

# Prognose der Überlebenswahrscheinlichkeit von Ich-AG's nach drei Jahren

Dr. Martin Treiber

03.02.2005

## 1 Ausgangsdaten

Die Daten vom Statistischen Bundesamt geben den monatlichen Zugang  $z_t$  und den Bestand  $B_t$  an Ich-AG's seit der Einführung dieser Betriebsform im Januar 2003 an. Der Index  $t$  gibt die Kalenderzeiten in Monaten seit Dezember 2002 an, also z.B.  $t = 5$  für "Mai 2003". Aus der Bilanzgleichung

$$B_t = B_{t-1} + z_t - a_t \quad (1)$$

erhält man daraus den monatlichen Abgang  $a_t$ , d.h. die Zahl der Ich-AGs, welche in diesen Monat -aus welche Gründen auch immer - aufgegeben haben, vgl. Abb. 1.

Zu bemerken ist, dass sich für den Oktober 2003 (Monat 8) eine Inkonsistenz ergibt, da für diesen Monat die Bestandsänderung größer als die Zugangsrate ist, einige Ich-AGs also ohne Zugangsmeldung entstanden sind, d.h.  $a_8$  negativ ist. Für die durchgeführte Prognose ist dies jedoch unerheblich. Außerdem scheinen jeweils die Daten des letzten Monats mit Unsicherheit behaftet zu sein. Im Januar 2005 wurde der Bestand an Ich-AGs für den Dezember 2004 nach oben korrigiert, wodurch sich rückwirkend die nach (1) errechnete Abgangsrate für Dezember von etwa 8000 auf unter 5000 erniedrigt (vgl. Abb. 1). Dies deutet darauf hin, dass die aktuelle Januar-Zahl möglicherweise ebenfalls nicht endgültig ist und u. U. noch nach unten korrigiert werden wird.

Die Frage besteht, welche Überlebenswahrscheinlichkeit für dreijährige Ich-AGs längerfristig zu erwarten ist und inwieweit man diese bereits aus den Daten abschätzen kann - noch ist ja die älteste Ich-AG (Stand Januar 2005) erst 25 Monate alt. Diese Frage, auf die mich freundlicherweise Dr. Andreas Lutz von [www.ueberbrueckungsgeld.de](http://www.ueberbrueckungsgeld.de) aufmerksam gemacht hat, wird in den Medien ziemlich kontrovers diskutiert. Für einen ziemlich drastischen Artikel siehe [einen Bericht von Spiegel-Online](#).

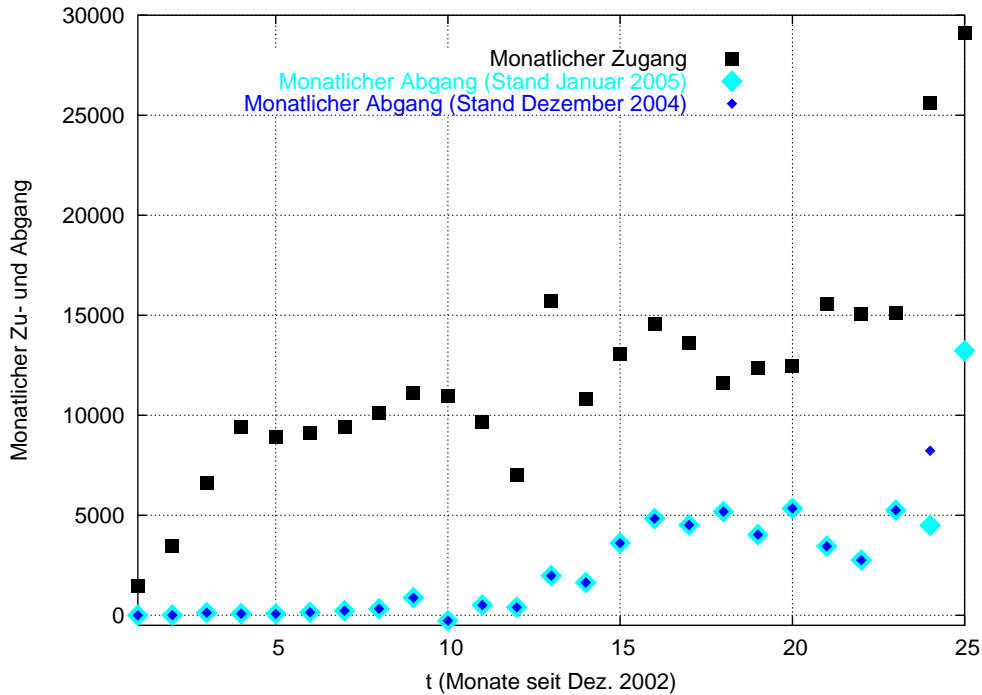


Abbildung 1: Ausgangsdaten für den monatlichen Zu- und Abgang von Ich-AGs.

