 1, 3, 6, 9 und 12 Monate berechnet. Dabei wird entweder die Modellspezifikationen, die auf Basis des gesamten Stichprobenzeitraums gewählt wurde verwendet oder die Spezifikation zu jedem Prognosezeitpunkt auf der Basis unseres Modellauswahlverfahrens neu bestimmt. Es versteht sich, dass bei der Berechnung der Prognosen die Koeffizienten der 0/1-Dummy-Variablen auf Null gesetzt sind. Als Maß zur Bewertung der Prognosegüte wird Theils U verwendet, also der Quotient aus der Wurzel aus dem mittleren quadratischen Prognosefehler (RMSE) des Prognosemodells und dem RMSE der autoregressiven Prognose (naive Prognose). Eine dem autoregressiven Modell überlegene Prognose weist dabei einen Wert kleiner als Eins auf. Außerdem wird die Modifikation des Diebold/Mariano-Tests auf Gleichheit der Prognosefehler von Harvey, Leybourne und Newbold (1997) berechnet, um zu erkennen, ob die Unterschiede zwischen den RMSE auch statistisch signifikant sind.

3.1 Prognosegüte bei konstanter Modellstruktur

Zunächst untersuchen wird die Prognosegüte der Modelle nach dem gängigen, auch von H&S verwendeten Verfahren bei dem die Modellstruktur konstant bleibt und

der für den gesamten Stichprobenzeitraum gefundenen entspricht. Es werden also die Spezifikationen aus Tabelle 1 verwendet.

Die für den Prognosezeitraum von Januar 1994 bis September 2000 ermittelten Ergebnisse sind in Tabelle 2 denen von H&S gegenübergestellt⁷. Wir stellen fest, dass sowohl das autoregressive Modell, als auch das für die ifo-Geschäftserwartungen deutlich geringere Prognosefehler aufweisen als die jeweiligen Modelle bei H&S. Für die Modelle auf Basis des ZEW-Indikators gilt dies nur für ein- und dreimonatige Prognosen; ab einem Prognosehorizont von 6 Monaten liefert die Modellspezifikation von H&S die niedrigeren RMSE. Allerdings gelingt es unseren indikatorgestützten Modellen nicht die Prognosegüte des autoregressiven Modell zu übertreffen. Die U-Statistiken liegen für jeden der fünf Prognosehorizonte über 1 und das marginale Signifikanzniveau des modifizierten Diebold/Mariano-Tests signalisiert, dass die Hypothese gleicher Prognosefehler nicht verworfen werden kann. In der Untersuchung von H&S ergaben sich dagegen signifikant geringere Prognosefehler für die Indikator gestützten Modelle für eine Reihe von Prognosehorizonten, insbesondere beim Modell auf der Basis des ZEW-Indikators.

Tabelle 2: RMSE und Theils U für den Zeitraum 1994:01–2000:09

Prognoseschritte	1	3	6	9	12
ifo	1,322	1,644	2,486	3,697	5,215
	1,01	1,00	1,09	1,24	1,53
	(0,936)	(0,988)	(0,686)	(0,437)	(0,290)
ZEW	1,329	1,700	2,824	4,193	5,643
	1,01	1,04	1,23	1,41	1,66
	(0,893)	(0,784)	(0,238)	(0,196)	(0,214)
Autoregressiv	1,314	1,641	2,291	2,971	3,402
ifo ¹ _{H&S}	1,736	2,111	2,933	4,730	7,405
	0,92	0,84	0,72	0,85	1,12
	(0,224)	(0,142)	(0,077)	(0,219)	(0,644)
ZEW ¹ _{H&S}	1,803	2,004	2,702	3,303	3,724
	0,95	0,80	0,66	0,60	0,56
	(0,361)	(0,058)	(0,048)	(0,069)	(0,095)
Autoregressiv ¹ _{H&S}	1,889	2,511	4,088	5,540	6,609

Für jeden Indikator geben die Werte in der ersten Zeile die Wurzel aus dem mittleren quadratischen Prognosefehler (RMSE) für den jeweiligen Prognoseschritt, die zweite Zeile der Quotient aus dem RMSE des Modells und dem RMSE der autoregressiven Prognose (Theils U) und die Werte in Klammern darunter das marginale Signifikanzniveau des modifizierten Diebold-Mariano-Tests an.

