

생성모형 기반 합성 시계열 자료 증강이 백테스팅 과적합 평가에 미치는 영향

이유섭*, 최인수**

Effects of Generative-Model-based Synthetic Time-Series Data Augmentation on Backtest Overfitting Evaluation

Yuseop Lee*, Insu Choi**

요 약

본 연구는 생성모형 기반 합성 시계열 자료를 활용한 증강이 백테스팅 과적합 확률(PBO, Probability of Backtest Overfitting)을 완화하는지를 실증한다. 비적대적 학습 기반 생성 모델인 일치 신경망(GMMN)과 변분 자동부호기(VAE), 그리고 비모수 기준선인 정상 부트스트랩(SB)을 GICS 11개 산업의 미국 ETF에서 비교하며, 시간순 분할(F-augment 대 W-augment)로 표본 외 정보 누출을 통제한다. W-augment 분기에서 GMMN이 PBO를 유의하게 완화하며 SB와 VAE보다 우월하였다. 한편 양식화 사실(Stylized facts) 보존에서는 SB가 우월하여, 시계열 충실성과 과적합 완화 사이에 상충관계(Trade-off)가 존재함을 보인다. 합성 자료 증강의 실효성은 단일 지표가 아닌 다중 패러다임 비교로만 평가 가능하다.

Abstract

This study examines whether synthetic time-series augmentation alleviates the Probability of Backtest Overfitting (PBO). Two non-adversarial generative models—Generative Moment Matching Networks (GMMN) and Variational Autoencoders (VAE)—are compared against the Stationary Bootstrap (SB) on U.S. industry ETFs across the eleven GICS sectors, with two temporal-split regimes (F-augment and W-augment) isolating information leakage. Under W-augment, GMMN significantly reduces PBO and outperforms SB and VAE. Meanwhile, SB best preserves stylized facts, yielding a trade-off between stylized-fact fidelity and overfitting-risk mitigation. Evaluating synthetic financial data thus requires multi-paradigm comparison aligned with the evaluation purpose.

Keywords

backtest overfitting, combinatorially symmetric cross-validation, synthetic data augmentation, generative moment matching networks, stationary bootstrap

* 한국과학기술원 산업및시스템공학과 학부연구원
- ORCID: <https://orcid.org/0009-0006-7930-4200>
** 가천대학교 금융·빅데이터학부 조교수(교신저자)
- ORCID: <https://orcid.org/0000-0003-2596-7368>

• Received: May 14, 2026, Revised: Aug. 16, 2026, Accepted: Aug. 19, 2026
• Corresponding Author: Insu Choi
Dept. of Finance and Big Data, Gachon University
Tel.: +82-31-750-5405, Email: ischoi@gachon.ac.kr

1. 서 론

체계적 투자전략의 개발은 본질적으로 자료 마이닝의 위험을 동반한다. 동일한 과거 자료에 다수의 사양을 시험할수록 표본 내 성과가 우월해 보이는 전략을 우연히 발견할 확률이 증가하며, 이렇게 선정된 전략은 표본 외 운영 단계에서 성과 저하를 경험할 가능성이 크다[1][2]. Bailey et al.[3]은 사양 탐색 위험을 정량화하기 위해 조합 대칭 교차검증(CSCV, Combinatorially Symmetric Cross-Validation)을 제안하고, 표본 외 성과가 표본 내 중앙값에 미달할 확률로 정의된 백테스팅 과적합 확률(PBO, Probability of Backtest Overfitting)을 도입하였다.

PBO 추정의 정밀도는 평가 시계열의 길이에 직접적으로 의존한다. 그러나 자산운용 실무에서 신뢰할 수 있는 평가창은 대개 5~10년 수준에 머물며, 이는 거시 국면별 강건성 진단에 필요한 표본 다양성을 확보하기에는 부족하다. 합성 시계열 자료를 통한 증강은 이러한 표본 부족을 완화하는 한 경로이며, 최근 다수의 연구가 적대적 생성 신경망(GAN, Generative Adversarial Networks) 계열 모형을 금융 시계열 합성에 적용해 왔다[4]-[6].

그러나 GAN 계열 모형은 (i) 학습 불안정성과 모드 붕괴(Mode collapse)로 합성 표본의 일관성이 저해되고[7], (ii) 발산·결측 표본의 사후 제거가 선택 편향을 유발하며, (iii) 적대적 손실이 양식화 사실(자기상관, 변동성 군집, 두꺼운 꼬리, 자산 간 상관구조)을 명시적으로 보존하지 않아 분포 일치와 사전적 보장이 부족하다는 제약이 있다.

이러한 배경에서 본 연구는 적대적 학습을 사용하지 않는 두 생성모형, 곧 최대 평균 불일치(MMD, Maximum Mean Discrepancy)를 직접 최소화하는 GMMN(Generative Moment Matching Networks)[8]과 명시적 증거하한(ELBO, Evidence Lower Bound)을 최대화하는 VAE(Variational Autoencoders)[9]를 비모수 대표본추출 기준선인 정상 부트스트랩(SB, Stationary Bootstrap)과 함께 비교한다. 분석 대상은 GICS(Global Industry Classification Standard) 11개 산업 분류를 상호배타적으로 포괄하는 산업별 상장지수펀드(Exchange-Traded Fund, ETF) 11종(XLK, XLF, XLV,

XYL, XLP, XLE, XLI, XLB, XLU, IYR, VOX)이며, 분석 기간은 2020년 1월부터 2025년 12월까지의 6년이다. 합성 자료의 학습 윈도우와 평가 윈도우의 관계가 결과에 미치는 영향을 통제하기 위해 동일한 분석 프레임워크가 두 시간 분할 분기, 즉 F-augment(전 기간 학습)와 W-augment(엄격 시간순 분할)에서 동시에 평가된다. W-augment 분기에서는 학습 구간(2020~2023)과 평가 구간(2024~2025)이 분리된다.

