

# MS-REPLAY-20COIN-180D-P0-FILTER-ABLATION-1

## — Report

**Datum:** 2026-06-05 01:30 UTC **Modus:** READ-ONLY, kein Bot-/Worker-/Container-Touch, 0x DB-Mutation, 0x Mainnet, 0x Push **Basis:** artifacts/ms-replay-20coin-180d — 491 Trades aus 20-Coin-Universe über 180 Tage (2025-12-29 bis 2026-05-25), Baseline-Filter MS-MTF-D 3-of-4 + Dedup + Stable-Block **Frage:** Können Regime-, BTC-Macro-, MTF- oder Circuit-Breaker-Filter die Strategie auf operatorische Akzeptanz-Schwelle bringen? **Akzeptanz (HART):**  $PF \geq 1.5$  UND  $Exp > 0$  UND VALIDATION+TEST beide positiv UND  $P(loss) << 10\%$  UND Max Streak  $<< 15$  UND MaxDD besser als -38% — alle ohne Coin-Allowlist/Denylist

### 0. TL;DR

**NEIN** — keine einzelne Filter-Variante erfüllt die HARTE Akzeptanz-Schwelle. **NEIN** — keine Combo passt sauber. **JA** — RANGE-only / no-TRENDS sind in TRAIN strong, kollabieren aber in VALIDATION (Feb-Apr 2026 BTC-Bärenmarkt). BTC-1d > EMA200 hätte alle Trades blockiert (BTC war 180d nie über EMA200) → faktisch perfekter „Macro-Off“-Schutz, aber Strategy-Edge dann = 0. Empfehlung: MS-Pause bleibt. Strategy braucht Re-Design, nicht Filter-Overlay.

### 1. Baseline (zur Erinnerung)

Metric	Value
N	491
Win-Rate	45.0 %
Expectancy	-0.026 %
Profit-Factor	0.95
Total	-12.55 %
Max-DD	-38.4 %
Max Loss-Streak	15
Bootstrap P(loss)	62.2 %
Bootstrap 95% CI	[-93.2 %, +58.7 %]

### 2. MTF-Ablation

Variant	N	WR	Exp	PF	Total	Δ	MaxDD	P(loss)
1a no MTF	491	45.0%	-0.026%	0.95	-12.55 %	+0.00	38.4%	62.0%
1b 1h+4h	491	45.0%	-0.026%	0.95	-12.55 %	+0.00	38.4%	64.0%
1c 2-of-4	491	45.0%	-0.026%	0.95	-12.55 %	+0.00	38.4%	64.8%
1d 3-of-4 (Baseline)	491	45.0%	-0.026%	0.95	-12.55 %	0	38.4%	62.2%
1e 4-of-4 strict	412	45.1%	+0.002%	1.005	+0.90 %	+13.46	26.4%	55.0%

**Befund MTF:** - 1a-1d sind identisch, weil die 491 Trades bereits durch MTF-D 3-of-4 selektiert wurden — Post-hoc Lockerung kann verworfene Signale nicht zurückbringen. - 1e (4-of-4 strict) reduziert auf 412 Trades und kommt auf +0.9 % Total — knapp breakeven, PF 1.005 weit unter 1.5. - MTF-Verschärfung bringt keinen profitablen Edge.

### 3. Regime-Aware (klassifiziert via existing market\_regime aus decision\_logs)

Variant	N	WR	Exp	PF	Total	Δ	MaxDD	Streak	P(loss)
2a RANGE only	261	47.9%	+0.100%	1.31	+26.11 %	+38.66	16.9%	8	13.6%
2b no WEAK_TREND	284	46.8%	+0.062%	1.16	+17.52 %	+30.07	20.1%	8	22.8%
2c no BULL_STRONG	468	45.5%	-0.009%	0.98	-3.96 %	+8.59	—	—	57.6%
2d RANGE OR WEAK	468	45.5%	-0.009%	0.98	-3.96 %	+8.59	—	—	58.4%
2e no TRENDS (=RANGE only)	261	47.9%	+0.100%	1.31	+26.11 %	+38.66	16.9%	8	12.6%

