

Die Steuerschätzung in Deutschland – eine Erfolgsgeschichte?

Robert Lehmann*

Einleitung

Im Zeitraum vom 4. bis 6. Mai 2010 fand die 136. Sitzung des ARBEITSKREISES „STEUERSCHÄTZUNGEN“ (AKS) in Lübeck statt. Im Mittelpunkt stand dabei die mittelfristige Schätzung der Steuereinnahmen für die Jahre 2010 bis 2014. Nach Angaben des BUNDESMINISTERIUMS DER FINANZEN werden die Steuereinnahmen in diesem sowie den kommenden Jahren insgesamt um rund 39 Mrd. € unter den Schätzergebnissen des vergangenen Jahres liegen. Das öffentliche Interesse an den Vorhersagen war in diesem Jahr unter anderem auch deshalb besonders groß, weil die Bundesregierung hiervon ihr weiteres Vorgehen in der Steuerpolitik abhängig gemacht hatte.

Doch auch unabhängig hiervon kommt den Vorhersagen des AKS große politische Bedeutung zu, weil deren Ergebnisse „eins zu eins“ in den Bundeshaushalt bzw. die mittelfristige Finanzplanung übernommen werden [BUNDESMINISTERIUM DER FINANZEN (2005); S. 11]. Zudem erfordern gesetzliche Verankerungen bei der Aufstellung des Haushaltsplans die Benennung aller zu erwartenden Einnahmen. Diese sind wegen ihrer Abhängigkeit von der weiteren wirtschaftlichen Entwicklung naturgemäß schwerer zu planen als die Ausgaben¹ [vgl. VON DER LIPPE (1986); S. 334], was eine besonders hohe Prognosequalität erforderlich macht, um den zuvor genannten Ansprüchen gerecht zu werden. Vor diesem Hintergrund wird im vorliegenden Artikel untersucht, wie treffsicher bzw. unverzerrt die Prognosen des AKS in der Vergangenheit waren und ob die Ergebnisse den Anforderungen gerecht werden.

Methodik

Im Rahmen dieses Beitrags werden anhand einer Treffsicherheitsanalyse die Ergebnisse des AKS evaluiert. Der untersuchte Zeitraum erstreckt sich über die Jahre 1973 bis 2008, wobei eine Einschätzung der Vorhersagequalität für zwei Prognosehorizonte durchgeführt wird. Die zu den verschiedenen Zeitpunkten veröffentlichten Schätzergebnisse werden den tatsächlichen Steuereinnahmen gegenübergestellt. Dabei ist darauf hinzuweisen, dass die Steuerschätzung immer von geltendem Recht ausgeht und nur jene Änderungen berücksichtigt, die bereits beschlossen sind. Insofern führen zwischenzeitlich eingetretene Steuerrechtsänderungen zu Abweichungen von Prognose- und Istwert. Somit ist die Bezeichnung „Treffsicherheitsanalyse“

– die eine periodengerechte Bereinigung um alle Rechtsänderungen erfordert – nicht ganz zutreffend; vielmehr wäre eher von einer „Abweichungsanalyse“ zu sprechen [vgl. KÖRNER (1983); S. 240]. Zudem erweist es sich bei einer Steuerrechtsänderung als schwierig, verschiedene, sich überlagernde Reaktionen² der Steuerpflichtigen zu quantifizieren [vgl. GEBHARDT (2001); S. 132].

Der AKS legt jeweils zweimal im Jahr (im Mai und im November) eine Prognose vor. Für die hier vorgenommene Analyse wurden jeweils die Ergebnisse der Sitzung im Mai herangezogen, da deren Prognose die Grundlage für die mittelfristige Finanzplanung, den Haushaltsentwurf des darauf folgenden Jahres und die Finanzplanung der einzelnen Gebietskörperschaften bildet [vgl. BUNDESMINISTERIUM DER FINANZEN (2005); S. 11]; als Prognosevariable wird das Gesamtsteueraufkommen gewählt. Die verwendeten Daten stammen aus Publikationen des BUNDESMINISTERIUMS DER FINANZEN und beinhalten alle für den Betrachtungszeitraum notwendigen Informationen über Schätzergebnisse und Realisationen [vgl. BUNDESMINISTERIUM DER FINANZEN (2010a) und (2010b)]. Zur Beurteilung der Güte von Prognosen stehen deskriptive Maße sowie statistische Testverfahren und regressionsanalytische Methoden zur Verfügung, welche in Box 1 näher beschrieben werden.³