본 연구의 학술적 기여를 다음 세 측면에서 정리할 수 있다. 첫째, 비적대적 생성 패러다임을 통한 안정성과 모멘트 일치의 동시 확보이다. GMMN의 학습 목적함수인 MMD는 무한 차원 재생 핵 힐베르트 공간(RKHS, Reproducing Kernel Hilbert Space)에서의 분포 간 거리이며, 이를 최소화하는 것은 모든 차수의 모멘트를 일치시키는 것과 동가적이다. 양식화 사실의 보존이 모형의 학습 목적함수 자체에 명시적으로 포함되는 구조이다. 둘째, 시간 분할의 엄격성을 분리한 이중 실험 설계이다. F-augment는 합성 자료가 평가 구간을 학습한 비현실적 상한을, W-augment는 표본 외 정보 누출이 원리적으로 불가능한 보수적 하한을 제공하며, 두 분기의 비교를 통해 합성 자료 평가에서 시간 분할이 차지하는 영향이 분리되어 식별된다. 셋째, 양식화 사실 보존과 PBO 완화 사이의 비단조적 관계 정량화이다. 통상 별개로 보고되어 온 두 평가 차원, 즉 합성 자료의 시계열 충실성과 과적합 평가에서의 실용적 효과가 동일 실험 위에서 동시에 측정되며, 시계열 충실성에서 우월한 SB가 PBO 완화 효과에서는 GMMN에 미치지 못하는 결과는 합성 자료의 평가 기준이 단일 지표에 의존해서는 안 됨을 시사한다. 본 연구의 차별성에 관한 선행연구 대비 상세 비교는 §2.4에 별도 절로 정리된다.

논문의 구성은 다음과 같다. 제2장에서는 백테스팅 과적합과 합성 자료 증강에 관한 선행연구가 비판적으로 종합된다. 제3장은 분석 자료, 시간 분할 설계, GMMN, VAE, SB 세 모형의 수리적 정식화를 제시한다. 제4장은 PBO 결과, 양식화 사실 진단, 민감도 분석을 보고한다. 제5장은 결과의 학술적 함의와 자산운용 실무 시사점을 논의하며, 제6장에서 결론과 후속 연구 방향이 제시된다.

II. 관련 연구

2.1 백테스팅 과적합과 CSCV 프레임워크

체계적 투자전략의 표본 내·외 성과 격차에 관한 문제는 Harvey and Liu[1], Harvey et al.[10], López de Prado[2]의 일련의 연구에서 학술적으로 정착되었다. Bailey et al.[11]은 다중 가설 검정 환경의 사양 탐색이 야기하는 의사 통계적 유의성을 지적하고, 표본 내 우위가 표본 외에서 어떻게 사라지는지에 관한 분석적 한계를 도출하였다. Bailey et al.[3]은 이 통찰을 운영 가능한 평가 절차로 발전시켜 CSCV를 제시하였다. CSCV는 표본을 S개의 동일 길이 블록으로 분할한 후 C(S, S/2)개의 IS-OOS 조합 위에서 표본 외 등수 분포를 추정하며, 이로부터 PBO를 도출한다. 이 절차는 시간 의존성을 부분적으로만 보존한다는 한계가 있으나, 사양 탐색 위험을 정량화하는 가장 널리 채택된 방법이다.

후속 연구는 CSCV의 특정 한계를 보완하는 데 집중되어 왔다. Harvey and Liu[12]는 다중 가설 검정 보정의 관점에서 백테스트 결과의 강건성을 평가하였고, Romano and Wolf[13] 및 White[14]는 자료 탐사(Data snooping) 보정의 통계적 절차를 제시하였다. Sullivan et al.[15]는 거래 규칙의 자료 탐사 효과를 부트스트랩으로 정량화한 초기 연구이다. López de Prado and Bailey[16]은 다중 검정 보정을 반영한 디플레이션된 샤프 비율(Deflated Sharpe ratio)을 제안하였다. 그러나 이러한 연구는 평가창 자체의 길이가 제약되어 발생하는 PBO 추정의 분산 문제에 직접 대응하지 않으며, 본 연구는 합성 자료 증강을 통한 평가창 확장의 가능성을 제시함으로써 이러한 공백을 보완한다.

