**KruegerAlgorithms**
*Quantitative Marktdaten-Analyse*

# Machine Learning im Intraday-Handel: Out-of-Sample-IC von +0,004, keine von 79 Kombinationen netto brauchbar

10 Märkte · 2015–2026

*Wir haben Ridge-Regression und XGBoost auf 98 lookahead-freien Merkmalen im Walk-Forward über 79 Kombinationen aus Markt, Zeitpunkt und Horizont trainiert. Die Out-of-Sample-Rangkorrelation liegt bei +0,004 (Ridge) und +0,006 (XGBoost): messbar über dem Zufallsmarkt, ökonomisch null. Keine der 79 Kombinationen erreicht netto t > 2, die Modellregel verliert im Schnitt 1,4 bis 1,6 bps pro Trade. Die aus dem Scan destillierte Regel (USDJPY an den letzten Monatstagen) zeigte im Holdout t 0,61.*

**Dokument-Metadaten**

| Feld | Wert |
|---|---|
| Herausgeber | KruegerAlgorithms |
| Titel | Machine Learning im Intraday-Handel: Out-of-Sample-IC von +0,004, keine von 79 Kombinationen netto brauchbar |
| Erscheinungsdatum | 29. September 2026 |
| Version | v1.0 |
| Verdikt | Kein Edge |
| Märkte | 10 Märkte |
| Zeitraum | 2015–2026 |
| Stichprobe | 98 Merkmale · 7.343 Tests |
| Kosten | netto, Spread + Slippage |
| Datengrundlage | **Datenbasis:** DAX, FTSE, NQ, SPX, Dow, HK50, JPN225 (ab 2018), Gold, EURUSD, USDJPY; Dukascopy-CFD-Minutendaten (BID). 83.717 Zeilen (Markt × Entscheidungszeitpunkt × Tag), 98 lookahead-freie Merkmale (Intraday-Pfad, Nachbarmärkte, öffentliche Tagesdaten, Kalender und Termine), 79 Kombinationen aus Markt, Zeitpunkt und Horizont (60 Minuten oder bis Handelsschluss). Walk-Forward mit wachsendem Trainingsfenster, Testjahre 2018–2022, ein Tag Embargo. Suchzeitraum 2015–2022, Holdout 2023 bis Juni 2026 nur für die eine destillierte Regel. Kosten: Spread plus Slippage je Markt, Round-Trip 1,2 bis 4,7 bps. Benchmarks: Zufallsmarkt (jede Minutenkerze zufällig gespiegelt: gleiche Volatilität, keine Richtung) und über Tage vermischte Ziele. **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

*Wir haben Ridge-Regression und XGBoost auf 98 lookahead-freien Merkmalen im Walk-Forward über 79 Kombinationen aus Markt, Zeitpunkt und Horizont trainiert. Die Out-of-Sample-Rangkorrelation liegt bei +0,004 (Ridge) und +0,006 (XGBoost): messbar über dem Zufallsmarkt, ökonomisch null. Keine der 79 Kombinationen erreicht netto t > 2, die Modellregel verliert im Schnitt 1,4 bis 1,6 bps pro Trade. Die aus dem Scan destillierte Regel (USDJPY an den letzten Monatstagen) zeigte im Holdout t 0,61.*

„KI findet Muster, die Menschen übersehen“ ist die häufigste Erwartung an Machine Learning im Trading. Sie hat einen wahren Kern: Modelle gewichten viele Merkmale gleichzeitig und können Wechselwirkungen finden. Sie hat ein Problem: Finanzdaten haben ein sehr kleines Signal-Rausch-Verhältnis, und jedes flexible Modell findet im Rauschen etwas.

Wir haben es mit den zwei robustesten Standardverfahren geprüft, mit starker Regularisierung und einem Vergleich gegen Märkte ohne Muster. Am Ende blieb eine Regel übrig, die wir im Holdout testeten.