Box 1: Prognosefehlermaße, Mincer-Zarnowitz-Regression und t-Test

Die hier verwendeten deskriptiven Maße basieren auf dem relativen Prognosefehler RPF_n , der sich für das betrachtete Jahr n ergibt aus der Differenz zwischen geschätztem ($\hat{T}_{n|v}$) sowie realisiertem (T_n) Steueraufkommen (jeweils in Relation zu den tatsächlichen Steuereinnahmen⁴).

$$RPF_n = \frac{\hat{T}_{n|v} - T_n}{T_n} \quad \forall v = n - 1, n$$

Dabei ist die Variable der geschätzten Steuereinnahmen ($\hat{T}_{n|v}$) folgendermaßen zu interpretieren: Sie bezeichnet die Prognose des Steueraufkommens für das Jahr n , welche entweder im Mai des letzten ($v = n - 1$) oder des aktuellen Jahres ($v = n$) erstellt wurde. Bei einem positiven Vorzeichen wurde überschätzt, ansonsten war die Einschätzung zu pessimistisch. Ein sehr häufig verwendetes Maß zur Messung

* Robert Lehmann ist Doktorand am ifo Institut, Niederlassung Dresden.

der Treffsicherheit ist der mittlere relative Prognosefehler *MRPF*.

$$MRPF = \frac{1}{N} \sum_{n=1}^N \frac{\hat{T}_{n|v} - T_n}{T_n}$$

Mit dieser Größe kann die mittlere relative Abweichung und die durchschnittliche Tendenz der Vorhersage gemessen werden. Im Rahmen dieser Analyse bezeichnet *N* die Anzahl der Prognosewerte. Das Problem an diesem Maß ist die Saldierung von positiven und negativen Werten über die Zeit; dennoch findet es speziell beim t-Test seine Verwendung. Dieser überprüft, ob der *MRPF* signifikant von Null verschieden ist. Exakte Ergebnisse werden aber nur dann geliefert, wenn die Vorhersagefehler einer Normalverteilung⁵ folgen. Ist es nicht möglich die zu testende Hypothese zu verwerfen, dann kann von Unverzerrtheit gesprochen werden.

Um dem Problem der Saldierung zu entgehen, dient der mittlere relative absolute Prognosefehler *MRAPF* zur Messung der mittleren prozentualen Entfernung zwischen Prognose und Realisation.

$$MRAPF = \frac{1}{N} \sum_{n=1}^N \left| \frac{\hat{T}_{n|v} - T_n}{T_n} \right|$$

Ein Maß für die Schwankung der Vorhersagefehler um ihren empirischen Mittelwert ist die korrigierte Standardabweichung *s**.⁶ Kleinere Werte der Standardabweichung bedeuten eine geringere Unsicherheit der Prognose bezüglich ihrer Treffsicherheit.

$$s^* = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (RPF_n - MRPF)^2}$$

Diese vorgestellten deskriptiven Maße reichen nicht aus, um die Treffsicherheit der Prognosen abschließend beurteilen zu können. Daher wird ergänzend ein Verfahren verwendet, dass die Leistung der Vorhersage unter statistischen Aspekten betrachtet. Dazu dient die Mincer-Zarnowitz-Regression [vgl. DÖPKE (2004); S. 12].