2.2 금융 시계열의 합성 자료 생성과 GAN의 작동 원리

본 연구의 비교 대상이 되는 GAN 계열 모형의 작동 메커니즘은 다음과 같이 정리된다. GAN은 두 신경망의 적대적 게임으로 자료의 분포를 학습한다. 생성기 G_θ 는 잠재 변수 $z \sim N(0, I)$ 로부터 합성

표본을 생성하고, 판별기 D_ϕ 는 입력이 실제 자료인지 합성 자료인지를 분류한다. 두 신경망의 손실은 서로의 성능을 꺾어내리는 방향으로 정의되어 미니맥스 게임을 형성한다. 이론적으로 균형에 도달하면 합성 분포가 실제 분포와 일치하나, 실무 학습에서는 (i) 판별기가 너무 강해져 생성기 기울기가 소실되거나, (ii) 생성기가 판별기를 속이기 위한 한 두 개의 모드만 학습하는 모드 붕괴, (iii) 학습 발산으로 인한 결측치 산출 등이 빈발한다. 이 한계는 Salimans et al.[7], Li et al.[17]을 비롯한 다수의 후속 연구에서 일관되게 보고되어 왔다.

금융 시계열 응용에서 Wiese et al.[4]는 시간 합성곱 신경망(TCN, Temporal Convolutional Network)에 기반한 QuantGAN을 제안하여 단변량 주가 수익률 합성에서 양식화 사실의 부분적 보존을 보고하였다. Yoon et al.[5]의 TimeGAN은 비지도 적대적 학습과 지도학습 손실을 결합한 하이브리드 구조이며, Lin et al.[6]의 DoppelGANger는 다변량 시계열의 속성과 시계열을 분리 모델링하는 구조를 도입하였다. Esteban et al.[18]은 의료 시계열 영역에서 RGAN을 제시하였고, Koshiyama et al.[19]는 거래 전략의 미세조정에 GAN을 활용하였다. 그러나 이들 연구는 모드 붕괴와 결측치 사후 제거의 선택 편향 문제를 공유하며, 본 연구의 비적대적 패러다임 채택의 동기가 된다.

본 연구가 적대적 학습 기반 모형(TimeGAN, DoppelGANger, QuantGAN 등) 대신 비적대적 모형을 채택한 직접적 근거는 다음과 같다. 첫째, 적대적 학습 동학의 구조적 불안정성으로 인해 합성 표본 추출에서 결측치가 빈번히 발생한다. 본 연구의 사전 점검에서 DoppelGANger를 동일 자료에 적용한 결과 100개 합성 표본 중 32개에서 결측치가 발생하였으며, 이를 사후적으로 제거하는 절차는 선택 편향을 야기한다. 둘째, 적대적 손실함수는 시계열의 양식화 사실(자기상관, 변동성 군집, 두꺼운 꼬리, 자산 간 상관구조)을 명시적으로 보존하지 않으므로 학습된 분포의 실증적 일치 정도에 관한 사전적 보장이 부족하다. 한편 Diffusion 기반 생성모형은 금융 시계열 응용 사례가 본 연구 진행 시점 기준 제한적이며, 이중 분기(F-augment·W-augment) 평가, K=20 시드 매칭, 11종목 부트스트랩 신뢰구간

(부트스트랩 10,000회) 산출 등 본 연구의 강건성 검증 절차를 가용 전산 자원 내에서 수행하기에 요구 자원이 상당하다. 따라서 본 연구는 비적대적 학습 패러다임 중 학습 동학의 안정성과 명시적 분포 일치(MMD 최소화) 혹은 명시적 변분 하한 최대화(ELBO)를 직접 보장하는 두 핵심 모형, 즉 GMMN과 VAE를 채택하였다.

2.3 비모수 부트스트랩과 비적대적 생성 패러다임

비모수 재표본추출 방법으로서 Politis and Romano[20]의 정상 부트스트랩(SB)은 평균 블록 길이를 매개변수로 하는 기하분포로부터 가변 블록 길이를 추출하여 자기상관 구조를 부분적으로 보존한다. Politis and White[21]은 자료 의존적 블록 길이 자동 선택 절차를, Patton et al.[22]는 SB의 점근적 성질에 대한 추가 분석을 제공하였다. Politis[23]은 시계열 부트스트랩의 응용을 종합하였다. Efron[24]의 본래 부트스트랩 정신과 같이, SB는 단순한 통계적 가정 하에서 강건한 추정을 가능케 한다는 장점이 있으나, 합성 표본이 본질적으로 원자료의 재구성에 머물러 새로운 결합 패턴을 생성하지 못한다는 구조적 제약이 있다.

비적대적 생성 패러다임의 첫 번째 축은 Li et al.[8]이 제안한 생성 모멘트 일치 신경망(GMMN)이다. GMMN은 적대적 판별기 없이 MMD를 직접 최소화하는 모형으로, 학습 동역학의 안정성과 모멘트 일치의 명시적 보장을 동시에 달성한다. Dziugaite et al. [25]는 MMD 기반 생성모형의 이론적 성질을 분석하였으며, Li et al.[17]은 MMD-GAN으로의 확장을 제시하였다. 두 번째 축은 Kingma and Welling[9]의 변분 자동부호기(VAE)이며, 잠재변수 모형의 우도 기반 학습이라는 또 다른 비적대적 경로를 제공한다. ELBO 최대화는 명시적이고 미분가능한 목적함수이며, 학습 안정성이 원리적으로 보장된다. 금융 응용에서 Bühler et al. [26]은 VAE 계열을 시장 시뮬레이션에 적용한 바 있으나, 백테스팅 과적합 평가의 맥락에서 PBO에 미치는 영향은 본격적으로 평가된 바 없다.