Diese Auswertung enthält keine Trading-Empfehlung.

# Inhaltsverzeichnis

**Zusammenfassung** **3**

**Inhaltsverzeichnis** **4**

Abbildungen . . . . . 4

**1. Ein Merkmal nach dem anderen: schwache, aber echte Struktur** **5**

**2. Ridge und XGBoost: über dem Zufallsmarkt, aber nahe null** **6**

**3. Die destillierte Regel: im Holdout null** **7**

**Was das heißt** **7**

**Grenzen** **8**

## Abbildungen

*Abb. 1: Die eine Regel, die der Scan übrig ließ: USDJPY long von 07:00 bis 16:00 Londoner Zeit an den letzten drei Handelstagen des Monats, ohne Signal und o…*
*Abb. 2: 31. Oktober 2018 (letzter Handelstag): −28,6 bps nach Kosten*
*Abb. 3: Anzahl der 6.921 Einzelmerkmals-Tests mit |t| über 2, 3 und 4 auf echten Daten (orange), auf dem Zufallsmarkt (grau) und bei reinem Zufall erwartet (…*
*Abb. 4: Mittlere Out-of-Sample-IC über 79 Kombinationen*
*Abb. 5: Nettoergebnis pro Trade der Regel in der Suchperiode (orange) und im Holdout (grau), jeweils mit allen Trades und ohne die fünf besten Tage, darüber…*




*Abb. 1: Die eine Regel, die der Scan übrig ließ: USDJPY long von 07:00 bis 16:00 Londoner Zeit an den letzten drei Handelstagen des Monats, ohne Signal und ohne Filter. 27. September 2019 (vorletzter Handelstag): +35,6 bps nach Kosten.*




*Abb. 2: 31. Oktober 2018 (letzter Handelstag): −28,6 bps nach Kosten. Beide Beispiele stammen aus der Suchperiode 2015–2022 und wurden mit festem Seed zufällig aus allen Handelstagen der Regel gezogen, nicht ausgesucht.*

# 1. Ein Merkmal nach dem anderen: schwache, aber echte Struktur

Vor dem Modell steht der einfache Scan: 6.921 Tests, jeweils ein Merkmal gegen die Rendite eines Marktes zu einem Zeitpunkt (Rangkorrelation, auch Information Coefficient oder IC genannt). Er findet mehr starke Zellen als der Zufallsmarkt: $|t| > 3$ bei 50 Tests, auf dem Zufallsmarkt bei 20, bei reinem Zufall wären 18,7 zu erwarten. Sechs Treffer überstehen die Mehrfachtest-Korrektur (Zufallsmarkt: 0), ihre IC liegt bei 0,09 bis 0,10. Es gibt also Struktur, aber sie ist klein.