$$T_n = a + b\hat{T}_{n|v} + \epsilon_n$$

Dabei wird das Steueraufkommen (T_n) erklärt durch die Prognose ($\hat{T}_{n|v}$), eine Konstante *a* und einen Fehlerterm ϵ_n .⁷ Diese Regressionsgleichung erlaubt es, konkrete Parameterschätzungen abzuleiten, welche mittels eines sog. Wald-Koeffiziententests auf zuvor definierte Werte getestet werden können. Mit der Nullhypothese, $H_0 : a = 0; b = 1$, wird überprüft, ob das Steueraufkommen lediglich durch die Prognose und einen zufällig schwankenden Fehlerterm erklärt werden kann. Ist es durch den vorgeschlagenen Test nicht möglich diese Hypothese zu verwerfen, dann ist die Steuerschätzung als unverzerrt zu bezeichnen.

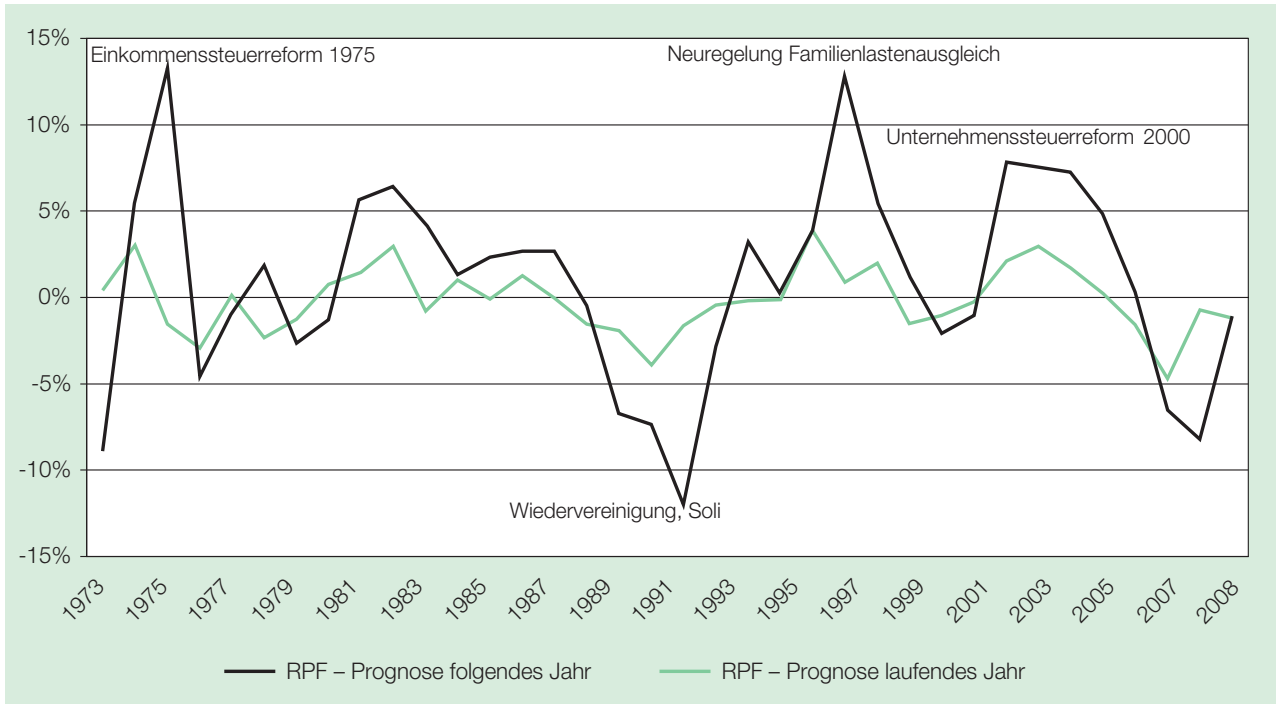
Für den Betrachtungszeitraum gilt, dass alle Beobachtungen vor dem Jahr 1990 auf Prognosen für die westdeutschen Bundesländer beruhen, während die Ergebnisse nach 1990 das gesamte Bundesgebiet umfassen. Außerdem wurden die Jahre 1989 und 1990 aus der Analyse ausgeschlossen, da die Schätzergebnisse durch die Wiedervereinigung verzerrt und diese Prognosefehler dem AKS als solche nicht anzulasten sind.