2.4 본 연구의 차별성

선행연구가 (i) 합성 자료의 양식화 사실 보존을 평가하거나(Wiese et al.[4], Cont[27]), (ii) 합성 자료를 활용한 단일 모형의 백테스트 성과를 보고하는데(Koshiyama et al.[19]) 그쳤다면, 본 연구는 두 차원을 통합하여 합성 자료의 시계열 충실성과 과적합 평가에서의 실용적 효과를 동일 실험 위에서 동시에 측정한다. 또한 시간 분할의 엄격성(F-augment 대 W-augment)이 별도의 통제 변수로 분리되며, 적대적 학습 패러다임을 벗어난 두 모형(GMMN, VAE)과 비모수 기준선(SB)의 직접 비교가 수행된다. 이러한 설계는 Prado[2]이 제기한 사양 탐색 위험관리의 관점에서 합성 자료 증강의 한계와 가능성을 동시에 평가하는 학술적 가치를 지닌다.

III. 자료 및 연구 방법론

3.1 분석 자료

분석 대상은 GICS 11개 산업 분류를 상호배타적으로 포괄하는 미국의 산업별 ETF 11종이며, 종목 구성은 표 1에 제시되어 있다. 종목 선정의 임의성을 제거하고 자산 간 상관구조를 분석가능한 형태로 정의하기 위해, 각 ETF가 GICS 분류 중 정확히 하나의 산업에 대응하도록 구성된다. 자산배분 의사결정의 실제 단위인 ETF가 분석 단위로 채택됨으로써 본 연구의 결과가 자산운용 실무에 직접 적용 가능하다는 운영적 의의를 확보한다.

표 1. 분석 대상 ETF 구성 (GICS 11개 상호배타적 산업)
Table 1. Composition of Industry ETFs (Eleven mutually exclusive GICS sectors)

Ticker	Sector (GICS)	Issuer
XLK	Information technology	SPDR
XLF	Financials	SPDR
XLV	Health care	SPDR
XLY	Consumer discretionary	SPDR
XLP	Consumer staples	SPDR
XLE	Energy	SPDR
XLI	Industrials	SPDR
XLB	Materials	SPDR
XLU	Utilities	SPDR
IYR	Real estate	iShares
VOX	Communication services	Vanguard

가격 자료는 일별 수정종가이며, 분석 기간은 2020년 1월 1일부터 2025년 12월 31일까지(약 1,508 영업일)이다. 일별 로그수익률 $r_{i,t} = \log(P_{i,t}/P_{i,t-1})$ 이 분석의 기본 단위이다. 변동성 산정에 사용되는 사전 표본은 분석 시작 이전 구간에서 추출되어 평가 정보의 누출을 방지한다.

3.2 시간 분할 설계

본 연구는 두 시간 분할 분기 위에서 동일한 분석을 병렬적으로 수행한다.

전 기간 학습(F-augment): 합성 모형이 분석 전 기간을 학습 자료로 활용하고, 동일 기간을 CSCV 평가에 사용하는 분기이다. 이는 합성 자료가 평가 구간 정보를 사전에 관찰한 비현실적 상한 시나리오에 해당하며, 표본 외 정보 누출의 영향을 의도적으로 허용한 비교 기준점으로 기능한다.

엄격 시간순 분할(W-augment): 합성 모형의 학습 자료는 2020년 1월 1일부터 2023년 12월 31일까지(약 1,006일)로 한정되고, CSCV 평가는 2024년 1월 1일부터 2025년 12월 31일까지(약 501일)의 미관측 구간에서만 수행된다. 표본 외 정보가 합성 자료에 누출될 가능성이 원리적으로 차단되며, 자산운용 실무의 전진 갱신 방식의 운영 시나리오와 일치한다.

두 분기의 결과 비교를 통해 합성 자료의 효과가 시간 분할의 엄격성에 어떻게 의존하는지가 분리되어 식별된다. 이러한 설계는 IS-OOS 분리 원칙의 엄정한 통제 하에서 합성 자료의 운영적 가치를 평가할 수 있게 한다.

3.3 전략 격자와 CSCV

전략 격자는 진입일·보유기간·손절 배수·매매 방향의 네 차원으로 구성된다. 손절은 진입 후 누적수익률이 사전표본 변동성에 손절 배수를 곱한 음의 임계값에 도달하면 발동된다. 각 종목·전략 조합에 대해 일별 성과 행렬 M (T 는 거래일 수, N 은 전략 수)이 산출된다.

CSCV의 블록 수는 $S=16$ 이고, 조합 수는 $16C_8 = 12,870$ 이다. 각 조합에 대해 표본 내(IS, In-Sample)

에 해당하는 8개 블록과 표본 외(OOS, Out-Of-Sample)에 해당하는 8개 블록의 통계량 텐서 $SS \in \mathbb{R}^{S \times N \times 3}$ 이 사전 계산되며, 이로부터 IS 최적 전략의 OOS 등수 분포가 추정된다. PBO는 OOS 등수가 중앙값 미만일 확률로 정의된다.

