

Osterloh, Steffen

Article

Wer prognostiziert die deutsche Konjunktur am besten?

ZEW Wachstums- und Konjunkturanalysen

Provided in Cooperation with:

ZEW - Leibniz Centre for European Economic Research

Suggested Citation: Osterloh, Steffen (2006) : Wer prognostiziert die deutsche Konjunktur am besten?, ZEW Wachstums- und Konjunkturanalysen, Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim, Vol. 9, Iss. 3, pp. 8-9

This Version is available at:

<https://hdl.handle.net/10419/125968>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Wer prognostiziert die deutsche Konjunktur am besten?

Die Erstellung von Konjunkturprognosen gehört zu den Tätigkeiten von Volkswirten, die am meisten im Blickpunkt der Öffentlichkeit stehen. Neben öffentlichen Institutionen und internationalen Organisationen sind insbesondere Wirtschaftsforschungsinstitute sowie Banken als Prognostiker tätig. In diesem Beitrag werden die kurzfristigen Prognosen für das deutsche Bruttoinlandsprodukt von 22 Instituten und Banken seit 1995 hinsichtlich der Frage analysiert, ob es Unterschiede in der Prognosegüte zwischen den verschiedenen Institutionen gibt.

Es liegen zahlreiche Untersuchungen vor, die die Qualität der kurzfristigen Prognosen der Wachstumsraten des deutschen Bruttoinlandsproduktes (BIP) untersuchen. Die aktuellste Untersuchung dazu liefern Döpke und Fritsche (2005), die einen über 30 Jahre zurückreichenden Datensatz bestehend aus Prognosedaten von 14 Institutionen analysieren. Betrachtet werden hier Konjunkturprognosen von Wirtschaftsforschungsinstituten, öffentlichen Institutionen und internationalen Organisationen. Dabei zeigen die Autoren ein Problem auf, das bei derartigen Untersuchungen anhand der von den Prognostikern veröffentlichten Werte häufig auftritt: Viele Institutionen veröffentlichen ihre Prognosen zu vorher festgelegten

re Modelle oder Fähigkeiten der Prognostiker zu erklären, wenn sie sich im Veröffentlichungszeitpunkt unterscheiden.

Die Daten

Um dieser Problematik zu begegnen, nutzt die vorliegende Untersuchung die in der Zeitschrift „Consensus Forecasts“ veröffentlichten Prognosen. Dabei handelt es sich um eine monatliche Umfrage des privaten britischen Institutes Consensus Economics, die seit 1989 erhoben und momentan für etwa 70 Länder durchgeführt wird. Die teilnehmenden Prognostiker werden jeweils zu einem gemeinsamen Stichtag hinsichtlich ihrer Prognosen für eine Vielzahl volkswirtschaftlicher Variablen für das aktuelle und das folgende Jahr befragt, wobei eine klare Definition der untersuchten Größen eine hohe Vergleichbarkeit der Prognosen sicherstellt. Somit liegen für jedes zu prognostizierende Jahr insgesamt 24 Prognosen vor: 12 Ein-Jahres- und 12 Zwei-Jahresprognosen.

Eine weitere Besonderheit der „Consensus Forecast“-Veröffentlichungen besteht in der Zusammensetzung ihrer Teilnehmer. So liegt hier der Schwerpunkt neben den in anderen Untersuchungen ausführlich analysierten Forschungsinstituten bei kommerziellen Prognostikern, hauptsächlich Banken.

Zusätzlich zu den individuellen Prognosen aller Teilnehmer wird auch der arithmetische Mittelwert für jede Variable ausgewiesen. Dieser Wert, der in der Öffentlichkeit als „Consensus“ bekannt ist, wird häufig von Prognostikern als Benchmark in Verbindung mit dem eigen-

nen Wert veröffentlicht, um die eigene Position im Verhältnis zu den übrigen Prognostikern aufzuzeigen.

