

2026.06.24

Strategy Issue



# DS 전략이슈

양형모 투자전략  
02-709-2660  
yhm@ds-sec.co.kr

## 기대를 증명해야 하는 시간

### 현실화된 Base 시나리오

6월 23일 미국 증시는 다시 기술주 중심으로 하락했다. 6월 8일 「과열 청구서」에서 진단한 멀티플 압축과 디레버리징의 흐름이 이어지고 있으며 펀더멘털 이슈가 점차 구체화되는 국면이다. 여전히 추세 하락(약세장)으로 단정하지 않는다. 다만 시장이 AI 수익화의 숫자를 다시 보기 시작한 만큼 변동성 확대는 이미 진행 중이라는 판단이다. 펀더멘털 이슈는 다음과 같다.

하반기 상장을 준비 중인 Anthropic과 OpenAI가 처한 구조적 딜레마다. 사모 시장에서의 외부 조달이 한계에 이른 상황에서 두 거대 랩은 증시 상장으로 향하고 있다. 상장 밸류를 정당화하려면 유저 수와 매출액 증가율, 혹은 최소한 토큰 사용량 성장률이라도 보여줘야 한다. 그러나 두 랩이 동시에 상장을 추진하면서 가격 경쟁이 불가피해졌고, 통제하기 어려운 가격 경쟁과 헤비유저 앵커링(토큰 소비를 키우기 위한)으로 상장 전 숫자를 만드는 작업 자체가 쉽지 않을 것으로 예상된다.

상장 이후도 문제다. 상장 후 자금 조달을 위해서는 이익 성장을 증명해야 하기 때문에 지금처럼 구독료 대비 과도한 사용량을 내주는 정책은 지속되기 어렵다. Anthropic이 6월 출시했다가 중순에 막힌 Fable5의 가격 정책에서 그 방향이 드러난다. Fable5는 6월 22일까지만 구독 사용료로 이용할 수 있고 이후로는 호출 비용으로 부과될 예정이었다. 상장 이후 Claude Max와 같은 20배 사용량 제공이 사라지면 헤비유저는 조금씩 이탈할 수 있겠다.

중국 모델의 추격은 더 가파르다. DeepSeek V4 Flash는 100만 토큰당 입력 \$0.14, 출력 \$0.28로 모델 중 가장 싸고, Kimi K2.7-Code 계열도 성능은 우수하면서 Claude Opus 4.8(\$5/\$25) 대비 약 1/6 수준이다. Claude와 ChatGPT는 에이전트 시스템의 24시간 가동을 전제로 API 호출에서 수익을 내야 하는데 가격을 중국 모델 수준으로 내리지 않으면 비용에 민감한 기업 수요를 붙잡기 어렵다.



두 거대 랩이 동시에 상장을 추진하면서 가격 경쟁이 불가피해졌고, 상장 전 숫자 만들기 자체가 쉽지 않다는 판단

여기에 6월 22일 공개된 Sakana Fugu는 가격이 싸지는 않지만 여러 프론티어 모델을 오케스트레이션해 SWE-Bench Pro, Terminal Bench 등에서 Claude와 ChatGPT 단일 모델을 능가하는 성능을 보였다. 성능 자체가 상향 평준화되는 국면에서 단일 랩이 차별적 경쟁력을 내기가 점점 어려워지고 있다.

이 모든 것이 데이터센터 수요 가정으로 수렴한다. 글로벌 IB에 의하면 2027년까지 추가로 약 15GW의 데이터센터가 지어져야 하고 기가와트당 매출 100~150억 달러를 가정하면 미국 내 AI 매출의 80~90%를 차지(제미나이 제외)하는 Anthropic과 OpenAI의 올해 추정 매출(각 300억 달러 안팎, 합산 약 600억 달러)에 더해 내년에만 1,500~2,250억 달러의 매출 증가가 수반되어야 한다. 가격 경쟁과 헤비유저 이탈, 중국 모델로의 대체가 겹치면 이 증가율 가정은 흔들리고, 데이터센터 Capex를 정당화하는 전제도 함께 흔들릴 수 있을 것이다.