3.4 생성 모형의 수리적 정식화

3.4.1 생성 모멘트 일치 신경망(GMMN)

Li et al.[8]이 제안한 GMMN은 잠재변수 $z \sim N(0, I)$ 를 자료 공간으로 사상하는 생성기 $G_\theta: \mathbb{R}^{d_z} \rightarrow \mathbb{R}^{w \cdot D}$ 를 학습한다. 여기서 w 는 윈도우 길이, D 는 종목 수이다. 학습 목적함수는 실제 분포 $\mathbb{P}$ 와 합성 분포 $\mathbb{Q}_\theta$ 사이의 제곱 최대 평균 불일치이며, 다음과 같이 정의된다.

$$MMD^2(\mathbb{P}, \mathbb{Q}_\theta) = E[k(X, X')] + E[k(Y, Y')] - 2E[k(X, Y)] \quad (1)$$

여기서 X, X' 는 실제 분포 $\mathbb{P}$ 에서 독립 추출된 표본, Y, Y' 는 합성 분포 $\mathbb{Q}_\theta$ 에서 독립 추출된 표본이며, 마지막 항의 X 와 Y 는 각각 $\mathbb{P}$ 와 $\mathbb{Q}_\theta$ 에서 독립 추출된 표본이다. k 는 커널 함수이다.

다중 척도 가우시안 핵 $k(x, y) = \sum_{\sigma \in S} \exp(-\|x - y\|^2 / 2\sigma^2)$ 이 사용되며, 척도 집합 S 는 자료의 쌍별 유클리드 거리의 중앙값 어림법(Median heuristic)을 기준으로 결정된다. 핵의 보편적 성질에 의해 식 (1)을 0으로 만드는 것은 모든 다항 모멘트의 일치와 동가적이며, 양식화 사실의 보존이 학습 목적함수에 직접 부여된다.

생성기는 Linear(64→512) - ReLU - Linear(512→512) - ReLU - Linear(512→512) - ReLU - Linear(512→660)의 4계층 다층 퍼셉트론으로, 잠재 차원 $d_z=64$, 은닉층 폭 $h=512$, 출력 차원 $w \cdot D=60 \cdot 11=660$ 이다. 은닉층에는 ReLU 활성화를 적용하고 출력층은 선형으로 두며, 정규화 없이 가중치는 Kaiming 균등 초기화로 초기화한다.

윈도우 길이 $W=60$ 은 본 연구의 실증적 점검 결과를 반영한 선택이다. 사전 GARCH(1,1) 스트레스

검증에서 $W=24$ 는 윈도우 경계의 시간 의존성 단절이 잦고, $W>120$ 은 가용 회전 윈도우 수 감소로 일반화 성능이 저하되었으며, $W=60$ (약 3개월)에서 실제 자료 대비 약 94%의 단기 자기상관이 보존되는 절충점이 관찰되었다.

3.4.2 변분 자동부호기(VAE)

Kingma and Welling[9]의 변분 자동부호기는 잠재 변수 모형 $p_\theta(x, z) = p_\theta(x|z)p(z)$ 하에서 다음의 증거하한(ELBO)을 최대화한다.

$$L_{ELBO}(\theta, \phi; x) = E_{q_\theta(z|x)}[\log p_\theta(x|z)] - \beta \cdot D_{KL}(q_\theta(z|x) \| p(z)) \quad (2)$$

우변 첫 항은 잠재변수 z 를 고정한 상태에서 자료 x 의 재구성 우도이고, 둘째 항은 인코더 분포 $q_\theta(z|x)$ 와 사전분포 $p(z)$ 사이의 쿨백-라이블러(KL, Kullback-Leibler) 발산이다.

인코더와 디코더는 모두 가우시안 분포를 가정하며, 디코더의 로그 분산은 $[-6, 2]$ 구간으로 클리핑된다. 인코더는 Linear(660→512) - ReLU - Linear(512→512) - ReLU 구조 뒤에 평균 헤드 Linear(512→32)와 로그 분산 헤드 Linear(512→32)로 분기하며, 디코더는 이와 대칭적인 2계층 공유 표현을 거쳐 평균과 로그 분산 두 헤드로 분기한다. 모든 은닉층에 ReLU 활성화 함수를 적용하고 출력 헤드는 선형으로 두며, 배치 정규화나 층 정규화는 사용하지 않고 가중치는 PyTorch 기본값인 Kaiming 균등 초기화로 초기화한다. 학습 안정성을 위해 매 스텝 최대 노름 5.0의 기울기 클리핑을 적용한다. 잠재 차원 $d_z = 32$ 는 GMMN의 절반 수준으로, VAE의 정규화된 사후분포 학습이 보다 압축적인 잠재 표현을 가능하게 한다는 Kingma and Welling[9]의 관찰에 기반한 설정이다. 은닉층 폭, 배치 크기, 에포크는 GMMN과 동일하게 부여하여 학습 예산을 맞추되, 학습률만 안정적 수렴을 위해 GMMN의 두 배인 1×10^{-3} ($\beta_1 = 0.9$, $\beta_2 = 0.999$)으로 설정한다. β 매개변수는 학습 초기 50 에포크에 걸쳐 0에서 1.0으로 선

형 증가시키는 KL 어닐링을 적용하여 사후 분포 붕괴(Posterior collapse)를 방지한다.