**Befund Regime:** - RANGE-only zeigt deutliche Verbesserung — Total +26.11 %, PF 1.31, MaxDD halbiert auf

16.9 %, Streak halbiert auf 8. - Akzeptanz-Test: PF 1.31 < **1.5 (FAIL)**, P(loss) 13.6% >> << **10% (FAIL)**. - Walk-Forward (siehe §6) zeigt: TRAIN +33.4 %, **VALIDATION -11.1 %** (Kollaps), TEST +3.8 %. Validation killt die Robustheit. - Strategie ist **Range-spezifisch**, nicht „trend\_follow“ wie der Name sagt.

## 4. BTC-Macro-Filter

Variant	N	WR	Exp	PF	Total	Δ	MaxDD	P(loss)
3a BTC-4h close>EMA20	431	45.2%	+0.018%	1.04	+7.95 %	+20.50	32.9%	41.0%
3b BTC-4h EMA20>EMA50	399	44.6%	-0.028%	0.94	-11.11 %	+1.44	—	62.6%
3c BTC-1d close>EMA50	316	43.7%	-0.044%	0.91	-14.07 %	-1.51	—	66.2%
<b>3d BTC-1d close&gt;EMA200</b>	<b>0</b>	—	—	—	—	—	—	—
3e BTC-4h full bullish	365	44.9%	-0.013%	0.97	-4.93 %	+7.62	—	58.6%
<b>3f BTC-1d EMA50 AND EMA200</b>	<b>0</b>	—	—	—	—	—	—	—

**Befund BTC-Macro:** - **3d/3f = 0 Trades**, weil BTC im **gesamten 180-d-Fenster nie über EMA200 (1d) schloss** (0 von 200 EMA200-validen Bars) — BTC-Macro-Bärenmarkt. - **3a BTC-4h>EMA20** zeigt **stärkste Test-Performance** (siehe §6, TEST +20.93 %), aber **TRAIN -4.33 %** → Overfitting auf Test-Period. - 3b/3c/3e helfen nicht oder verschlechtern. - Implikation: Wenn man strikt nur in „BTC bullish macro“ gehandelt hätte, wäre der Bot 180 Tage off gewesen → Total = 0 %. Das ist faktisch **besser als -12.55 %**, aber nicht profitabel.

## 5. Circuit-Breaker / Daily-Stop

Variant	N	WR	Exp	PF	Total	Δ	Streak	Worst-Day
4a pause 4h nach 3×SL_LOSS	485	45.6%	-0.000%	0.999	-0.19 %	+12.36	<b>12</b>	-13.19 %
4b pause 24h nach 5×SL_LOSS	490	45.1%	-0.021%	0.96	-10.35 %	+2.21	14	-13.19 %
<b>4c daily stop –2 %</b>	424	46.9%	+0.042%	1.10	<b>+17.94 %</b>	<b>+30.50</b>	12	<b>-4.59 %</b>

**Befund Circuit:** - **4c daily-stop –2 %** ist die effektivste mechanische Bremse: kappt Worst-Day auf -4.59 % (vorher -13.19 %), reduziert Trades um 13.6 %, holt **+30.5 % vs Baseline**. - Aber: PF 1.10 < **1.5 (FAIL)**, P(loss) 27% >> << **10% (FAIL)**. - Walk-Forward: TRAIN +23.27 %, **VALIDATION -7.34 %**, TEST +2.01 % — gleiche VAL-Kollapse-Pattern.