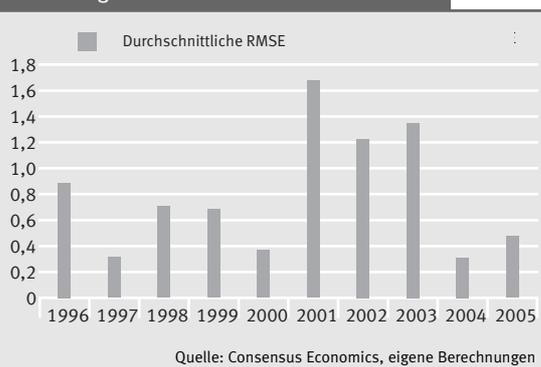
Der Test

Zur Beantwortung der Frage, ob sich die Prognosegenauigkeit der verschiedenen Prognostiker signifikant unterscheidet, schlägt Stekler (1987) einen Rangsummentest vor. Er zieht diesen einer Untersuchung der Prognosefehler, bspw. mittels eines F-Testes, vor, da nicht alle Jahre mit gleicher Präzision zu prognostizieren sind. Bei einer Untersuchung auf Basis der Fehlersumme über die untersuchten Jahre würden automatisch die schwer zu prognostizierenden Jahre mit einer höheren Varianz der Fehler die Untersuchung dominieren. Abb. 1 zeigt dieses für die durchschnittlichen Root Mean Squared Error (RMSE)¹, die die Consensus-Teilnehmer in den letzten 10 Jahren für das deutsche BIP begangen haben. Es zeigt sich, dass gerade in den Jahren der Stagnation von 2001 bis 2003 die Fehler außergewöhnlich hoch waren.

Der von Stekler (1987) eingeführte nichtparametrische Test berücksichtigt dagegen allein die Positionen einer Rangfolge, die auf Basis der Fehlerhöhe für jedes Jahr erstellt wird. Der Test wird wie folgt durchgeführt: Zuerst wird für jeden Prognostiker für jedes zu prognostizierende Jahr der RMSE berechnet. Auf Basis dieser Werte wird anschließend jedem Prognostiker ein Rang zugeordnet, beginnend mit demjenigen mit dem geringsten Fehler, welcher den Rang 1 erhält, bis zum höchsten Fehler, dem der letzte Rang zugewiesen wird.

Dieses Vorgehen mittels der RMSE erscheint für den vorliegenden Datensatz jedoch nicht angebracht, da dieser viele fehlende Beobachtungen aufweist. So geben manche Teilnehmer ihre erste Prognose für ein Zieljahr regelmäßig nicht im Januar des vorherigen Jahres ab, sondern erst einige Monate später. Da jedoch der RMSE in der Regel im Zeitverlauf abnimmt, würde eine Bewertung anhand

Abbildung 1: Durchschnittliche RMSE



Zeitpunkten, die sich selten überschneiden. Da der Prognosefehler im Zeitverlauf jedoch in der Regel abnimmt, zeigen diejenigen Prognosen geringere Fehler, die eine späte Veröffentlichung aufweisen und somit bevorzugt werden. Daher ist es kaum möglich, signifikante Unterschiede der Prognosegüte durch besse-

¹ Eine Übersicht über die verwendeten statistischen Maße findet sich im Beitrag „Mittelfristige gesamtwirtschaftliche Projektionen“ auf Seite 6.

Tabelle 1: Positionen in den Rangfolgen

	BfG Bank/ SEB	MM Warburg	HSBC Trinkaus	...	Nr. 20	Nr. 21	Nr. 22
1995	9	3	4		17	19	22
1996	12	2	11		21	22	19
1997	12	10	13		6	9	18
1998	3	10	4		20	19	22
1999	3	11	21		22	19	8
2000	11	5	9		6	22	8
2001	8	9	1		15	18	22
2002	4	12	1		14	6	3
2003	3	8	2		15	1	19
2004	8	6	20		15	17	19
2005	8	10	1		12	19	21
Summe	81	86	87	...	163	171	181
Erwartete Summe*	126,5	126,5	126,5		126,5	126,5	126,5

* siehe Text für Details

der RMSE diejenigen Prognostiker bevorzugen, die weniger Prognosen mit einem längeren Horizont abgeliefert haben.

Stattdessen sind die Rangfolgen mittels einer Variation des Theil's U gebildet worden. Dabei wurde der RMSE nicht in Relation zum RMSE eines naiven Prognosemodells gesetzt, sondern in Relation zum RMSE der Consensus-Prognosen. Für den Consensus RMSE wurden jedoch nur jeweils diejenigen Beobachtungen berücksichtigt, für die auch eine Beobachtung des jeweiligen Prognostikers verfügbar war. So kann das Ergebnis als die relative Genauigkeit im Vergleich zum Consensus interpretiert werden. Ein Prognostiker, der in den Monaten, in denen er teilgenommen hat, genauso gut wie der Consensus abgeschnitten hat, erreicht somit einen Wert von 1. Die beste Prognose erreicht den geringsten Wert. Ihr wird deshalb für das jeweilige Jahr der Rang 1 zugeordnet. Aus diesen Werten lässt sich für jeden Prognostiker i die Rangsumme als $r_i = \sum_{t=1}^T r_{it}$ berechnen, also der Summe der Ränge über alle prognostizierten Jahre T .