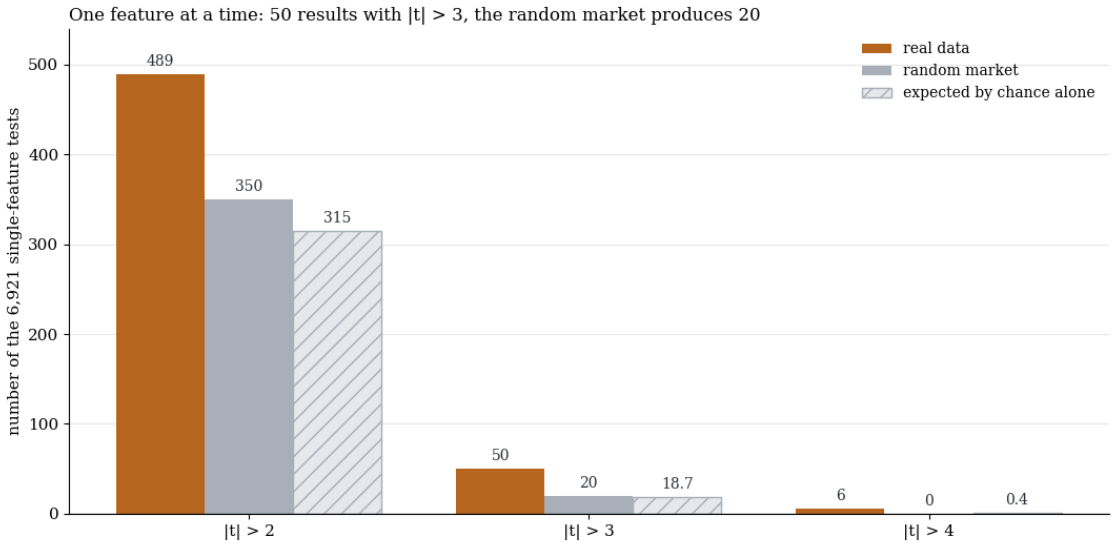



*Abb. 3: Anzahl der 6.921 Einzelmerkmals-Tests mit |t| über 2, 3 und 4 auf echten Daten (orange), auf dem Zufallsmarkt (grau) und bei reinem Zufall erwartet (schraffiert).*

## 2. Ridge und XGBoost: über dem Zufallsmarkt, aber nahe null

Die Modelle: Ridge-Regression und flaches XGBoost (Tiefe 2, 150 Bäume) auf allen 98 Merkmalen, Ziel ist die Rendite geteilt durch die ATR der letzten 20 Tage. Das Training wächst Jahr für Jahr, getestet wird 2018 bis 2022 mit einem Tag Embargo. Gehandelt wird, wenn die Vorhersage betragsmäßig über dem Median der Vorhersagen im Trainingsfenster liegt.

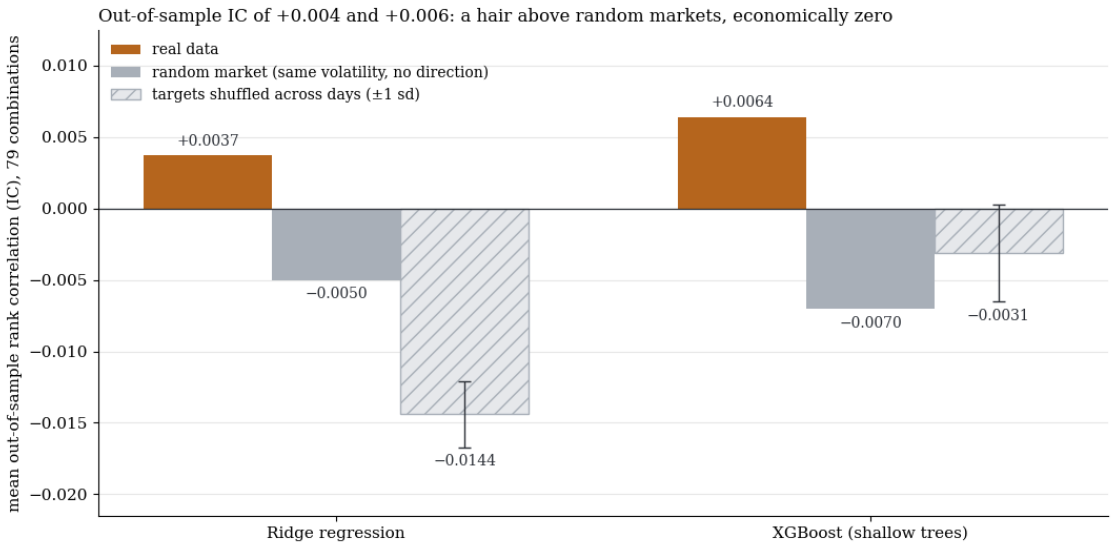



*Abb. 4: Mittlere Out-of-Sample-IC über 79 Kombinationen. Orange: echte Daten. Grau: Zufallsmarkt mit identischer Pipeline. Schraffiert: Ziele über Tage vermischt (Fehlerbalken: eine Standardabweichung über die Wiederholungen).*

