KruegerAlgorithms
Quantitative Marktdaten-Analyse

Reinforcement Learning für den Ausstieg, sauber neu trainiert: Der Agent verliert gegen einen starren Trailing-Stop

DAX FTSE Dow NQ · 2015–2019 Training · 2020–2022 Test

Im Juli haben wir offengelegt, dass unsere Reinforcement-Learning-Modelle auf Merkmalen trainiert waren, die in die Zukunft schauten. Jetzt ist derselbe Agent auf der korrigierten Datenbasis neu trainiert, ohne jede Nachbesserung. Auf 22.636 Trades im Test 2020–2022 liegt er bei −0,062 R je Trade und damit 0,063 R hinter einem starren Trailing-Stop (t −2,9). Im Training schlägt er die Regel, auf einem Zufallsmarkt sogar deutlicher: Er lernt Rauschen auswendig. Eine Korrektur der Zielfunktion ändert das Urteil nicht. Der frühere Befund war ein Artefakt.

Dokument-Metadaten

Feld	Wert
Herausgeber	KruegerAlgorithms
Titel	Reinforcement Learning für den Ausstieg, sauber neu trainiert: Der Agent verliert gegen einen starren Trailing-Stop
Erscheinungsdatum	11. November 2026
Version	v1.0
Verdikt	Kein Edge
Märkte	DAX FTSE Dow NQ
Zeitraum	2015–2019 Training · 2020–2022 Test
Stichprobe	22.636 Trades im Test · 11 Läufe
Kosten	netto, Spread + Slippage je Markt
Datengrundlage	**Datenbasis:** 17 Ausbruchsregeln auf DAX, FTSE 100, Dow und Nasdaq 100, 15-Minuten-Kerzen, korrigierte Rechenkette. Training 2015–2019 (37.540 Trades), Test 2020–2022 (22.636 Trades). Der Holdout ab 2023 wurde nicht geladen. Derselbe Ablauf auf zwei Zufallsmärkten. Kosten je Round-Trip wie in allen Studien dieser Reihe, Standardfehler nach Tag geclustert. Agent, Vergleichsregel und Tor standen vor dem ersten Lauf fest (registriert am 11. November 2026, SHA-256 5ab9e565…, Korrekturlauf 524969f5…). **Keine Handelsempfehlung.**
Dokumenttyp	Statistische Auswertung (keine Trading-Empfehlung)

Rein statistische Auswertung historischer Marktdaten. Keine Trading-Empfehlung, keine Finanzberatung. Historische Performance ist keine Garantie für die Zukunft. Trading mit Hebelprodukten kann zum Totalverlust führen.

Zusammenfassung

Im Juli haben wir offengelegt, dass unsere Reinforcement-Learning-Modelle auf Merkmalen trainiert waren, die in die Zukunft schauten. Jetzt ist derselbe Agent auf der korrigierten Datenbasis neu trainiert, ohne jede Nachbesserung. Auf 22.636 Trades im Test 2020–2022 liegt er bei −0,062 R je Trade und damit 0,063 R hinter einem starren Trailing-Stop (t −2,9). Im Training schlägt er die Regel, auf einem Zufallsmarkt sogar deutlicher: Er lernt Rauschen auswendig. Eine Korrektur der Zielfunktion ändert das Urteil nicht. Der frühere Befund war ein Artefakt.

Reinforcement Learning verspricht, was sich mit festen Regeln schwer fassen lässt: einen Agenten, der in jedem Moment eines laufenden Trades entscheidet, ob er hält, den Stop nachzieht, teilweise schließt oder ganz aussteigt, und der diese Entscheidungen aus Zehntausenden Trades selbst lernt. Wir hatten einen solchen Agenten im Sommer 2026 trainiert. In einem Walk-Forward-Test lag er in allen 20 Fenstern vorn. Dann fanden wir vier Merkmale, die in die eigene Zukunft schauten, und später einen zweiten Fehler in der Rechenkette, der selbst auf einem Zufallsmarkt Gewinne auswies. Beide Fehler sind behoben. Was fehlte, war die Gegenprobe: derselbe Agent, auf der sauberen Basis neu trainiert.

Diese Auswertung enthält keine Trading-Empfehlung.

Inhaltsverzeichnis

Zusammenfassung **3**

Inhaltsverzeichnis **4**

Abbildungen 4

1. Der Aufbau **5**

2. Ergebnis: Tor in allen drei Punkten verfehlt **5**

3. Was der Agent gelernt hat **6**

4. Die Zielfunktion war falsch, die Korrektur hilft nicht **6**

Was das heißt **6**

Grenzen **7**

Abbildungen

Abb. 1: Links der Abstand des Agenten zur besten starren Regel im Test 2020–2022, je Trade in R, mit einem Standardfehler: wie im Sommer trainiert (orange) u…