매크로 변수도 다시 부각됐다. 케빈 워시 연준 의장 체제에서 추가 인상 우려가 커지며 연말 인상 베팅이 확대됐다. 우리는 현재 금리 인상 자체에 동의하지 않으나, 시장이 이를 우려하는 한 주가에 프라이싱될 수 있다고 본다. 다만 본질은 다시 펀더멘털이다. AI 인프라(PCB, DRAM, CPU 등) 투자가 물가를 자극할 수 있다는 시각은 파월과 워시 모두 공유하고 있어 인프라 확장에 일부 제동이 걸릴 수 있고 민주당은 이를 전기료 상승 등 정치적 쟁점으로 활용할 여지가 있다.

결론적으로 AI가 혁명이라는 전망은 변하지 않았다. 현실에서는 기술이 발전하고 인프라가 깔리며 진행은 순조롭다. 문제는 주식 시장이 그 장밋빛 베스트 시나리오를 가격에 반영해온 데 있다. 현실의 진행 속도와 주가가 선반영한 기대 사이의 간극이 주가 변동성을 키우는 본질이며 6월의 조정은 그 간극을 좁히는 과정으로 이해한다.

## 펀더멘털 점검

### 이제 주식 발행 국면

AI 기업의 주식 발행은 자력으로 자금을 댈 수 있는지 시장이 의심하기 직전에 꺼내는 유동성 카드라는 판단

AI 구축 비용이 내부 현금과 부채로는 감당되지 않는 수준에 이르자 AI 산업은 보통주 같은 주식 발행 단계로 넘어왔다. AI 기업의 주식 발행은 자력으로 자금을 댈 수 있는지 시장이 의심하기 직전에 꺼내는 유동성 카드라는 판단이다.

Oracle(ORCL): OCI 확장 명목으로 2026년까지 450~500억 달러를 조달하겠다는데 그 절반을 의무전환우선주와 최대 200억 달러 ATM 보통주로 채우려 한다.

Super Micro(SMCI): 같은 문제의 하드웨어 버전이다. 총 70억 달러 규모 주식·주식 연계 거래를 내놨다. 50억 달러 공모(보통주 약 12.5억 + 의무전환우선주 연계 예탁 증권 약 37.5억)에 3분기 이후 최대 20억 달러 ATM 보통주를 더한 구조다. 시장은 희석 신호를 곧바로 읽어 발표 후 5거래일 동안 약 -30% 빠졌다.

Alphabet(GOOG): 세계 최대 현금창출 기업조차 AI 인프라를 위해 외부 자금을 끌어온다. 조달 목표를 800억→약 847.5억 달러로 늘렸고 구조는 보통주+의무전환우선주+버크셔 해서웨이 사모 100억 달러+ATM 400억 달러다.

Meta(META): 사실관계는 아직 초기 단계다. 블룸버그는 수백억 달러 규모 주식 공모 검토를 보도했고 메타는 순전히 추측이라며 부인했다. 그럼에도 광고 캐시카우조차 주식형 조달이 거론될 만큼 신빙성이 생겼다는 점이 핵심이다.

AI Lab: OpenAI는 6/8 8,520억 달러 밸류로 S-1을 비공개 제출하고 600억 달러 조달을 목표로 한다. Anthropic은 약 650억 달러 시리즈 H를 ~9,650억 달러 밸류로 마감(5/28)했음에도 불구하고 연말 상장을 목표로 하고 있다.