## 2 Verwendetes Modell und Kalibrierung

Eine modellgestützte Prognose beruht i.A. darauf, dass es in der zu prognostizierenden Dynamik auch einige mehr oder weniger unveränderliche Größen und dazugehörige Modelle mit diesen Größen als Parameter gibt. Die modellgestützte Prognose wird dann in zwei Schritten durchgeführt:

1. *Kalibrierung*: Festlegung der Werte der Modellparameter, so dass die gegenwärtigen Daten möglichst gut beschrieben werden,
2. *Prognose*: Die kalibrierten Modellparameter werden unverändert in die Zukunft fortgeschrieben und die Dynamik mit dem Modell berechnet.

Als relativ unveränderliche Größen habe ich die *bedingten Aufgabewahrscheinlichkeiten*

$$\alpha_\tau = P(\text{Aufgabe der Ich-AG im Monat } \tau) = \text{const.} \quad (2)$$

identifiziert. Die Bedingung für diese Wahrscheinlichkeit ist, dass die Ich-AG im Vormonat existierte. Diese Wahrscheinlichkeit hängt nur vom Alter  $\tau$  der Ich-AG, nicht jedoch von deren Gründungszeitpunkt  $t$  ab.

Des Weiteren nehme ich auch für die Abhängigkeit vom Alter der Ich-AG nur zwei Parameter an:

$$\alpha_\tau = \begin{cases} \alpha_{\text{crit}} & \tau = 12 \text{ oder } 13 \\ \alpha_0 & \text{sonst.} \end{cases} \quad (3)$$

Hintergrund ist, dass die Reduzierung der Förderung nach Abschluss von 12 Monaten ein “kritisches” Alter für Ich-AGs darstellt, so dass im letzten Monat der vollen Förderung und im ersten Monat der reduzierten Förderung eine erhöhte Ausfallwahrscheinlichkeit  $\alpha_{\text{crit}}$  gilt, während in allen anderen Monaten davor und danach die geringere Ausfallwahrscheinlichkeit  $\alpha_0$  angenommen wird. Dieser geringere Wert wird auch für die Prognose angenommen.

Zwar weisen die Daten vom Januar 2005 auf ein weiteres “kritisches” Alter nach zweijährigem Bestehen hin. Wegen der oben diskutierten Unsicherheit der Daten wird dies jedoch noch nicht im Prognosemodell berücksichtigt.

Bei Kenntnis der Zugangsraten  $z_t$  im Monat  $t$  kann man aus dem Modell (2) alle relevanten statistischen Größen errechnen, z.B.

- die Überlebenswahrscheinlichkeit einer  $\tau > 1$  Monate alten Ich-AG,

$$u_\tau = u_1 \prod_{\tau'=2}^{\tau} (1 - \alpha_{\tau'}) = u_1(1 - \alpha_2)(1 - \alpha_3) \dots (1 - \alpha_\tau) \quad (4)$$

mit  $u_1 = 1$ , da die innerhalb ein -und desselben Monats an- und abgemeldeten Ich-AGs nicht in der Statistik auftauchen,