## 6. Walk-Forward 60/20/20 — Top-Kandidaten

**TRAIN = älteste 60 % (Dez 2025-Feb 2026), VALIDATION = mittlere 20 % (Feb-Apr 2026), TEST = letzte 20 % (Apr-Mai 2026).**

Variant	Split	N	WR	PF	Total	DD	P(loss)	CI 95%
<b>2a RANGE-only</b>	TRAIN	156	51.3%	1.80	<b>+33.42 %</b>	8.3%	<b>2.8% ✓</b>	[-0.2, +72.3]
	VAL	52	44.2%	0.47	<b>-11.13 %</b>	13.9%	<b>96.8% ✗</b>	[-22.6, +0.8]
	TEST	53	41.5%	1.17	+3.81 %	10.9%	33.6%	[-15.6, +26.6]
2b no_WEAK	TRAIN	170	48.8%	1.55	+31.84 %	12.5%	5.4%	—
	VAL	57	49.1%	0.52	-10.14 %	13.2%	96.0% ✗	—
	TEST	57	38.6%	0.86	<b>-4.18 % ✗</b>	17.3%	64.4%	—
3a BTC-4h>EMA20	TRAIN	258	43.4%	0.97	<b>-4.33 % ✗</b>	31.9%	55.4%	—
	VAL	86	44.2%	0.73	-8.65 %	14.6%	81.4%	—
	TEST	87	51.7%	1.74	<b>+20.93 %</b>	9.7%	<b>6.6% ✓</b>	[-4.7, +54.7]
4c daily-stop	TRAIN	254	47.6%	1.22	+23.27 %	18.9%	16.8%	—
	VAL	85	45.9%	0.77	-7.34 %	15.0%	83.8%	—
	TEST	85	45.9%	1.05	+2.01 %	19.3%	52.0%	—
5b BTC4h_full+RANGE	TRAIN	110	48.2%	1.76	+22.54 %	7.8%	11.0%	—
	VAL	37	43.2%	0.47	-8.18 %	9.4%	93.6%	—
	TEST	37	40.5%	1.18	+2.64 %	5.1%	33.8%	—

**Kernbeobachtung:** Jede einzelne Filter-Variante kollabiert in **VALIDATION** (Feb-Apr 2026, BTC-Bärenmarkt-Phase). Die VAL-Periode hat strukturell anderes Marktverhalten als TRAIN/TEST.

## 7. Akzeptanz-Matrix (HARTE Operator-Kriterien)

Variant	PF≥1.5	Exp>0	VAL+TEST positiv	P(loss)<<10%	Streak<<15	DD<-38%	Result
Baseline 1d	✗ 0.95	✗	✗/✗	✗ 62%	✗ 15	✗ -38.4%	<b>FAIL</b>
1e 4-of-4	✗ 1.005	✓ 0.002	✗ -15/+5.9	✗ 55%	✗ 15	✓ -26.4%	<b>FAIL</b>

2a RANGE only	X 1.31	✓	X -11.1/+3.8	X 13.6%	✓ 8	✓ -16.9%	FAIL
2e no TRENDS	X 1.31	✓	X -11.1/+3.8	X 12.6%	✓ 8	✓ -16.9%	FAIL
3a BTC-4h>EMA20	X 1.04	✓	X -8.6/+20.9	X 41%	—	✓ -32.9%	FAIL
4c daily-stop	X 1.10	✓	X -7.3/+2.0	X 27%	✓ 12	✓ -19.3%	FAIL
5b BTC4h+RANGE	X 1.28	✓	X -8.2/+2.6	X 19.6%	✓ 8	✓ -15.4%	FAIL
5c BTC1d50+noTRENDS	X 1.06	✓	X -8.7/+4.8	X 45.4%	✓ 8	✓ -16.9%	FAIL

Keine einzelne Variante besteht alle 6 Kriterien gleichzeitig.