표1 종목별 구간 수익률

종목	렐리 3/31~5/29(%)	6월 조정 6/1~6/12(%)	6/12 단일(%)
ORCL	53.0	-25.8	0.02
AVGO	45.0	-16.9	-0.91
NVDA	21.0	-8.5	0.16
MU	187.0	-5.2	-1.43
INTC	160.0	13.9	6.51
WDC	97.0	3.0	6.35
STX	124.0	1.0	7.25
KLAC	31.0	31.2	5.55
AMAT	32.0	23.8	2.64

자료: DS투자증권 리서치센터

### 쓸수록 손해인 AI 랩의 비즈니스 모델

헤비유저가 한도를 채우면  
마진이 claude-max-20x  
의 경우 -900%

AI 랩의 비즈니스 모델은 현재 헤비유저가 쓰면 쓸수록 손해다. SemiAnalysis에 따르면 월 \$200 플랜이 최대 \$8,000(claude)에서 \$14,000(chatgpt) 어치 토큰을 유저에게 내준다. 가격의 40~70배다. API 총마진 75%를 가정해도 헤비유저가 한도를 채우면 마진이 claude-max-20x의 경우 -900%이고 chatgpt-pro-20x는 -1,650%까지 떨어진다.

토큰 비용은 Anthropic의 경우 다음의 네 가지다. 입력(\$5/100만토큰), 출력(\$25, 가장 비쌌), 캐시 쓰기(반복 내용 저장, \$6.25~10), 그리고 캐시 읽기(재사용, \$0.50, 90%↓)다.

긴 작업일수록 캐시 읽기가 비용을 지배한다. 한 개발자의 Claude Code 30일 실사용을 Opus 4.7 list-price로 환산하면 입력 120만 토큰×\$15=\$18, 출력 3,994만 토큰×\$75=\$2,995, 캐시 쓰기 73.4억 토큰×\$1.5=\$11,014, 캐시 쓰기 2.15억 토큰×\$30=\$6,462로 총 \$20,489였다. 캐시 읽기 73억 토큰이 전체의 54%를 차지했다. \$200 구독의 102배다. 다만 이는 Opus 4.7 단가 기준이며 현행 Opus 4.8 단가로 같은 사용량을 환산하면 약 \$6,830(34배)으로 내려온다. 어느 단가든 가격이 저렴한 캐시 읽기도 양이 많으면 거대한 비용이 되며 이 구조가 헤비유저(개발자) 적자의 근원이다.

적자를 줄려면 가격을 올리거나 사용을 제한해야 하는데 상장을 앞둔 지금은 정반대로 가격 경쟁 중이다. 상장 밸류를 정당화할 런레이트와 유저 수를 키워야 하기 때문이다. OpenAI는 Anthropic의 인하에 맞서 토큰 가격을 급격히 인하할 필요가 있고 Anthropic도 상장 전까지 유저, 런레이트를 끌어올려야 한다. 그러나 상장 후(혹은 이전)에는 이 그림이 뒤집힐 수 있겠다. 즉 이미 Anthropic의 Fable5(6/22 종료였으나 현재 사용 제한)가 예고한 대로 종량제로 옮겨가는 것이다. 그러면 유저의 토큰 사용이 줄고 예상 대비 GW 데이터센터 수요가 하향될 수 있다.

표2 SemiAnalysis가 공개한 구독 플랜별 마진

플랜	가격(\$)	최대지출(\$)	1.0%	5.0%	5.7%	10.0%	11.4%	15.0%	20.0%	25.0%	50.0%	100.0%
claude-pro	\$20	\$400	95%	75%	72%	50%	43%	25%	0%	-25%	-150%	-400%
claude-max-5x	\$100	\$2,000	95%	75%	72%	50%	43%	25%	0%	-25%	-150%	-400%
claude-max-20x	\$200	\$8,000	90%	50%	43%	0%	-14%	-50%	-100%	-150%	-400%	-900%
chatgpt-plus	\$20	\$700	91%	56%	50%	13%	0%	-31%	-75%	-119%	-338%	-775%
chatgpt-pro-5x	\$100	\$3,500	91%	56%	50%	13%	0%	-31%	-75%	-119%	-338%	-775%
chatgpt-pro-20x	\$200	\$14,000	83%	13%	0%	-75%	-100%	-163%	-250%	-338%	-775%	-1650%