3.4.3 정상 부트스트랩(SB)

Politis and Romano 20]의 정상 부트스트랩은 평균 블록 길이를 매개변수로 하는 기하분포 $\geq om(1/\bar{b})$ 로부터 가변 블록 길이를 추출하여 자기상관 구조를 보존한다. 다변량 시계열 $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_T)$ 에 대해, t 번째 합성 표본 $\tilde{\mathbf{r}}_t$ 는 시점 인덱스 I_t 에 따라 다음과 같이 결정된다.

$$\tilde{\mathbf{r}}_t = \mathbf{r}_{I_t}, I_t = \begin{cases} U_t & \text{if } t = 1 \text{ or } B_t = 1 \\ (I_{t-1} + 1) \bmod T & \text{otherwise} \end{cases} \quad (3)$$

여기서 $U_t \sim \text{Uniform}\{0, \dots, T-1\}$ 는 새 블록의 시작점이고, $B_t \sim \text{Bernoulli}(1/\bar{b})$ 는 블록 종료 표지이다. 동일한 인덱스 I_t 가 모든 종목에 적용되므로, 본 SB 구현은 자산 간 상관구조를 표본 단위에서 정확히 보존한다. 기본 매개변수는 $\bar{b} = 10$ 이며, 민감도 분석에서는 $\bar{b} \in \{5, 10, 15, 20\}$ 값을 활용하여 비교한다.

3.5 매칭된 표본 수와 평가 절차

세 방법 모두에 대해 합성 표본 수는 $K = 20$ (시드 2개 $\times$ 시드당 10회 추출)으로 매칭된다. 결국 치는 모든 방법에 대해 K 추출 전반에 걸쳐 단 한 건도 발생하지 않았으므로 사후 선택 편향이 발생하지 않는다.

다중 가중 PBO는 실제 자료와 합성 자료의 통계량을 가중 결합하여 정의된다. 가중치 $w \in [0, 1]$ 에 대해 결합된 IS 통계량은 다음과 같이 산출된다.

$$SS_w^{combined} = w \cdot SS^{real} + (1 - w) \cdot \frac{1}{K} \sum_{k=1}^K SS^{aug, (k)} \quad (4)$$

IS 최적 전략 선정에는 네 가지 점수화 방법이 비교된다. M1은 각 전략이 IS에서 1등(최상위 SR)

을 기록한 횃수를, M2는 IS 등수가 중앙값을 상회한 횃수를 실제·합성 데이터에 걸쳐 가장 결합한 점수이다. M3는 가장 결합된 IS 평균 등수가 가장 높은 전략을 선택하며(단순 최우수 IS 통계량에 해당), M4는 정규화된 IS 등수를 가장 결합한 연속 보르다 점수이다.

표본 수준 변동성은 11개 종목을 모집단으로 한 95% 표본 부트스트랩 신뢰구간(부트스트랩 횃수 10,000)으로 정량화된다. 모든 무작위 결정은 단일 마스터 시드(seed = 20,260,424)로부터 SHA-256 해시 기반의 결정적(Deterministic) 절차로 도출되며, 동일 마스터 시드에 대해 비트 수준 재현이 가능하다. 본 연구의 모든 생성 모형은 PyTorch 2.x로 구현되었으며, 학습은 NVIDIA GeForce RTX 3090 GPU 환경에서 수행되었다.

IV. 실증 분석 결과

본 장은 세 분석 결과를 차례로 보고한다. 제4.1절은 다중 가중 PBO의 주요 결과를, 제4.2절은 합성 자료의 양식화 사실 보존도를, 제4.3절은 SB 블록 길이에 대한 민감도 분석을 제시한다. 모든 결과는 11개 종목 평균에 대한 95% 표본 부트스트랩 신뢰구간을 동반한다.

4.1 다중 가중 PBO 결과

표 2는 M3 점수화 방법 하에서 두 시간 분할 분기, 세 합성 방법론, 가중치 $w \in \{1.0, 0.7, 0.5, 0.3, 0.1\}$ 에 대한 PBO 평균을 보고한다.

표 2. 다중 가중 PBO 결과 (M3 점수, 11개 종목 평균; F = F-augment, W = W-augment)

Table 2. Multi-Weight PBO results (M3 score, mean over 11 tickers; F = F-augment, W = W-augment)

Regime	Method	w=1.0	w=0.7	w=0.5	w=0.3	w=0.1
F	GMMN	0.304	0.181	0.172	0.168	0.179
F	VAE	0.304	0.184	0.176	0.165	0.147
F	SB	0.304	0.196	0.186	0.172	0.160
W	GMMN	0.376	0.244	0.227	0.214	0.206
W	VAE	0.376	0.294	0.283	0.281	0.284
W	SB	0.376	0.291	0.284	0.275	0.281

핵심 결과의 95% 표본 부트스트랩 신뢰구간(부트스트랩 횃수 10,000)은 다음과 같다. F-augment 분기 $w=1.0$ 기준값 0.304는 신뢰구간 [0.260, 0.348]을, W-augment 분기 $w=1.0$ 기준값 0.376은 [0.291, 0.464]를 가진다. W-augment 분기 $w=0.5$ 에서 GMMN의 PBO 0.227은 [0.162, 0.296], VAE의 0.283은 [0.214, 0.345], SB의 0.284는 [0.206, 0.370]의 신뢰구간을 가진다. GMMN 신뢰구간의 상한 0.296이 W-augment 기준값 신뢰구간의 하한 0.291을 거의 정확히 만나는 임계 위치에 있어, 두 분포가 미중첩에 가까운 형태로 분리됨을 의미한다.