¹Nach Hüfner und Schröder (2002), Tabelle 4

3.2 Prognosegüte bei rekursiv bestimmter Modellstruktur

Tabelle 3 zeigt die Prognoseergebnisse der Modelle bei rekursiver Auswahl der Modellstruktur. Die Modellspezifikation wurde dabei für jeden Monat im untersuchten Zeitraum von Januar 1994 bis September 2000 neu bestimmt wurde. Ferner ist berücksichtigt, dass die Konjunkturindikatoren etwa einen Monat vor der Industrieproduktion des jeweiligen Monats bekannt gegeben werden.

⁷Angegeben sind jeweils RMSE und Theil' s U, sowie das Ergebnis des Diebold/Mariano-Tests.

Zunächst fällt auf, dass die Prognosefehler aller Modelle deutlich höher sind als bei konstanter Modellstruktur. Dies ist ein Indiz dafür, dass eine Untersuchung der Prognosegüte mit einer konstanten, auf Basis der gesamten Stichprobeninformation festgelegten Modellspezifikation, den Prognosefehler unterschätzt, da sie zwar die mit der Parameterschätzung, nicht jedoch die mit der Modellauswahl verbundene Unsicherheit berücksichtigt. Es gelingt auch hier nicht, mit Hilfe der Indikatoren die Veränderung der Industrieproduktion im Vorjahresvergleich signifikant besser zu prognostizieren als mit dem autoregressiven Modell.

Tabelle 3: RMSE und Theils U für den Zeitraum 1994:01–2000:09

Prognoseschritte	1	3	6	9	12
ifo	1,458	1,844	2,677	3,707	4,489
	0,98	0,98	0,93	0,94	0,96
	(0,676)	(0,765)	(0,553)	(0,527)	(0,696)
ZEW	1,558	1,877	2,826	3,979	5,214
	1,05	1,00	0,99	1,01	1,12
	(0,364)	(0,987)	(0,933)	(0,978)	(0,503)
Autoregressiv	1,483	1,880	2,867	3,957	4,674
Diebold/Mariano-Test					
ifo vs. ZEW	0,126	0,766	0,540	0,558	0,116

Den Berechnungen von H&S zufolge ist das Prognosemodell auf Basis der ZEW-Erwartungen dem Modell auf Basis der ifo-Erwartungen überlegen; lediglich für 1-Schritt-Prognose weist das ifo-Modell in ihrer Untersuchung einen etwas geringeren Prognosefehler aus als das ZEW-Modell. Diesen Schluss lässt unsere Analyse nicht zu, vielmehr zeigt sich, dass das von uns ermittelte Prognosemodell auf Basis der ifo-Geschäftserwartungen durchgängig niedrigere Prognosefehler liefert als das Modell auf Basis der ZEW-Konjunkturerwartungen; einen signifikanten Unterschied in der Prognosegüte gibt es jedoch nicht.

3.3 Variationen des Prognosezeitraums

In einem weiteren Schritt haben wir den Untersuchungszeitraum ausgeweitet. Eine genauere Untersuchung der Prognosefehler der Modelle ergibt, dass diese für die Jahre 1994 und 1995 deutlich höher sind als für den späteren Zeitraum. Dies ist nicht überraschend, denn da der Datensatz erst im Dezember 1991 beginnt, stützen sich die Parameterschätzwerte der Modelle zunächst nur auf sehr wenige Beobachtungen, was die Prognosegenauigkeit beeinträchtigt. Um zu vermeiden, dass die Ergebnisse unserer Untersuchung dadurch verzerrt werden, etwa weil der sehr kurze Stützzeitraum zu Beginn der Prognosen Modelle mit wenigen Parametern stark begünstigt, soll nun der Prognosezeitraum erst ab Januar 1996 beginnen. Dafür kommen die mittlerweile verfügbaren Werte bis Oktober 2002 hinzu, die nicht zuletzt deshalb interessant sind, weil es im Jahr 2001 zu einer Rezession in Deutschland kam, der zweiten im gesamten untersuchten Zeitraum. Ein Vergleich der Prognosefehler in diesem Zeitraum mit jenem im zuvor analysierten Zeitraum kann Auskunft darüber geben, wie es den Indikatoren gelang, die rezessive Entwicklung frühzeitig anzuzeigen. Die Modellsélection erfolgt abermals rekursiv.