## 8. Root-Cause der VALIDATION-Kollapse

- **Zeitraum:** ca. 2026-02-15 bis 2026-04-15 — fällt mit BTC-Crash + Verstärkter Volatilität zusammen.
- **Per-Regime Behaviour in VAL** ist dort wo bei TRAIN RANGE = profitabel war, VAL RANGE-classified Trades plötzlich verlieren → die **Regime-Klassifikation alleine reicht nicht** — das Marktverhalten innerhalb derselben Regime-Klasse ändert sich.
- **Strategie reagiert nicht** auf den Regime-Wechsel weil sie keine adaptiven Exit-Schwellen hat.
- Daher: **Filter-Overlay = Patch, nicht Heilung.**

## 9. Empfehlung

### 9.1 Sofort (READ-ONLY, keine Runtime-Änderung)

1. **MS-Pause beibehalten.** Keine Live-Aktivierung. T1 (Binance only) bleibt aktiv.
2. **MS-MTF-D-SHADOW-LOGGER-1** — Plan bleibt valide, aber **Verschiebung um 1 Phase:** erst Strategy-Re-Design, dann Shadow.
3. **REGIME-AWARE-PARAMS-1** Plan bleibt valide, aber **Eindampfung:** nur Phase P1.A (Shadow) zuerst, P1.B/C (Hard-Gate, Per-Regime-Params) zurückstellen bis Re-Design-Frage geklärt.

### 9.2 Strategische Frage an Operator (vor weiterer Arbeit)

Soll die Strategie als **RANGE-Trader umgebaut werden** (passt zu Validation-Data, +26 % TRAIN, aber Vol-Spike-Resistenz fehlt) — **oder als Trend-Follow mit Regime-Adaptive-Sizing weitergebaut** (passt zum ursprünglichen Namen, aber Performance war −30 % in WEAK\_TREND)?

### 9.3 Was NICHT empfohlen wird

- **X Keine BTC-1d-EMA200-Gate** als sole-source-of-truth (war 180 Tage = 100 % off → wäre verloren).
- **X Keine MTF-Verschärfung auf 4-of-4** (PF 1.005, knapp breakeven, kein Edge).
- **X Keine Coin-Allowlist/Denylist** (Operator-Boundary).
- **X Keine Aktivierung von 2a/2e RANGE-only** ohne adaptive Exits — würde in nächster Bärenphase wieder kollabieren.

### 9.4 Aussagekräftigster Single-Filter (wenn unbedingt einer)

**4c daily-stop −2 %** — mechanisch, modell-frei, kappt nur Worst-Days, bringt +30 % vs Baseline.  
**Empfehlung:** als Add-on einbauen, nicht als sole-source-of-truth.

## 10. Verbleibende Risiken & Follow-ups

Item	Prio	Status
STRATEGY-REDESIGN-1 (RANGE-or-Adaptive Decision)	P1	offen — Operator-Entscheidung needed
MS-MTF-D-SHADOW-LOGGER-1 Plan-Aktivierung	P2	offen — Plan published, GO ausstehend
REGIME-AWARE-PARAMS-1 Phase P1.A only	P2	offen — Plan published
CIRCUIT-BREAKER-DAILY-STOP-2-PCT-1	P3	optional, mechanisch sicher
VAL-PERIOD-DEEP-DIVE (Was war in 2026-02 bis -04 anders?)	P3	open — würde Daten-Drift erklären

## 11. Boundaries (verified)

- 0x Bot-Touch
- 0x Worker-Touch
- 0x GUI-Touch
- 0x DB-Mutation
- 0x Container-Restart
- 0x Mainnet

- 0× Push
  - 0× Secrets
  - 0× Allowlist/Denylist
  - READ-ONLY auf bestehende artifacts/ms-replay-20coin-180d Files
  - BTC 1d-Cache (200→400 bars) ist neue externe Public-API-Pull, kein State
- 

## 12. Files

---

- Code: `/tmp/ms_replay/run_ablation.py` (Ablation-Engine, 391 LoC)
  - Code: `/tmp/ms_replay/show_wf.py` (Walk-Forward-Show, 25 LoC)
  - Results: `/tmp/ms_replay/ablation_results.json` (vollständige Metriken, alle 21 Varianten)
  - Cache: `/tmp/ms_replay/cache/BTCUSDT_1d.json` (400 bars, neu)
- 

**STOP** — Operator-Entscheidung nötig: STRATEGY-REDESIGN-1 (Q: RANGE-Bot vs adaptive Trend) **vor** weiterer Filter-Arbeit. Bis dahin **MS-Pause bleibt**.