Die folgende Untersuchung der Treffsicherheit der deutschen Prognostiker umfasst die Ein-Jahresprognosen für die Jahre von 1995 bis 2005 und die Zwei-Jahresprognosen für 1996 bis 2005. Dieses Vorgehen erfordert jedoch, dass die Prognostiker während des gesamten Zeitraums teilgenommen haben müssen. Jedoch zeigt es sich, dass die Zusammensetzung der Teilnehmer, in der Regel etwa 30 jeden Monat, im Zeitverlauf stark variiert hat. So sind einige Insti-

tutionen ausgeschieden, bspw. durch Fusionen oder Übernahmen, andere sind neu hinzugekommen. Dieses führt dazu, dass der Datensatz auf 22 Prognostiker reduziert werden musste. Dieser spiegelt jedoch weiterhin eine große Varietät der Teilnehmer wider: So enthält er 4 öffentliche und 1 privates Forschungsinstitut, 3 Großbanken, 2 genossenschaftliche und 5 öffentlich rechtliche Banken, 3 Privatbanken, 3 Töchter ausländischer Banken und eine internationale Investmentbank. Als realisierte Werte wurden die in der jeweiligen Juni-Ausgabe „Consensus Forecast“ des Folgejahres veröffentlichten Realisationen verwendet.

Tabelle 1 zeigt die Ränge der auf Basis der Rangsummen ermittelten drei besten und drei schlechtesten Prognostiker, unter Berücksichtigung sowohl der Ein- als auch der Zwei-Jahresprognosen, also aller 24 Prognosehorizonte.

Es lässt sich beobachten, dass manche Prognostiker häufig vordere Plätze belegen während andere regelmäßig schlechter abschneiden. Dass dies nicht zufällig ist, soll nun statistisch abgesichert werden. Die Nullhypothese, dass alle Prognostiker gleich sind, bedeutet, dass bei Betrachtung der Ränge über mehrere Perioden keine systematischen Unterschiede in der Rangverteilung auftreten sollten. Mit anderen Worten: Der durchschnittliche Rang eines Prognostikers über T Perioden sollte dem theoretischen Durchschnittsrank entsprechen, der im vorliegenden Fall bei $n=22$ Teilnehmern dem Wert 11,5 entspricht. Empirisch wird die Nullhypothese, dass die Rangsum-

men r_i der Teilnehmer i ($i=1,\dots,n$) der Summe der Durchschnittsränge $T(n+1)/2$ entsprechen, mit folgender χ_{n-1}^2 -verteilter Teststatistik von Batchelor (1990) überprüft:

$$f = \sum_{i=1}^n \frac{\{r_i - T(n+1)/2\}^2}{Tn(n+1)/12}$$

Der Zähler dieser Statistik gibt die Abweichung der jeweiligen Rangsumme von ihrem unter der Nullhypothese erwarteten Wert an. Im Nenner wird die Varianz einer individuellen Rangstatistik ($n(n+1)/12$) für die Summe von T individuellen Rängen verwendet.

Ergebnisse

Die Ergebnisse des Tests finden sich in Tabelle 2. Es sind drei Teststatistiken berechnet worden: Neben der Betrachtung aller 24 Prognosehorizonte ist der Test auch jeweils für die Ein-Jahres-, sowie die Zwei-Jahresprognosen durchgeführt worden. Es zeigt sich, dass die Betrachtung beider Jahre die höchste Signi-

Tabelle 2: Ergebnisse

Gesamter Zeitraum	Ein-Jahres- Prognosen	Zwei-Jahres- Prognosen
40,41***	34,85**	29,99*

***, **, *: Signifikanz zum 1 v.H.-, 5 v.H.-, 10 v.H.-Niveau; Teststatistik ist χ^2 -verteilt mit 21 Freiheitsgraden

fikanz aufweist: Die Nullhypothese, dass die Prognosegüte aller Teilnehmer gleich ist, kann auf dem 1 v.H.-Niveau verworfen werden. Für die Ein-Jahresprognosen kann die Nullhypothese auf dem 5 v.H.-Niveau verworfen werden, für die Zwei-Jahresprognose immerhin noch auf dem 10 v.H.-Niveau.

Abschließend kann festgestellt werden, dass die gefundenen unterschiedlichen Positionen in den Rangfolgen nicht zufällig sind. Nicht alle Prognostiker sind gleich, manche von ihnen haben signifikant besser abgeschnitten als es eine zufällige Verteilung der Ränge erwarten ließe und somit eine höhere Prognosegenauigkeit aufgewiesen als andere Teilnehmer im Consensus Forecast.

Steffen Osterloh, osterloh@zew.de

Literatur

- Batchelor, R. A. (1990), All forecasters are equal, *Journal of Business and Economic Statistics*, 8 (1), 143-144.
- Döpke, J. und U. Fritsche (2006), Growth and Inflation Forecasts for Germany. A Panel-based Assessment of Accuracy and Efficiency, *Empirical Economics* 31 (3), 777-798.
- Stekler, H. O. (1987), Who forecasts better?, *Journal of Business and Economic Statistics*, 5 (1), 155-158.