Ergänzt wird die Analyse durch den von der Zeitschrift "Wirtschaftswoche" in Zusammenarbeit mit der "Commerzbank" veröffentlichte "Earlybird"-Konjunkturindikator. In der Untersuchung von Breitung und Jagodzinski (2001) stellte dieser

Indikator hinsichtlich der Prognosegüte sowohl den ZEW- Konjunkturerwartungen als auch den ifo-Geschäftserwartungen in den Schatten. Da dieser Indikator erst seit 2001 veröffentlicht wird, haben wir eine eigene Rückberechnung bis 1991 vorgenommen.⁸ Die Ergebnisse der Analyse sind in Tabelle 4 abgebildet.

Tabelle 4: RMSE und Theil's U für den Zeitraum 1996:01–2002:10

Prognoseschritte	1	3	6	9	12
ifo	1,422	1,538	2,059	2,433	2,910
	1,11	0,95	0,87	0,78	0,77
	(0,055)	(0,448)	(0,213)	(0,077)	(0,91)
ZEW	1,421	1,630	2,386	2,981	3,531
	1,12	1,00	1,01	0,95	0,94
	(0,061)	(0,965)	(0,939)	(0,806)	(0,701)
Ebird	1,371	1,621	2,308	3,021	3,628
	1,08	1,00	0,98	0,97	0,96
	(0,307)	(0,993)	(0,904)	(0,880)	(0,901)
Autoregressiv	1,272	1,623	2,356	3,122	3,769

Es ist ersichtlich, dass der RMSE für drei- bis zwölfmonatige Prognosen geringer ist als im Prognosezeitraum 1994:1-2000:9. Anders als für jenen Zeitraum, ist der Prognosefehler auf Basis des ifo-Modells jetzt ab der dreimonatigen Prognose kleiner als der des autoregressiven Modells. Für den Earlybird-Indikator ist dies ab sechs Monaten und für den ZEW-Indikator ab neun Monaten der Fall. Lediglich die ifo-Geschäftserwartungen liefern jedoch für neunmonatige Prognosen signifikant bessere Ergebnisse als das naive Modell.

Im direkten Vergleich der Prognosefehler der drei Indikator-Modelle zeigen sich allerdings kaum signifikante Differenzen (Tabelle 5). Über einen Prognosehorizont von 12 Monaten ergibt sich eine Überlegenheit des Modells auf Basis der ifo-Geschäftserwartungen gegenüber dem ZEW-Modell. Insgesamt sind die Unterschiede in der Prognosegüte der Modelle aber gering. Eine Überlegenheit des ZEW-Indikators gegenüber dem ifo-Indikator bei der Prognosefähigkeit, wie sie H&S feststellen können wir nicht belegen. Das Resultat von Breitung und Jagodzinski (2001)⁹, nach dem der Earlybird deutlich bessere Prognosen liefert als die übrigen Indikatoren, können wir ebenfalls nicht bestätigen. Immerhin vermag er aber gut mit den rein umfragebasierten Indikatoren mitzuhalten. Überdies decken sich die Ergebnisse mit denen von Breitung und Jagodzinsky insofern, als sie eine sehr ähnliche Prognosegüte von ZEW- und ifo-Indikator feststellen, wobei der mittlere Fehler für das Modell auf Basis der ifo-Geschäftserwartungen geringer ausfällt als für das Modell auf Basis der ZEW-Konjunkturerwartungen.