자료: SemiAnalysis, DS투자증권 리서치센터

### 경쟁 구도 — 가격은 중국, 성능은 상향 평준화

DeepSeek V4 Flash  
\$0.14/\$0.28로 1선급 최저  
가. Sakana Fugu는 단일  
프론티어 모델을 능가

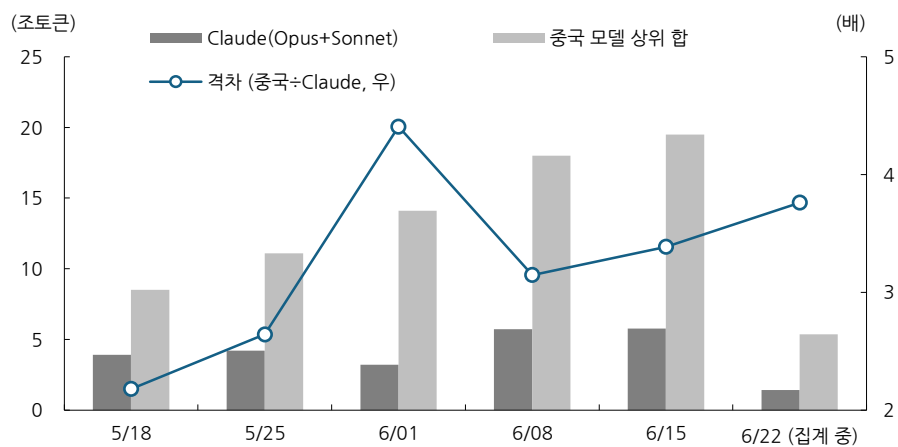
수요 측 균열의 핵심은 가격이다. DeepSeek V4 Flash는 100만 토큰당 입력 \$0.14·출력 \$0.28로 현재 1선급 모델 중 가장 싸며, Kimi K2.7-Code 계열도 입력 약 \$0.95로 Claude Opus 4.8(\$5·\$25) 대비 약 1/5~1/6 수준이다.

OpenRouter 토큰 사용량에서 중국 모델 계열이 상위권을 매우는 흐름이 이어지는 가운데, 에이전트 시스템의 24시간 가동을 전제로 API 호출에서 수익을 내야 하는 미국 랩은 가격을 중국 모델 수준으로 내리지 않는 한 비용에 민감한 기업 수요를 방어하기 어렵다. 기업이 토큰 종량제로 전환한 이후 비용 통제가 본격화된 점도 같은 방향이다.

성능 측면의 변화도 위협적이다. 6월 22일 Sakana AI가 공개한 Fugu-Fugu Ultra는 여러 프론티어 모델(Gemini 3.1 Pro·Claude Opus 4.8·GPT-5.5)을 단일 API로 오케스트레이션하는 구조로, SWE-Bench Pro(73.7), Terminal Bench(82.1), LiveCodeBench(93.2) 등 다수 벤치마크에서 개별 단일 모델을 능가했다.