Ergebnisse

Dieser Abschnitt beinhaltet die Evaluation der Steuerschätzung des AKS. Erstens erfolgt die Analyse der Schätzungen mit einem einjährigen Prognosehorizont und zweitens ein Vergleich mit den Prognosen, die im laufenden Jahr erstellt wurden.

Es zeigt sich, dass die Treffsicherheit in verschiedenen Jahren deutlich schwankt. Die stärkste Überschätzung der späteren Steuereinnahmen lässt sich mit 13,3 % für das Jahr 1975 erkennen, während das Steueraufkommen 1991 mit -12,0 % die deutlichste Unterschätzung aufwies (vgl. Abb. 1). Das Verlaufsbild deutet an, dass der AKS häufiger über- als unterschätzt. Im gesamten Betrachtungszeitraum war er 21-mal zu optimistisch und 13-mal wurde eine zu pessimistische Prognose erstellt. Es wird deutlich, dass gerade in den Jahren mit einschlägigen Steuerrechtsänderungen die Prognosequalität der Steuerschätzung nachlässt. Beispielsweise führte die Einführung des steuerlichen Kindergeldes im Jahr 1996 zu einer Aufkommensminderung, wodurch die Steuereinnahmen drastisch überschätzt wurden [vgl. GEBHARDT (2001); S. 131]. Ein ähnliches Bild zeigt sich bei der im Jahr 2000 verabschiedeten Unternehmenssteuerreform. Aufgrund von übermäßigen Erstattungen im Jahr 2001 war das Aufkommen der Körperschaftsteuer sogar negativ [vgl. BUNDESMINISTERIUM DER FINANZEN (2010b)], wodurch es in den Jahren 2001 und 2002 zu deutlichen Überschätzungen gekommen ist. Durch die Absenkung des Körperschaftsteuersatzes ergab sich für altes Eigenkapital der Unternehmen ein Steuersparpotenzial. Dieser Anreiz konnte nur durch eine Gewinnausschüttung realisiert werden, was zu den angesprochenen übermäßigen Erstattungen in den Folgejahren führte. Eine starke Unterschätzung lag im Jahr 1991 vor. Dies ist zum einen mit der Wiedervereinigung erklärbar, wodurch es den Steuerschätzern noch nicht gänzlich gelang die Steuerkraft der neuen Bundesländer einzuschätzen. Zum anderen spielt auch der Solidaritätszuschlag eine gewichtige Rolle. Seine Einführung wirkte aufkommenssteigernd und konnte nicht vollkommen in der Steuerschätzung antizipiert werden [vgl. SCHOOF (1998); S. 210]. Der Vergleich beider Prognosehorizonte

Abbildung 1: Relative Prognosefehler des Gesamtsteueraufkommens 1973–2008 und wichtige steuerrechtliche Änderungen



Quelle: Bundesministerium der Finanzen, eigene Berechnung und Darstellung.

lässt erkennen, dass es dem Arbeitskreis im laufenden Jahr wesentlich besser gelingt die Steuereinnahmen vorherzusagen. Dies ist auf die größere Anzahl an Informationen zurückzuführen, die dem AKS ein Jahr zuvor noch nicht vorliegen.

Die relativen Prognosefehler der Schätzung im laufenden Jahr bewegen sich in einem Intervall zwischen +4,0 % und –5,0 %. Dennoch sind auch hier noch Ausreißer, bspw. durch die Unternehmenssteuerreform im Jahr 2000, sichtbar (vgl. Abb. 1). Dies spiegelt sich auch in den Gütemaßen wider (vgl. Tab. 1). Im Mittel weisen die Prognoseergebnisse der Mai-Steuerschätzung für das folgende Jahr eine Überschätzung in Höhe von 1,4 % auf, wodurch sich die zuvor erwähnte Tendenz bestätigt. Für das laufende Jahr hingegen zeigt sich im