분석 결과는 다음 네 핵심 사실을 정량적으로 보고한다.

첫째, 합성 자료를 활용한 모든 방법론에서 $w < 1$ 영역의 PBO가 $w=1.0$ 기준값보다 일관되게 낮게 추정되었다. F-augment 분기의 $w=0.5$ 에서 PBO 하락 폭은 0.118 - 0.132포인트이며, W-augment 분기에서는 0.092 - 0.149포인트이다. 합성 자료의 활용이 사양 탐색 위험을 통계적으로 의미 있게 완화함을 의미한다.

둘째, 시간 분할의 엄격성에 따라 합성 방법론 사이의 상대적 우위가 변화한다. 세 방법론 사이의 PBO 격차(최댓값-최솟값)를 가중치별로 비교하면, F-augment 분기에서는 평균 0.017포인트, 최대 0.032포인트에 머무른다($w=0.7, 0.5, 0.3, 0.1$ 에서 각각 0.015, 0.014, 0.007, 0.032). 반면 표본 외 정보 누출이 차단된 W-augment 분기에서는 평균 0.063포인트, 최대 0.078포인트에 달한다(각각 0.050, 0.057, 0.067, 0.078). 이는 11개 종목 표본 부트스트랩에서 통계적으로 유의한 격차가 있음을 의미한다.

셋째, M1 점수화 방법 하에서 PBO는 모든 가중치에 걸쳐 $w=1.0$ 값과 거의 동일한 수준에서 유지된다(F-augment 0.304, W-augment 0.376). 이는 가중치 w 가 변하더라도 M1이 선택하는 IS 최적 전략이 대체로 유지되는 데 따른 것으로 보인다. 실제 자료와 합성 자료 양쪽에서 일관되게 상위를 차지하는 전략이 가중 결합 이후에도 우위를 유지하는 경향이 있어, 선택되는 IS 최적 전략과 그 OOS 등수 분포가 w 에 크게 민감하지 않은 것으로 해석된다. 개별 전략 점수의 단조적 경향만으로 최적 전략의 불

변이 보장되는 것은 아니므로, 본 결과는 단조성 자체보다 IS 최적 전략의 상대적 안정성과 관련된 것으로 보인다. 반면 M3와 M4의 추정값 차이는 0.001 포인트 이내로 본질적으로 일치하며(표 3), M2의 이산 점프 문제가 본 연구의 결론에 영향을 미치지 않는다.

표 3. M2·M3·M4 점수화 비교 및 M3 - M4 일치성 검증 (w=0.5, 11개 종목 평균; F = F-augment, W = W-augment)

Table 3. Comparison of M2/M3/M4 Scoring and M3 - M4 Consistency Check (w=0.5, Mean over 11 Tickers; F = F-augment, W = W-augment)

Regime	Method	M2	M3	M4	M3-M4
F	GMMN	0.205	0.172	0.172	<0.001
F	VAE	0.166	0.176	0.176	<0.001
F	SB	0.179	0.186	0.186	<0.001
W	GMMN	0.235	0.227	0.227	<0.001
W	VAE	0.290	0.283	0.283	<0.001
W	SB	0.277	0.284	0.284	<0.001

넷째, w가 0.5에서 0.1로 감소함에 따라 PBO가 추가로 하락하는 정도는 모형별로 상이하다. GMMN의 W-augment PBO는 w=0.5의 0.227에서 w=0.1의 0.206으로 하락하나, VAE와 SB는 거의 변화가 없다(0.283 → 0.284, 0.284 → 0.281). GMMN이 합성 자료의 비중이 늘어남에 따라 추가적 정보를 제공하는 반면, VAE와 SB는 일정 수준 이상에서 한계 정보가 포화됨을 시사한다.

그림 1은 두 시간 분할 분기에서의 PBO와 가중

치의 관계를 시각화한다. W-augment 분기에서 GMMN 곡선이 다른 두 방법론과 명확히 분리됨을 확인할 수 있다.

4.2 양식화 사실 보존도

표 4는 세 합성 방법론의 양식화 사실 보존도를 다섯 지표로 비교한다. JSD는 한계 분포 간 Jensen-Shannon 발산이고, KS는 Kolmogorov-Smirnov 통계량이며, kurt는 합성 자료의 첨도(실제 자료 첨도: F-augment 12.36, W-augment 11.05), VaR(5%)는 5% 분위 VaR(실제 자료: 약 -0.022 및 -0.024), ACF(r^2)는 합성 제공수익률의 1 - 20차 표본 자기상관에 대한 절댓값 평균으로 변동성 군집의 보존도를 측정한다.

세 방법론은 한계 분포 일치(JSD, KS) 측면에서 모두 매우 작은 거리를 보고한다(JSD 최댓값 0.017). 그러나 첨도와 변동성 군집 보존에서 SB는 두 신경망 방법보다 명백히 우월하다. 실제 자료 첨도 11.05 - 12.36 대비 SB는 7.91 - 10.67을 보존(약 64% - 86% 수준)하는 반면, GMMN과 VAE는 각각 0.31 - 0.34와 0.62 - 0.79에 머문다. 신경망 생성기의 평균 출력이 정규분포에 가까운 형태로 수렴하면서 두꺼운 꼬리가 위축되는 현상으로 해석되며, Cont [27]이 지적한 금융 시계열의 두꺼운 꼬리 보존이 통상의 신경망 생성모형에서 어렵다는 점과 정합적이다.