- die Abgänge im Monat  $t \geq 2$ <sup>1</sup>(es gilt  $a_1 = 0$ ),

$$a_t = \sum_{t'=1}^{t-1} z_{t'} u_{t-t'} \alpha_{t-t'+1} \quad (5)$$

- den Bestand,

$$B_t = \sum_{t'=1}^t (z_{t'} - a_{t'}) \quad (6)$$

- und die kumulierten Abgänge

$$A_t = \sum_{t'=1}^t a_{t'}. \quad (7)$$

Zur Kalibrierung verwende ich die kumulierte Abgangsrate  $A_t$ , da die monatlichen Einzeldaten für Kalibrierungszwecke zu stark schwanken. Abbildung 2 zeigt eine Kalibrierung der vom Modell (2)-(5) ermittelten Größe (7) an die Daten. Im vorliegenden Fall kann man  $\alpha_0$  und  $\alpha_{\text{crit}}$  weitgehend separat bestimmen, da für die ersten 11 Monate nur  $\alpha_0$  relevant ist und die Kalibrierung von  $\alpha_{\text{crit}}$  mit dem vollen Datensatz wegen der Kleinheit von  $\alpha_0$  weitgehend unabhängig vom Wert  $\alpha_0$  ist. Zur Kalibrierung wird die Least-Square Methode angewandt, d.h. die Fehlerquadratsumme

$$F = \sum_{t=1}^{24} (\tilde{A}_t - A_t)^2 \stackrel{!}{=} \min_{\alpha_0, \alpha_{\text{crit}}} \quad (8)$$

---

<sup>1</sup>der Buchstabe  $t$  gibt Kalenderzeiten in Monaten seit Dezember 2002 an, während  $\tau$  das Alter von Ich-AGs in Monaten bezeichnet