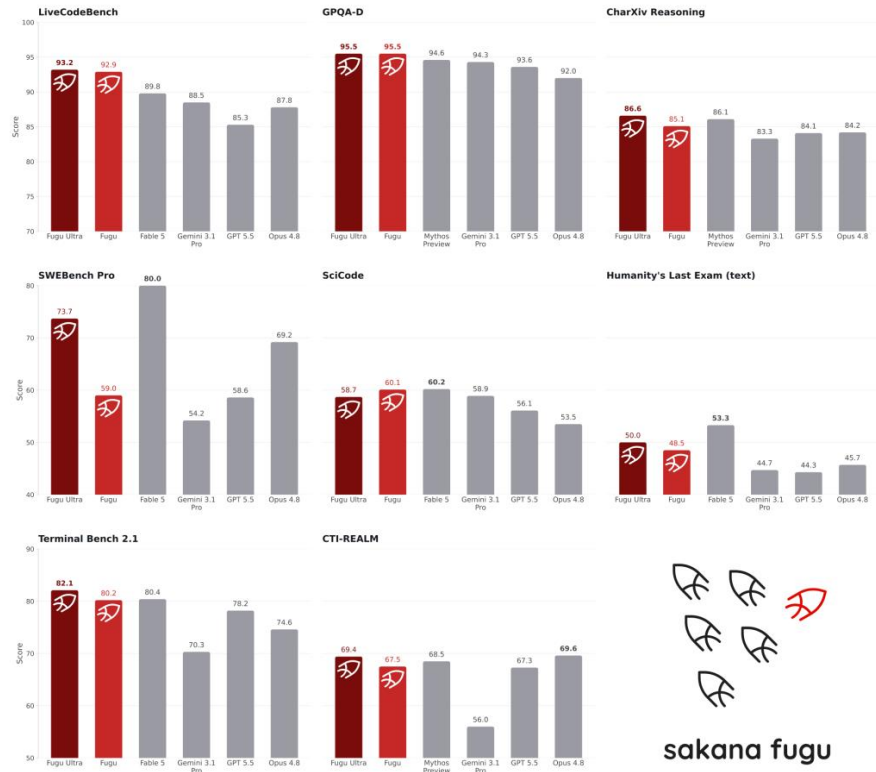
단일 모델만 선택하는 기본 Fugu조차 Terminal Bench에서 풀 내 최고 단일 모델을 앞섰다. 가격이 싸지 않더라도 성능이 상향 평준화되는 국면에서는 단일 랩이 차별적 우위를 유지하기가 어려워진다. 더욱이 Fugu의 종량제는 활성 에이전트 풀에 대한 단일 블렌드 레이트로 과금돼, 여러 모델을 쓰더라도 청구액이 곱해지지 않는다는 점에서 비용 경쟁을 한층 자극한다.

그림1 OpenRouter 토큰 사용량



자료: OpenRouter, DS투자증권 리서치센터

그림2 Sakana Fugu 성능



자료: Sakana Fugu, DS투자증권 리서치센터

그림3 Sakana Fugu 성능

Benchmark	Fugu	Fugu Ultra	Opus 4.8 <sup>†</sup>	Gemini 3.1 Pro <sup>†</sup>	GPT 5.5 <sup>†</sup>
SWE Bench Pro <sup>*</sup>	59.0	73.7	69.2	54.2	58.6
TerminalBench 2.1	80.2	82.1	74.6	70.3	78.2
LiveCodeBench	92.9	93.2	87.8	88.5	85.3
LiveCodeBench Pro	87.8	90.8	84.8	82.9	88.4
Humanity's Last Exam	47.2	50.0	49.8	44.4	41.4
CharXiv Reasoning	85.1	86.6	84.2	83.3	84.1
GPQA-D	95.5	95.5	92.0	94.3	93.6
SciCode	60.1	58.7	53.5	58.9	56.1
τ <sup>3</sup> Banking	21.7	20.6	20.6	8.4	20.6
Long Context Reasoning	74.7	73.3	67.7	72.7	74.3
MRCrv2	86.6	93.6	87.9	84.9	94.8

자료: Sakana Fugu, DS투자증권 리서치센터

### 데이터센터 병목과 단기 성장 둔화 가능성

결국 앞으로는 AI 랩의 압도적인 매출 성장이 실제로 나오지 않으면 앞서 살펴본 리스크가 현실화 가능

글로벌 IB는 AI 데이터센터 전력 사용량이 현재 7.7GW인데 22.7GW까지 성장할 것으로 추산했다. 또한 전체 미국 데이터센터 전력은 2025년 31GW→2026년 41GW→2027년 66GW이며 AI 비중 약 27%(~18GW)로 전망했다. 이는 Bull 시나리오에 미치지 못하는 베이스 시나리오와 맞아떨어진다.