| | Ridge | XGBoost |
|---|---|---|
| mittlere Out-of-Sample-IC | +0,0037 | +0,0064 |
| Zufallsmarkt, gleiche Pipeline | −0,0050 | −0,0070 |
| Ziele über Tage vermischt | −0,0144 | −0,0031 |
| z gegen den Zufallsmarkt (79 Kombinationen / effektiv 40 unabhängige) | 2,6 / 1,8 | 4,4 / 3,1 |
| Modellregel, Netto pro Trade im Mittel | −1,37 bps | −1,64 bps |

| | Ridge | XGBoost |
|---|---|---|
| Kombinationen mit Netto über null | 16 von 79 | 15 von 79 |
| Kombinationen mit Netto-t über 2 | 0 | 0 |

Die IC liegt messbar über der Null des Zufallsmarkts, das ist der Rest echter Struktur aus dem Einzelmerkmals-Scan. Ökonomisch ist eine Rangkorrelation von 0,004 nichts: Die Handelsregel auf den Vorhersagen verliert im Schnitt 1,4 bis 1,6 bps pro Trade, und keine der 79 Kombinationen erreicht netto t > 2. Die besten Zellen sind US-Indizes in der letzten Handelsstunde (SPX mit XGBoost IC +0,082, NQ mit Ridge +0,066), netto aber nur +0,8 bis +2,2 bps bei t unter 1,3.

Bei Gold von 13:30 bis 14:30 Londoner Zeit lagen beide Modelle systematisch falsch herum: IC −0,09 (Ridge) und −0,11 (XGBoost), in allen fünf Testjahren negativ, die Modellregel verliert 6,7 bps pro Trade (t −5,06). Die Ursache ist ungeklärt. Das Modell umzudrehen wäre nachträglich und kein Test. Lesart: Das Machine Learning bestätigt nur, was ein Merkmal allein schon zeigt, und findet keine handelbare Wechselwirkung.

## 3. Die destillierte Regel: im Holdout null

Aus den Scan-Treffern haben wir einfache Regeln destilliert. Die stärkste ist die aus den Beispielen oben: USDJPY long von 07:00 bis 16:00 Londoner Zeit an den letzten drei Handelstagen des Monats, ohne Signal.