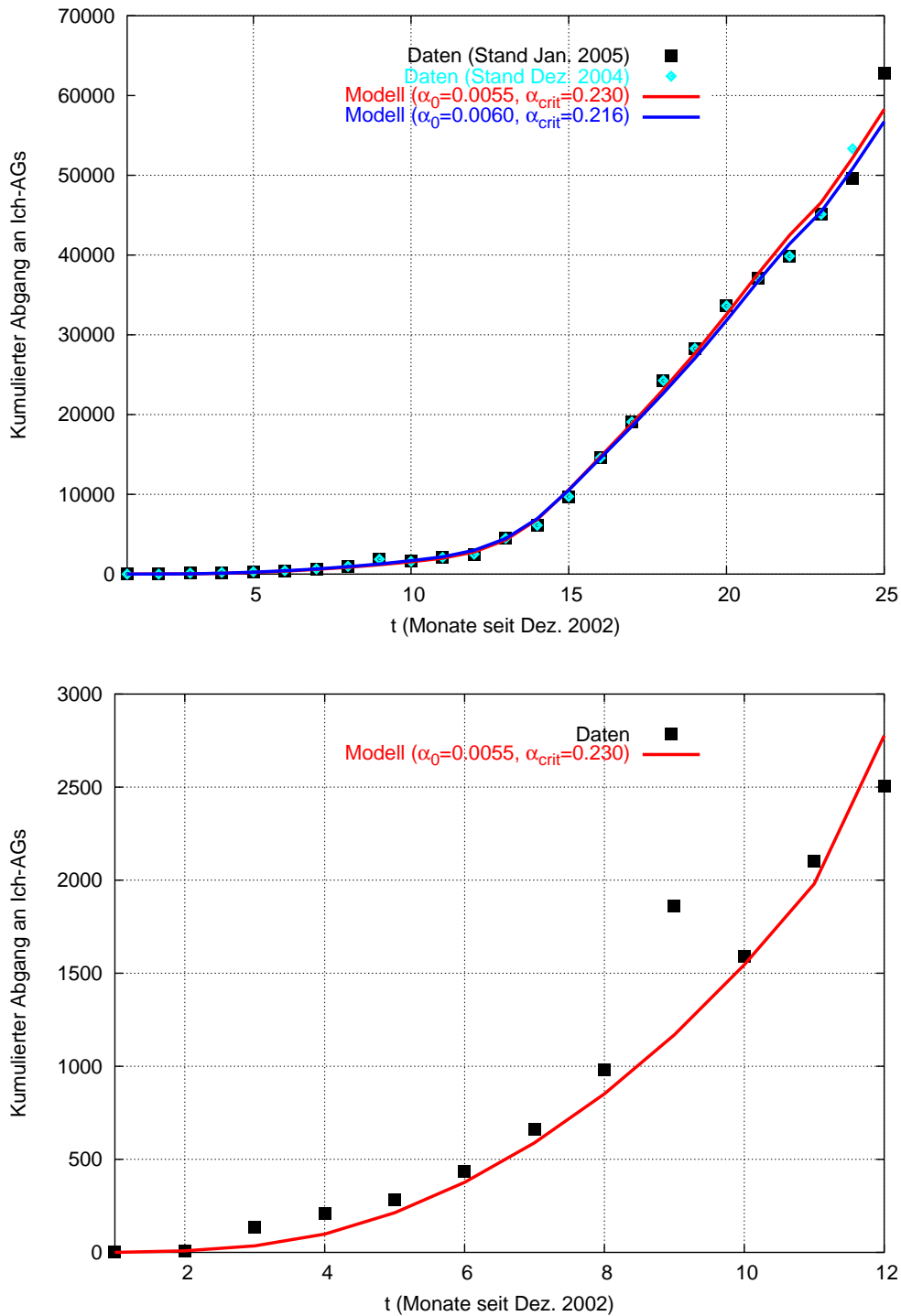


Abbildung 2: Kalibrierung der Modellparameter  $\alpha_{crit}$  und  $\alpha_0$  anhand der Zeitreihe der kumulierten Abgangszahl. Oben: Gesamter Kalibrierungszeitraum. Unten: Das erste Jahr, in dem, abgesehen vom letzten Monat, nur der Modellparameter  $\alpha_0$  wirksam ist. Die Kalibrierung wurde für die im Haupttext näher spezifizierten Szenarien 1 (rote Kurven) und 3 (blaue Kurve) durchgeführt.

zwischen den aus den Daten mit (1), (7) und dem Modell (2), (7) errechneten Werten  $\tilde{A}_t$  bzw.  $A_t$  bezüglich der Parameter  $\alpha_0$  und  $\alpha_{\text{crit}}$  minimiert. Die kalibrierten Werte der Modellparameter betragen

$$\alpha_0 = 0.55\%, \quad \alpha_{\text{crit}} = 23.0\%. \quad (9)$$

Als Maß für die Güte der Kalibrierung und des Modells kann das Unbestimmtheitsmaß

$$U = \frac{\sum_{t=1}^{24} (\tilde{A}_t - A_t)^2}{\sum_{t=1}^{24} (\tilde{A}_t - \bar{A})^2} = \frac{F}{\sum_{t=1}^{24} (\tilde{A}_t - \bar{A})^2} \quad (10)$$

dienen. Würde man den Verlauf von  $A_t$  durch eine Konstante beschreiben,  $A_t = c$ , so würde das Optimum dieser Konstante beim arithmetischen Mittel  $c = \bar{A}$  erreicht und das dazugehörige Unbestimmtheitsmaß wäre  $U = 1$ . Mit dem vorliegenden Modell wird hingegen ein sehr viel kleineres Unbestimmtheitsmaß von  $U = 0.0048$  erreicht.

### 3 Prognose

Die Prognose hängt im Wesentlichen vom Modellparameter  $\alpha_{\text{crit}}$  und in wesentlich geringerem Maße vom Parameter  $\alpha_0$  sowie der monatlichen Zugangsrate ab. Letztere wurde für den Prognosezeitraum als konstant angenommen.

Um zu überprüfen, inwieweit die möglicherweise in der Zukunft noch zu korrigierenden letzten Datenpunkte vom Januar 2005 das Ergebnis beeinflussen, wurde die Kalibrierung und Prognose auch ohne Berücksichtigung der Januar-Werte durchgeführt, wobei für den Dezember 2004 der korrigierte Wert des Bestands verwendet wurde. Die Kalibrierung ergibt für diesen Fall die Werte

$$\alpha_0 = 0.60\%, \quad \alpha_{\text{crit}} = 21.6\%, \quad U = 0.0022. \quad (11)$$

Um allfällige Fehler bei den Daten, der Modellierung und der angenommenen Zugangsrate abzuschätzen, wurde die Prognose für drei *Szenarien* durchgeführt:

- Szenario 1: Es werden alle Datenpunkte und die im vorigen Abschnitt gewonnenen kalibrierten Werte für die bedingten Ausstiegswahrscheinlichkeiten  $\alpha_0$  und  $\alpha_{\text{crit}}$  verwendet, vgl. (9). Ferner wurde für den Prognosezeitraum eine konstante monatliche Zugangsrate von 15 000 angenommen.
- Szenario 2: Wie oben, aber für die Kalibrierung wurden nur Daten bis einschließlich Dezember 2004, also die Werte von Gl. (11) verwendet.
- Szenario 3: Wie Szenario 1, aber mit einer verminderten extrapolierten Zugangsrate von 10 000 pro Monat.