현재 발표된 데이터센터 용량은 전력 기준 190GW다. 젠슨 황이 언급한 GW당 800~1,000억달러를 감안하면 신규 건설 시 약 15~19조 달러, 보수적으로 절반만 신규로 잡아도 약 9.5조 달러가 필요하다. 즉 실제 구축 5~10GW 기준 연 5,000억~1조 달러가 필요할 것으로 예상되는데 자금 조달 자체가 병목이 되어 데이터센터가 예정대로 안 지어질 수 있다는 우려가 커지고 있다. 또한 송전, 변압기, 계통연결의 리드타임이 수년에 달하고 있어 자본이 있어도 전력이 제때 들어오지 않는 리스크도 상존한다.

최근 Uber가 연간 토큰 예산을 4개월 만에 소진하고 사용자당 월 \$1,500로 묶은 사례가 있다. 또한 한 기업이 한 달에 Anthropic에 수억 달러를 쓴 사례가 발생했다. 기업 수요는 가격에 민감(탄력적)할 수 밖에 없다. 즉 단가가 비싸면 예산을 닫을 수 있다는 것이다. 여기에 OpenRouter에서 확인할 수 있는 바와 같이 더 싼 중국 모델로 옮겨가는 대체 위험까지 겹치면 AI 매출이 매년 두 배라는 전제가 흔들릴 수 있겠다. 또한 부채로 버티던 생태계가 주식 발행으로 넘어온 것은 부채 여력이 한계에 다다랐다는 신호이며 밸류에이션 하락으로 발행 창구마저 좁아지면 순환금융 고리가 역전될 가능성도 배제할 수 없다. 결국 앞으로는 AI 랩의 압도적인 매출 성장이 실제로 나오지 않으면 앞서 살펴본 리스크가 현실화될 수 있다.

결론적으로 모든 균열은 데이터센터 수요 가정으로 수렴한다. 2027년까지 약 15GW의 신규 데이터센터가 더 지어져야 하고 기가와트당 매출 100~150억 달러를 가정하면 이를 채울 매출 규모가 필요하다. 현재 미국 내 AI 매출의 80~90%를 차지하는 Anthropic과 OpenAI의 올해 추정 매출은 각각 300억 달러 안팎, 합산 약 600억 달러에 그친다. 그렇다면 내년에만 1,500~2,250억 달러의 매출 증가가 추가로 수반되어야 데이터센터 투자가 정당화된다. 가격 경쟁, 헤비유저 이탈, 중국 모델로의 대체가 겹치면 AI 랩 매출이 매년 두 배로 큰다는 전제가 흔들리고, 수익화 곡선이 Capex 곡선을 따라가지 못한다는 문제가 구체화된다.

단, 하드웨어가 변수다. 만약 발표된 수치가 정확하여 Vera Rubin이 토큰 원가를 1/10로 낮춘다면 (NVIDIA GTC 2026), 같은 사용량의 원가가 \$2,000 → \$200/월로 내려와 구독료와 손익분기에 도달 가능하다. 즉 하드웨어가 적자를 흑자로 바꿀 수 있다. 그러나 이는 구독료가 \$200으로 유지된다는 전제이며 치킨게임으로 가격이 더 내려가면 상쇄될 수 있다.

표3 원가와 손익분기 계산

단계	값(\$/월)	산식/비고
list 가치	\$8,000	claude-max-20x API 평가 환산
원가(Blackwell)	\$2,000	list 가치 × (1 - 총마진)
헤비유저 손익	(\$1,800)	구독료 - Blackwell 원가
원가(Rubin 1/10)	\$200	Blackwell 원가 ÷ 10
Rubin 손익	\$0	구독료 - Rubin 원가 → 손익분기

자료: DS투자증권 리서치센터

### 6월 23일 — 반도체가 다시 진앙

시장이 ROE 높은 고퀄리티 AI 종목을 팔아 ROE 낮은 디펜시브를 사는 디레버리징 구도

6월 23일 다우는 견조했으나 대형 기술주에 매도가 집중됐다. 방어 구조는 지난 6월 초와 동일한 성격이다. Microsoft(+1.80%, +7.8bp)가 단독으로 방어했고 Johnson & Johnson, Walmart, Amazon, IBM, Merck, Berkshire Hathaway, AbbVie 등 헬스케어/필수소비재/보험이 뒤를 이었다. 순수 디펜시브로의 이동이 재차 확인된다. 다만 이날 하락의 진앙은 AI 인프라 자본 지출이 그에 걸맞은 수익으로 이어지지 못한다는 의구심이며, 알파벳은 딥마인드 노벨상 수상자 존 점퍼 부사장의 Anthropic 이적 소식까지 겹치며 약세를 보였다.