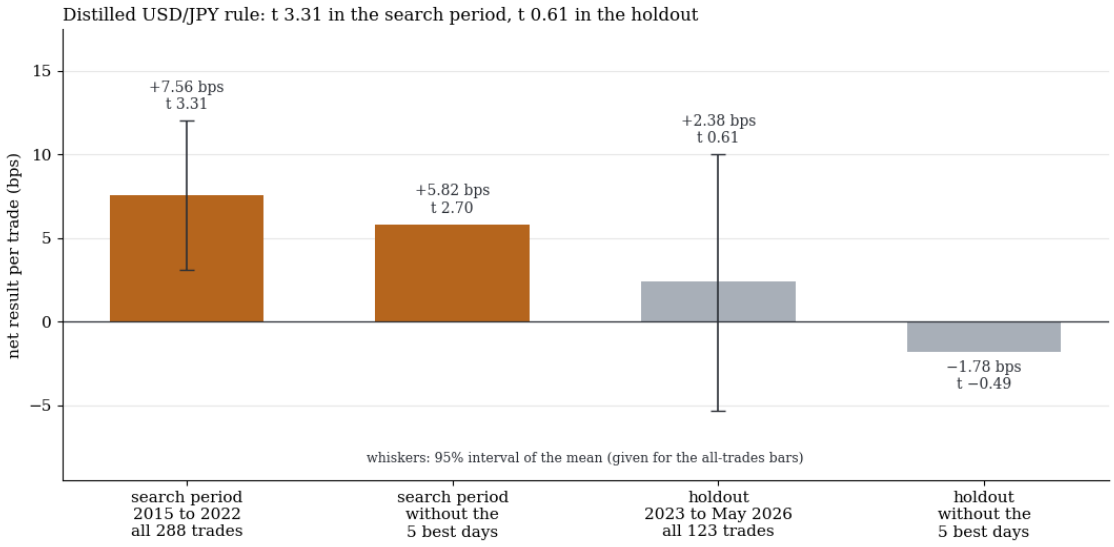



*Abb. 5: Nettoergebnis pro Trade der Regel in der Suchperiode (orange) und im Holdout (grau), jeweils mit allen Trades und ohne die fünf besten Tage, darüber der t-Wert. Fehlerbalken: 95-%-Intervall des Mittels.*

Suchperiode: 288 Trades, +7,56 bps netto (t 3,31), sieben von acht Jahren positiv, beide Hälften +7,5 und +7,7 bps, Zufallsmarkt −1,8 bps. Mit q = 0,18 verfehlt die Regel die Mehrfachtest-Korrektur über die 7.343 Tests des Scans.

Holdout, ein Lauf: 123 Trades in 41 Monaten, +2,38 bps (t 0,61, p = 0,27), 95-%-Intervall −5,3 bis +10,0 bps. Ohne die fünf besten Tage −1,78 bps, nur zwei von vier Jahren positiv, und ein Placebo (Monatsmitte, +3,7 bps) liegt über der Regel. Bei 1,5-fachen Kosten sind es +1,84 bps (t 0,47).

Der Holdout kann die Regel nicht widerlegen (Teststärke höchstens 62 %, selbst wenn der volle Effekt der Suchperiode echt wäre), aber er belegt sie auch nicht. Der Punktschätzer schrumpft auf ein Drittel, wie es bei einem aus 7.343 Tests ausgewählten Fund zu erwarten ist.

## Was das heißt

Machine Learning ist hier kein Zauber. Auf 98 Merkmalen und 79 Kombinationen bleibt eine Rangkorrelation von 0,004 bis 0,006, und sie ist kleiner als die Kosten. Das Modell bestätigt die Einzelmerkmale und findet keine handelbare Wechselwirkung. Das heißt nicht, dass ML nie hilft. Es heißt: Mit diesen Minutendaten, diesen Merkmalen und diesen Modellen bringt es keinen handelbaren Vorteil. Die zweite Lehre ist die Suchbreite: Ein Scan mit 7.343 Tests liefert immer einen schönen t-Wert (3,31), und im Holdout fällt er auf 0,61. Genau dafür gibt es den Holdout.

## Grenzen

- **CFD-Minutendaten (BID), festes Kostenmodell.** Round-Trip 1,2 bis 4,7 bps, reale Kosten sind eher höher. Das Tickvolumen ist über lange Strecken null und taugt nur eingeschränkt als Merkmal.
- **Einfaches ML-Design.** Feste, starke Regularisierung, keine Hyperparameter-Suche, nur fünf Testjahre. Nicht getestet: ElasticNet, Pooling über Märkte, andere Handelsschwellen, tiefere Modelle, neuronale Netze.
- **Kleine Stichprobe.** Die Regel hat 36 Trades pro Jahr, der Holdout 123 Trades. „Nicht bestätigt“ heißt nicht „bewiesen null“.
- **Auswahl.** Die Regel wurde aus 7.343 Tests ausgewählt, t 3,31 ist selektionsverzerrt. Eine zweite Beobachtung ähnlicher Stärke (94 Ereignistage) ging nicht in den Holdout.
- **Ein Holdout für eine Regel.** Das Design kann Muster, die erst ab 2023 entstanden sind, nicht finden. Das war der Preis für einen sauberen Holdout.

Alle Musterfamilien des Scans im Überblick. Verwandt: KI im Trading: Realität und Hype, Fehler bei KI-Handelsstrategien, Lookahead-Bug im Rückblick, 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.*