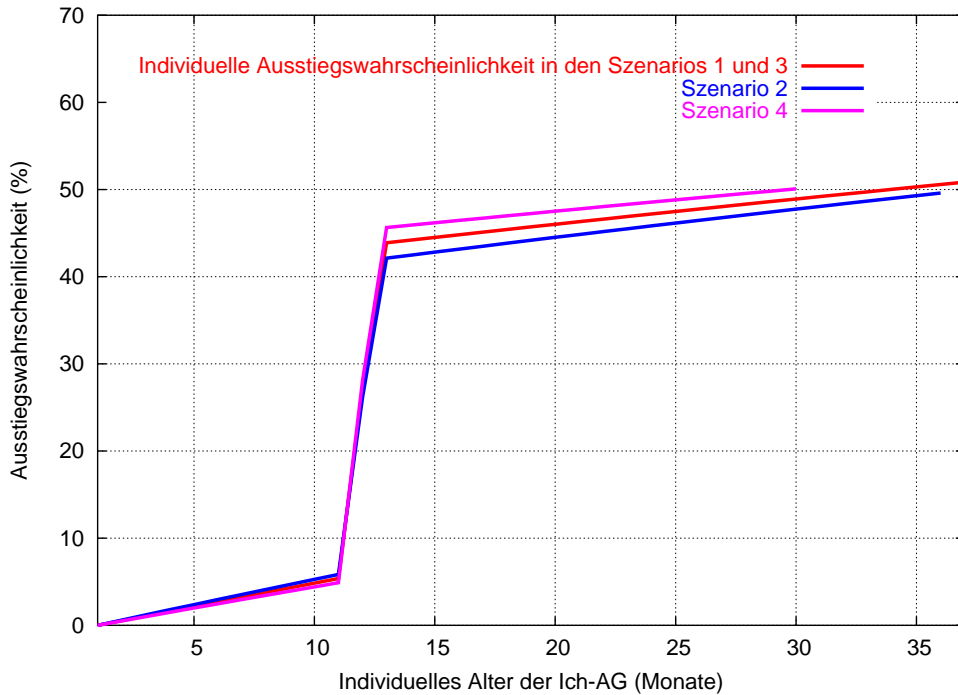


Abbildung 3: Prognose der Ausfallwahrscheinlichkeit einer einmal gegründeten Ich-AG als Funktion ihres Alters für die drei im Haupttext definierten Szenarien.

Die Abbildungen 3 - 5 stellen Ergebnisse der Prognose dar. Szenario 1 ist dabei das Referenz-Szenario für die bestmögliche Prognose, während die anderen Szenarios für die Fehlerabschätzung dienen.

Als zusätzlicher *Backtest* der Prognosegüte wurden dem Modell in einem vierten Szenario nur die Daten bis einschließlich Juni 2004 zur Verfügung gestellt und die Kalibrierung und Prognose für weitere 12 Monate wie oben durchgeführt. Es ergaben sich geringfügig andere kalibrierte Werte für die bedingten Ausfallwahrscheinlichkeiten und natürlich auch andere prognostizierte Zugangsraten (vgl. Bildunterschrift zu Abb. 6). Für den Dezember 2004 ergab der Vergleich der prognostizierten und tatsächlichen kumulierten Ausfallzahl

$$A_{24}^{\text{Daten}} = 52183, \quad A_{24}^{\text{Modell}} = 53521.6,$$

d.h. der Backtest ergab einen Prognosefehler für eine 6-Monats-Prognose von etwa 3%. Da es bis Juni 2004 nur 24259 Ausfälle gab, also in den 6 prognostizierten Monaten mehr Ausfälle als in den 18 Monaten davor, ist dies nichttrivial. Abb. 6 zeigt, dass der Prognosefehler für alle sechs prognostizierten Monate immer unter 5% liegt.