Schließlich führen wir eine weitere Variation des Prognosezeitraums durch. Hintergrund dieser Analyse ist, dass die Daten für die ifo-Geschäftserwartungen in der gegenwärtig verwendeten Branchenabgrenzung schon seit Februar 1968 vorliegen. Es liegt nahe, den gesamten zur Verfügung stehenden Stützzeitraum für die Modellselektion und die Schätzung der Modellparameter zu verwenden. Der Earlybird-index kann bis Februar 1973 rekonstruiert werden, für frühere Zeiträume fehlen die Daten des von der Deutschen Bundesbank berechneten Indikators der preislichen Wettbewerbsfähigkeit der deutschen Wirtschaft. Da der ZEW-Index noch nicht so

⁸In den Earlybird-Indikator gehen gewichtet ein der kurzfristige Realzins (50 Prozent), der Indikator der preislichen Wettbewerbsfähigkeit der deutschen Wirtschaft (25 Prozent) sowie der Einkaufsmanagerindex für das verarbeitende Gewerbe in den Vereinigten Staaten (25 Prozent).

⁹Allerdings betrachten die Autoren nur den rechten kurzen Zeitraum von Anfang 1999 bis Mitte 2001 und ermitteln nicht die statistischen Signifikanz der von ihnen festgestellten Unterschiede zwischen den Prognosefehlern.

Tabelle 5: Diebold/Mariano-Test 1996:01-2002:10

	1	3	6	9	12
ifo vs. ZEW	0,866	0,567	0,198	0,119	0,032
ifo vs. Ebird	0,496	0,646	0,301	0,234	0,445
ZEW vs. Ebird	0,590	0,979	0,827	0,941	0,909

lange erhoben wird, ist für diesen Indikator keine Ausweitung des Schätzzeitraums möglich.

Der Prognosezeitraum beginnt für beide Indikatoren wiederum im Januar 1996. Es zeigt sich, dass sich die Prognoseeigenschaften der Konjunkturindikatoren erheblich verbessern. Die Prognosefehler sind für alle drei Modelle deutlich besser als unter Verwendung des kürzeren Stützzeitraums, die der indikatorgestützten Modell verringern sich jedoch stärker als die des autoregressiven Modells. Für Prognosen über einen bis sechs Monate verfügen die ifo-Geschäftserwartungen über signifikant bessere Prognoseeigenschaften als das autoregressive Modell (Tabelle 6). Der Earlybird führt ab drei Monaten zu Prognosen, die dem naiven Modell deutlich überlegen sind. Eine Verlängerung des Stützzeitraums führt offenbar zu einer deutlichen Verbesserung der Prognoseeigenschaften der Indikatoren.

Tabelle 6: RMSE und Theil's U für einen langen Zeitraum

Prognoseschritte	1	3	6	9	12
02.1968-10.2002					
ifo	0,982	1,165	1,580	2,196	2,893
	0,93	0,83	0,75	0,78	0,87
	(0,080)	(0,046)	(0,053)	(0,125)	(0,333)
Autoregressiv	1,052	1,405	2,113	2,815	3,325
02.1973-10.2002					
Ebird	1,047	1,280	1,775	2,301	2,745
	1,00	0,93	0,86	0,82	0,82
	(0,935)	(0,019)	(0,014)	(0,014)	0,030)
Autoregressiv	1,045	1,383	2,068	2,791	3,336

4 Fazit

Kerngedanke des vorliegenden Papiers ist, dass Untersuchungen zur Prognosegüte nicht nur Prognosefehler, die auf der Schätzung der Parameter beruhen, berücksichtigen sollten, sondern auch solche, die aus der stichprobenabhängigen Auswahl des Prognosemodells resultieren. Sollen die Prognosefehler, wie es gängige Praxis ist, durch rekursive Out-of-Sample Prognosen geschätzt werden, so ist es unserer Ansicht nach wichtig, dass dabei nicht nur die Parameterschätzung, sondern auch die Modellselektion rekursiv vorgenommen wird.