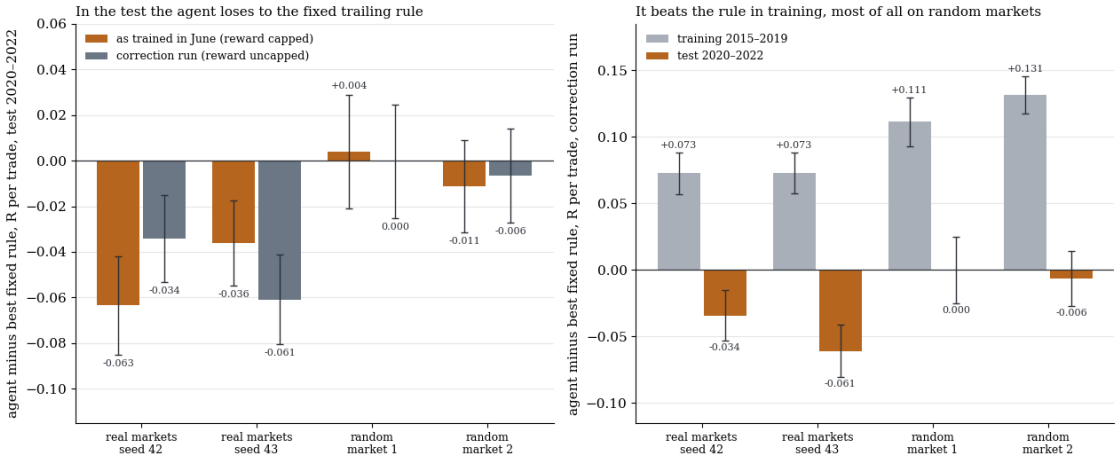


Abb. 1: Links der Abstand des Agenten zur besten starren Regel im Test 2020–2022, je Trade in R, mit einem Standardfehler: wie im Sommer trainiert (orange) und im Korrekturlauf mit ungekappter Belohnung (blaugrau). Rechts derselbe Abstand im Korrekturlauf, Training gegen Test.

1. Der Aufbau

Baustein	Festlegung
Agent	PPO mit einem Netz aus zwei Schichten zu je 128 Neuronen, 3 Millionen Trainingsschritte, unverändert gegenüber dem Sommer, kein Tuning
Aktionen	sieben je Kerze: halten, Stop auf Einstand, enger Trailing-Stop, weiter Trailing-Stop, schließen, aufstocken, teilweise schließen
Merkmale	nur Information, die zum Zeitpunkt der Entscheidung bekannt ist
Vergleichsregel	die beste von vier starren Regeln (halten, weiter Trailing-Stop, enger Trailing-Stop, Stop auf Einstand), gewählt im Training
Variante A (primär)	jeder Trade endet spätestens am Sitzungsende
Variante B	Trades laufen bis zu fünf Handelstage, wie im Sommer
Zufallsmarkt	derselbe Ablauf auf zwei künstlichen Märkten ohne Richtung

Das Tor für Variante A: Der Agent muss die beste starre Regel im Test mit $t \geq 2$ schlagen, dieser Vorsprung muss mindestens zwei Standardfehler über dem auf beiden Zufallsmärkten liegen, und der Agent selbst muss nach Kosten im Plus sein. Auf einem Zufallsmarkt hat jede Ausstiegsregel denselben Erwartungswert, nämlich minus Kosten. Ein Vorsprung dort wäre ein Hinweis auf einen weiteren Fehler.

2. Ergebnis: Tor in allen drei Punkten verfehlt

Test 2020–2022, Variante A	beste starre Regel	Agent	Agent minus Regel
echt, Startwert 42	weiter Trailing-Stop, +0,001 R (t 0,0)	**−0,062 R (t −8,1)**	**−0,063 R (t −2,9)**
echt, Startwert 43	weiter Trailing-Stop, +0,001 R	−0,035 R (t −2,2)	−0,036 R (t −1,9)
Zufallsmarkt 1	halten, −0,085 R	−0,081 R	+0,004 R (t 0,2)
Zufallsmarkt 2	weiter Trailing-Stop, −0,073 R	−0,084 R	−0,011 R (t −0,6)

22.636 Trades je Lauf auf den echten Märkten. Der Startwert ist die Zufallszahl, mit der das Training beginnt, der zweite Lauf zeigt, wie stark das Ergebnis davon abhängt.

Der Agent ist nicht besser als die Regel, sondern schlechter, und nach Kosten klar im Minus. Auf den Zufallsmärkten liegt der Abstand bei null, wie er muss, die Rechnung ist also sauber. Gegen die Zufallsmärkte gerechnet beträgt der Abstand −0,067 R (t −2,0) und −0,052 R (t −1,8). In Variante B mit bis zu fünf Tagen

Haltedauer liegt der Agent bei −0,044 R (t −3,0) und 0,056 R hinter der besten Regel (t −1,4).

Nebenbei zeigt die Tabelle, was die Regeln selbst wert sind: Mit dem besten starren Ausstieg liegen die 17 Ausbruchsregeln im Test nach Kosten bei null.