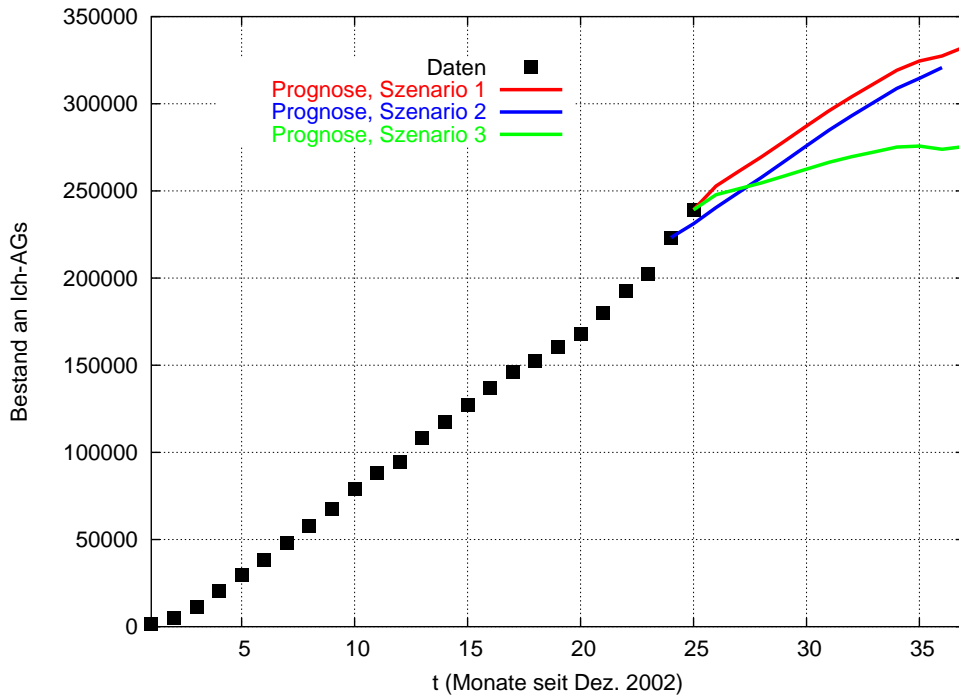


Abbildung 4: Prognose des Bestands  $B_t$  an Ich-AGs für die drei im Haupttext definierten Szenarien.

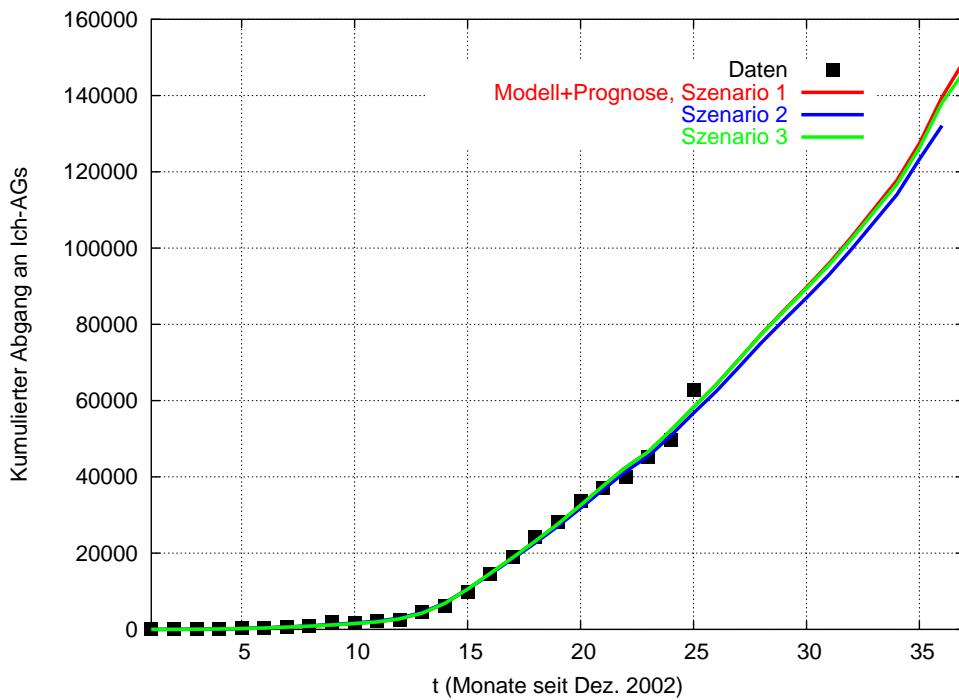


Abbildung 5: Prognose der absoluten kumulierten Abgänge für die drei im Haupttext definierten Szenarien.