Durchschnitt weder eine Über- noch Unterschätzung (0,0 %). Auch anhand der Koeffizienten der Regressionsgleichung werden diese beiden Tendenzen deutlich (0,98 bzw. 1,00). Die Betrachtung der Schwankungen um die Prognosefehlermittelwerte zeigt, dass die Unsicherheit im laufenden Jahr abnimmt. Während die Prognosen des folgenden Jahres eine Streuung in Höhe von 5,6 % aufweisen, wird diese Volatilität im laufenden Jahr reduziert (1,9 %). Dies ist zumeist auf sich verändernde Rahmenbedingungen im Steuerrecht und den steigenden Informationsgehalt zurückzuführen. Die Prognose des AKS weist für die verschiedenen Prognosehorizonte außerdem eine deutlich geringere absolute Entfernung zu den tatsächlichen Steuereinnahmen (1,4 % zu 4,5 %) auf.

Tabelle 1: Deskriptive Maße sowie Regressions- und t-Test Ergebnisse

Prognosehorizont	MRPF	MRAPF	s*	Regressionsgleichung ^a	F-Wert ^b	t-Wert ^c
$\hat{T}_{n n-1}$	1,4%	4,5%	5,6%	$T_n = 1,83 + 0,98\hat{T}_{n n-1} + \epsilon_n$	0,5268 (0,5955)	1,4674 (0,1517)
$\hat{T}_{n n}$	0,0%	1,4%	1,9%	$T_n = -0,56 + 1,00\hat{T}_{n n} + \epsilon_n$	0,0440 (0,9570)	-0,4344 (0,6667)

a) Schätzung mit OLS sowie heteroskedastiekonsistenter Kovarianzmatrix nach Newey und West, Werte für die Konstante in Mrd. €.
 b) Teststatistik des Wald-Koeffiziententests
 c) t-Statistik für die Unverzerrtheit der Prognosefehler

Quelle: eigene Berechnung und Darstellung.

Nach Analyse der deskriptiven Maße ist es noch nicht möglich, eine ausreichend fundierte Aussage über die Prognosequalität zu treffen. Die Anforderung der Unverzerrtheit verlangt, dass die Nullhypothese, $H_0 : a = 0; b = 1$, statistisch nicht verworfen werden kann. Für diese Beurteilung dient der in Spalte 5 von Tabelle 1 dargestellte F-Wert. Für den Wald-Koeffiziententest ergeben sich p-Werte in Höhe von 0,5955 bzw. 0,9570, was im Umkehrschluss bedeutet, dass die Schätzergebnisse beider Prognosehorizonte als unverzerrt bezeichnet werden können. Dieses Ergebnis wird durch die Untersuchung anhand der relativen Vorhersagefehler nochmals bestätigt. Dabei wird getestet, ob die mittleren relativen Prognosefehler (1,4 % und 0,0 %) statistisch von Null verschieden sind. Die Ergebnisse sind in Spalte 6 der Tabelle 1 dokumentiert. Auch bei diesem Verfahren ergeben sich p-Werte (0,1517 und 0,6667) die größer sind als das hier vorgegebene 5 %-Niveau. Somit kann die Hypothese, die Prognosefehler seien im Mittel gleich Null, in beiden Fällen nicht verworfen werden. Es zeigt sich, dass die Schätzergebnisse, die der AKS im Mai für das laufende bzw. darauffolgende Jahr erstellt, den Anforderungen der Unverzerrtheit genügen.

Fazit

Zu Beginn stellte sich die Frage, ob die Ergebnisse der deutschen Steuerschätzung als treffsicher bzw. unverzerrt zu bezeichnen sind. Nach den vorherigen Ausführungen kann diese Fragestellung mit einem „ja“ beantwortet werden. Allerdings muss auch darauf hingewiesen werden, dass es dem AKS in einigen Jahren nicht gelungen ist, einschlägige Steuerrechtsänderungen adäquat in seinen Schätzungen zu berücksichtigen [vgl. GEBHARDT (2001); S. 144]. Dies beweist auch die hohe Standardabweichung der Prognosefehler, die mit 5,6 % für die Vorhersagen des jeweils folgenden Jahres doch recht hoch ist. Dennoch sollten diese kritischen Einwände auch nicht überinterpretiert werden. Somit kann abschließend gesagt werden, dass der AKS einen wichtigen Beitrag für die fiskalische Projektion der deutschen Steuereinnahmen liefert.