6월 1일~23일로 넓히면 진앙이 달라진다. 하락을 주도한 것은 Microsoft(-18.8%), NVIDIA(-10.8%), Broadcom(-17.4%), Alphabet(-8.0%), Amazon(-10.4%), Oracle(-33.4%) 등 AI 플랫폼 및 하이퍼스케일러 메가캡이다. Palantir(-27.4%), Netflix(-15.2%), Salesforce(-26.8%), Accenture(-35.4%), ServiceNow(-29.4%) 등 고밸류 소프트웨어도 큰 폭으로 빠졌다. 6월 8일 보고서가 지적한 가장 인기 있던 종목의 멀티플 압축이 구간 전체로 확장된 모습이다.

하락군과 상승군의 펀더멘털을 집계하면 6월 8일 진단이 그대로 유지된다. 6월 23일 하락군(192종목)의 종합 ROE는 29.5%로 상승군(262종목) 17.4%를 크게 상회한다. 시장이 ROE 높은 고퀄리티 AI 종목을 팔아 ROE 낮은 디펜시브를 사는 디레버리징 구도다.

하락군의 펀더멘털 자체는 훼손되지 않았다. 하락군의 ROE는 6월 내내 26~29%대로 견조하게 유지됐고 매출·순이익 컨센서스도 1개월 기준 상향을 이어갔다. 변한 것은 가격(멀티플)이다. 다만 6월 8일 보고서에서 변수로 제시한 리비전 브래드스의 둔화, 낙수효과 반전(Top7 제외 종목군의 성장률이 Top7을 다시 하회), 사이클 정점 신호는 여전히 유효하다. 피크 아웃 일 수 있다는 가정 자체가 멀티플 압축의 빌미가 되는 국면이다.

표4 하락군과 상승군의 펀더멘털 점검

구분	6/23 하락군	6/23 상승군	6/1~23 하락군	6/1~23 상승군
종목수	192	262	185	269
ROE	29.5%	17.40%	26.20%	18.30%
PER	22.3x	18.0x	20.1x	20.2x
구간 수익률	-2.5%	1.60%	-8.8%	6.20%
1M 수익률	-2.9%	0.90%	-5.3%	5.50%
해석	디레버리징	디펜시브	멀티플 압축	로테이션 수혜

자료: DS투자증권 리서치센터

## Compliance Notice

---

- 동 자료는 기관투자가 등 제 3자에게 사전 제공한 사실이 없습니다.
  - 동 자료에 게시된 내용들은 본인의 의견을 정확하게 반영하고 있으며, 외부의 부당한 압력이나 간섭 없이 작성되었음을 확인합니다. (작성자: 양형모)
  - 동 조사자료는 고객의 투자에 참고가 될 수 있는 각종 정보제공을 목적으로 제작되었습니다. 이 조사자료는 당사의 리서치센터가 신뢰할 수 있는 자료 및 정보로부터 얻어진 것이나, 당사가 그 정확성이나 완전성을 보장할 수 없으므로 투자자 자신의 판단과 책임하에 종목 선택이나 투자시기에 대한 최종 결정을 하시기 바랍니다. 따라서 이 조사자료는 어떠한 경우에도 고객의 증권투자 결과에 대한 법적 책임소재의 증빙자료로 사용될 수 없습니다.
  - 동 조사자료의 지적재산권은 당사에 있으므로 당사의 허락 없이 무단 복제 및 배포 할 수 없습니다.
-