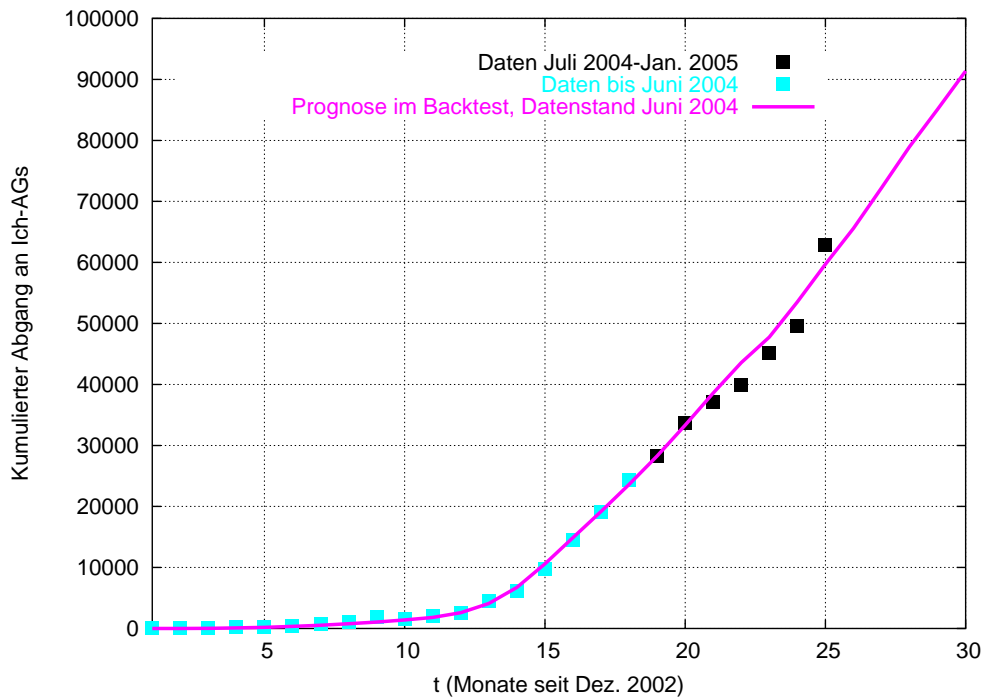


Abbildung 6: Backtest: Prognose der absoluten kumulierten Abgänge für den Datenstand von Juni 2004 bis zum Juni 2005. Das Ergebnis der Kalibrierung für diesen Datenstand ist  $\alpha_0 = 0.0050$  (statt 0.0055) sowie  $\alpha_{\text{crit}} = 0.244$  (statt 0.230). Das resultierende Unbestimmtheitsmaß beträgt  $U = 0.0030$ . Die prognostizierte Zugangsrate  $z_t = 13\,500$  wurde als das gerundete Mittel der letzten drei Monate (April-Juni 2004) angenommen.

## 4 Die Ergebnisse

Die Prognoseszenarien 1-3 sowie das Backtest-Szenario 4 erlauben folgende Aussagen:

- Der Ansatz (“das Modell”) mit bedingten Wahrscheinlichkeiten und nur zwei Parametern ist erfolgsversprechend, wie das geringe Unbestimmtheitsmaß der Kalibrierung zeigt. Dies wird durch den Backtest der Prognosegüte bestätigt.
- Die Prognose der Ausfallszahlen hängt wegen des geringen Wertes der anfänglichen Ausfallwahrscheinlichkeit nur sehr schwach von der unbekanntem Zahl der Neuzugänge ab. (Dies ist ganz analog zur Dynamik der Alterstruktur!)
- Die größte Unsicherheit liegt in der möglichen Existenz einer erhöhten bedingten Ausfallwahrscheinlichkeit  $\alpha_{\text{crit}2} > \alpha_0$  nach 24 und 25 Monaten, da dann, analog wie das Jahr zuvor, die Förderung sinkt. Die nächsten beiden Monatszahlen werden dies zeigen. Die Prognosen wurden unter der Prämisse gemacht, dass es keine derartige Erhöhung gibt, stellen also in dieser Hinsicht einen *Best-Case* dar.
- Die 12-Monats-Prognose sagt  $87\,000 \pm 3\,000$  ausgefallenen Ich-AGs für das Jahr 2005 voraus.
- Eine individuelle Ich-AG feiert nur mit einer Wahrscheinlichkeit von 49% den dreijährigen Geburtstag. Dies ist das Hauptergebnis.