Literatur

ANDRES, P.; SPIWOKS, M. (2000): Prognosegütemaße – State of the art der statistischen Expost Beurteilung von Prognosen. Sofia-Studien zur Institutionenanalyse Nr. 00-1, Darmstadt 2000.

BUNDESMINISTERIUM DER FINANZEN (2005): 50 Jahre Arbeitskreis „Steuerschätzungen“. Berlin 2005.

BUNDESMINISTERIUM DER FINANZEN (2010a): Veröffentlichung zu den „Ergebnissen des Arbeitskreises „Steuerschätzungen“ (Pressemitteilungen) seit 1971“. Homepage des Bundesministeriums der Finanzen, Berlin 2010.

BUNDESMINISTERIUM DER FINANZEN (2010b): Veröffentlichung zu den „kassenmäßigen Steuereinnahmen nach Gebietskörperschaften 1970 bis 2008“. Homepage des Bundesministeriums der Finanzen, Berlin 2010.

DÖPKE, J. (2004): Zur Qualität von Konjunkturprognosen. In: WiSt Wirtschaftswissenschaftliches Studium Heft 1, Januar 2004; S. 8–13.

GEBHARDT, H. (2001): Methoden, Probleme und Ergebnisse der Steuerschätzung. In: RWI-Mitteilungen, Jahrgang 52 (2); S. 127–147.

KÖRNER, J. (1983): Probleme der Steuerschätzung. In: HANSMEYER, K. H. (Hrsg.): Staatsfinanzierung im Wandel, Schriften des Vereins für Socialpolitik Band 134; Berlin 1983; S. 215–252.

SCHOOF, D. (1998): Von den Schwierigkeiten der Steuerschätzung. In: Deutsche Steuerzeitung Nr. 6; S. 209–214.

VON DER LIPPE, P. (1986): Steuerschätzungen und ihre wirtschaftspolitische Bedeutung. In: Steuer und Studium Heft 11/1986; S. 334–340.

¹ In diesem Fall sollte von Ausgaben, welche bspw. im Zuge einer Wirtschaftskrise notwendig sind, abstrahiert werden.

² Beispielsweise ist es denkbar, dass eine Lohnsteuersenkung und die damit einhergehende Steigerung des verfügbaren Einkommens, das Aufkommen der Umsatzsteuer durch gesteigerten Konsum erhöht.

³ Verschiedene Maße finden sich in der Publikation von ANDRES UND SPIWOKS (2000).

⁴ Eine Normierung erfolgt aus Gründen der Vergleichbarkeit zwischen zwei Prognosefehlern. Des Weiteren sind solche Normierungen insofern notwendig, da eine absolute Fehlerbetrachtung keine ausreichende qualitative Beurteilung der Güte zulässt [vgl. KÖRNER (1983); S. 239].

⁵ Es wurde anhand von verschiedenen Testverfahren die empirische Verteilungsfunktion auf die Form der Normalverteilung hin überprüft. Alle betrachteten Methoden kamen zu dem gleichen Ergebnis, dass die Prognosefehler dieser Anforderung genügen.

⁶ Die korrigierte Stichprobenstandardabweichung wird als Maß verwendet, da in endlichen Stichproben die korrigierte Stichprobenvarianz einen erwartungstreuen Schätzer darstellt.

⁷ Dieser Fehlerterm bzw. Residuum folgt den üblichen Annahmen eines sog. White-Noise-Prozesses. Dabei besitzen die zufälligen Störterme zu jedem Zeitpunkt den Erwartungswert Null, eine konstante Varianz und sind nicht miteinander korreliert.