Wir wenden dieses Prinzip auf die Analyse der Prognosegüte dreier wichtiger Indikatoren für die Konjunktur in Deutschland an. Es zeigt sich, dass die Prognosefehler bei rekursiver Modellauswahl wie erwartet größer sind als bei konstanter Modellstruktur. Hinsichtlich der Frage, ob die betrachteten Indikatoren bessere Prognosen liefern als ein einfaches autoregressives Modell, sind die Ergebnisse abhängig von dem für die Modellselektion und die Parameterschätzung zur Verfügung stehenden Stützzeitraum. Für die ifo-Geschäftserwartungen und den Earlybird-Konjunkturindikator, für die lange Reihen vorliegen, läßt sich zeigen, dass sie die Industrieproduktion signifikant besser zu prognostizieren vermögen, wenn die Modellselektion und -schätzung auf einem langen Stützzeitraum beruht. Werden dagegen nur kurze Zeitreihen für die Modellauswahl verwendet, so sind beide Indikatoren, wie auch die ZEW-Konjunkturerwartungen, nicht in der Lage, bessere Prognosen zu erzeugen als das autoregressive Modell.

Literatur

- [1] Breitung, J. und D. Jagodzinski (2001). Prognoseeigenschaften alternativer Indikatoren für die Konjunkturentwicklung in Deutschland. *Konjunkturpolitik*, 47. Jg. H.4, 292-314.
- [2] Chen, C. und L.-M. Lui (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association* 88, 284-297.
- [3] Diebold, F.X. und R.S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13, 253-263.
- [4] Döpke, J., J.W. Krämer und E. Langfeldt (1994). Konjunkturelle Frühindikatoren in Deutschland. *Konjunkturpolitik*, 40, 133-153.
- [5] Franses, P.H. (1998). *Time Series Models for Business and Economic Forecasting*. Cambridge U.K.
- [6] Fritsche, U. (1999). Vorlaufeigenschaften von ifo-Indikatoren für Westdeutschland. *DIW-Diskussionspapier*, 179.
- [7] Fritsche, U. (2001). Do Probit Models Help in Forecasting Turning Points in German Business Cycles? *DIW-Diskussionspapier*, 241.
- [8] Fritsche, U. und S. Stephan (2000). Leading Indicators of German Business Cycles: An Assessment of Properties. *Jahrbücher für Nationalökonomie und Statistik*, Band 222, 3, 289-315.
- [9] Funke, N. (1997). Predicting Recessions: Some Evidence for Germany. *Weltwirtschaftliches Archiv*, 133, 91-102.
- [10] Hansen, B. E. (1999). Discussion of 'Data Mining Reconsidered', *Econometrics Journal* 2, 192-201.
- [11] Hüfner, F.P. und M. Schröder (2002). Prognosegehalt von ifo-Geschäftserwartungen und ZEW-Konjunkturerwartungen: Ein ökonometrischer Vergleich. *Jahrbücher für Nationalökonomie und Statistik*, Band 222, 3, 316-336.
- [12] Harvey, D.I., S.J. Leybourne und P. Newbold (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, Vol. 13, 281-291.
- [13] Kilian, L. (2001). Impulse Response Analysis in Vector Autoregressions with Unknown Lag Order. *Journal of Forecasting* 20, 161-179.
- [14] Langmantel, E. (1999). Das ifo-Geschäftsklima als Indikator für die Prognose des Bruttoinlandsproduktes. *ifo Schnelldienst*, 16/17, 16-21.
- [15] Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg, New York, Tokyo.
- [16] Maddala, G. S. (1992). *Introduction to Econometrics*. Second Edition, New York.
- [17] Meese, R. A. und K. Rogoff (1983). Empirical Exchange Rate Models of the Seventies: Do they Fit out of Sample? *Journal of International Economics* 14, 3-24.

- [18] Solveen, R. (2001). Konjunktur vor der Wende? Deutschland: Konjunktur aktuell, Commerzbank, Juli/August.
- [19] Brüggemann, R. und H. Lütkepohl (2001). Lag Selection in Subset VAR Models with an Application to a U.S. Monetary System. In: Econometric Studies: A Festschrift for Joachim Frohn, Münster, S. 107-128.