3. Was der Agent gelernt hat

Der Agent hält im Mittel 2,7 Kerzen, also rund 40 Minuten. Er schließt früh. Damit schneidet er den Teil der Verteilung ab, von dem ein Trailing-Stop lebt: Beim weiten Trailing-Stop enden 3,9 % der Trades mit mehr als 5 R, beim Agenten 0,2 %. Die Kosten zahlt er trotzdem bei jedem Trade.

Im Training sieht das anders aus. Dort schlägt der Agent die Regel um 0,055 R je Trade (t 3,5), mit dem zweiten Startwert um 0,114 R (t 7,4). Auf dem ersten Zufallsmarkt, auf dem es nichts zu lernen gibt, schlägt er die dort beste Regel im Training um 0,085 R (t 4,7). Ein Vorsprung, der auf einem Markt ohne Struktur genauso entsteht, ist auswendig gelerntes Rauschen. Im Test ist davon nichts übrig.

4. Die Zielfunktion war falsch, die Korrektur hilft nicht

Die Nachprüfung fand einen eigenen Fehler im Trainingsaufbau, der schon im Sommer bestand. Die Belohnung wurde beim Training normalisiert und dabei gekappt, umgerechnet bei rund 5 R. Rund 4 % der Trades mit weitem Trailing-Stop liegen darüber und tragen etwa 0,10 R je Trade zum Mittel bei. Der Agent hat also nie gesehen, was die großen Läufer wert sind. Mit gekappten Ergebnissen gerechnet war er im Test sogar besser als die Regel (−0,066 gegen −0,098 R). Er hat gelernt, was die Zielfunktion verlangte.

Deshalb gab es genau einen Korrekturlauf mit ungekappter Belohnung, vorab registriert, sonst identisch:

Test 2020–2022, ungekappt	Agent	Agent minus Regel	im Training
echt, Startwert 42	−0,033 R (t −2,6)	−0,034 R (t −1,8)	+0,073 R (t 4,6)
echt, Startwert 43	−0,060 R (t −5,8)	−0,061 R (t −3,1)	+0,073 R (t 4,8)
Zufallsmarkt 1	−0,085 R	0,000 R (t 0,0)	+0,111 R (t 6,1)
Zufallsmarkt 2	−0,079 R	−0,006 R (t −0,3)	+0,131 R (t 9,3)

Nach der vorab festgelegten Lesart ist das mit dem ersten Startwert ein Gleichstand und mit dem zweiten ein schlechteres Ergebnis als die Regel. Auch ohne Kappung schneidet der Agent die großen Läufer ab (0,8 bis 1,2 % der Trades über 5 R statt 3,9 %), und auch ohne Kappung lernt er im Training Rauschen, auf den Zufallsmärkten am stärksten. Einen dritten Lauf gibt es nicht, das war vorab so festgelegt.

Was das heißt

Der Befund vom Sommer war ein Artefakt aus Zukunftsinformation und einer fehlerhaften Rechenkette. Ohne beides lernt derselbe Agent, früh zu schließen, und verliert gegen eine Regel, die sich in einer Zeile beschreiben lässt. Drei Dinge nehmen wir daraus mit:

- **Der Vorsprung im Training sagt nichts.** Erst der Zufallsmarkt zeigt, dass er auch ohne jede Struktur entsteht.
- **Die Zielfunktion prüfen.** Eine übliche Voreinstellung der Bibliothek hat genau die Trades unsichtbar gemacht, auf die es ankommt.
- **Mehr als ein Startwert.** Zwei Trainingsläufe auf denselben Daten liegen im Test 0,03 R auseinander. Ein einzelner Lauf ist eine Ziehung.

Das Ergebnis deckt sich mit unserer Messung zu Machine Learning im Intraday-Handel. Auf diesen Merkmalen trainieren wir kein weiteres Reinforcement Learning.

Grenzen

- **Ein Agent, eine Einstellung.** Kein Tuning, keine andere Architektur, mit Absicht. Dass Reinforcement Learning für Ausstiege grundsätzlich nicht funktioniert, ist damit nicht gezeigt, nur dass dieser Agent auf diesen Merkmalen nichts findet.
- **Merkmale aus Kerzen.** Orderbuch-, Options- oder Nachrichtendaten sieht der Agent nicht.
- **Drei Testjahre, vier Märkte.** 15-Minuten-Kerzen, ohne Finanzierungskosten und ohne Slippage beim Ausstieg.
- **Zweiter Blick auf den Test.** Der Korrekturlauf nutzt denselben Testzeitraum mit einer nachträglich gefundenen Änderung. Er war vorab registriert, ist aber kein unabhängiger Test.
- **Startwerte.** Zwei Trainingsläufe auf denselben Daten kommen zu merklich verschiedenen Ergebnissen. Die Aussage stützt sich darauf, dass keiner die Regel schlägt.

Alle Musterfamilien und Holdouts im Überblick. Verwandt: Brute-Force-Suche gegen Zufallsmärkte.

Rein statistische Auswertung historischer Marktdaten. Keine Trading-Empfehlung, keine Finanzberatung. Historische Performance ist keine Garantie für die Zukunft. Trading mit Hebelprodukten kann zum Totalverlust führen.