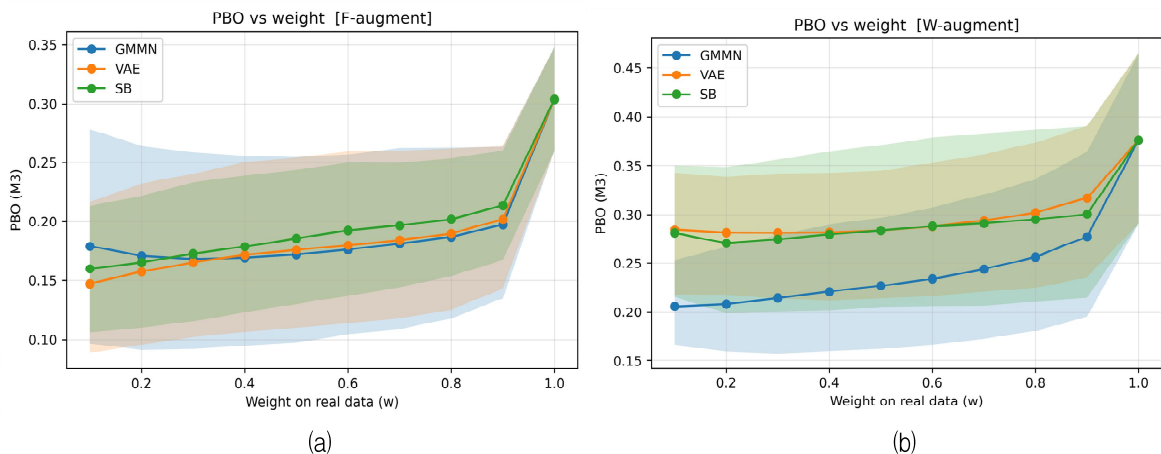


그림 1. 가중치에 따른 PBO 변화

(a) F-augment 분기, (b) W-augment 분기. M3 점수, 11개 종목 평균과 95% 신뢰구간 동반

Fig. 1. PBO as a Function of Weight

(a) F-augment regime, (b) W-augment regime. M3 score, mean over 11 tickers with 95% confidence interval

표 4. 양식화 사실 보존도 비교 (11개 종목 평균; F = F-augment, W = W-augment)

Table 4. Comparison of stylized facts preservation (Mean over 11 tickers; F = F-augment, W = W-augment)

Regime	Method	JSD ↓	KS ↓	kurt	VaR(5%)	ACF(r^2)
F	GMMN	0.011	0.048	0.339	-0.0204	0.063
F	VAE	0.009	0.038	0.794	-0.0195	0.058
F	SB	0.003	0.024	10.668	-0.0221	0.098
W	GMMN	0.017	0.057	0.315	-0.0217	0.065
W	VAE	0.015	0.050	0.622	-0.0213	0.059
W	SB	0.009	0.043	7.905	-0.0254	0.085

표 5. 자산 간 상관구조 보존도 (프로베니우스 거리, 작을수록 우수; 괄호 안은 표본 표준편차)

Table 5. Cross-Asset correlation preservation (Frobenius distance, smaller is better; parentheses are sample standard deviations)

Regime	GMMN	VAE	SB
F-augment	6.76 (0.07)	6.60 (0.07)	0.53 (0.33)
W-augment	7.12 (0.08)	6.84 (0.11)	0.58 (0.38)

자산 간 상관구조 보존도는 표 5에 보고된다. 실제 상관행렬과 합성 상관행렬의 프로베니우스 (Frobenius) 거리는 SB가 0.53 - 0.58로 가장 작고, VAE(6.60 - 6.84)와 GMMN(6.76 - 7.12)이 한 자리 수 이상 큰 격차를 보인다. SB는 동일한 시점 인덱스를 모든 종목에 적용하는 구조에서 자산 간 동시적 상관구조를 본질적으로 보존하므로 이 결과는 예측 가능하다. 신경망 모형의 상대적 약점은 다변량 확

률 동학에서 한계 분포에 비해 결합 분포 학습이 어렵다는 잘 알려진 특성을 반영한다[7].

핵심 발견은 양식화 사실 보존이 가장 우수한 SB가 PBO 완화 효과에서는 GMMN에 미치지 못한 다는 비단조적 결과이다. 합성 자료의 평가 기준이 단순한 분포 일치 지표만으로 환원될 수 없음을 의미하며, 자세한 해석은 제5장에서 논의된다.

4.3 민감도 분석

표 6은 SB 블록 길이 $\bar{b} \in \{5, 10, 15, 20\}$ 에 대한 PBO 변화를 보고한다($w=0.5$, M3 점수 기준).

표 6. SB 블록 길이 민감도 ($w=0.5$, M3, 11개 종목 평균)

Table 6. Sensitivity to SB Block Length ($w=0.5$, M3, mean over 11 tickers)

Regime	$\bar{b}=5$	$\bar{b}=10$	$\bar{b}=15$	$\bar{b}=20$
F-augment	0.156	0.231	0.204	0.109
W-augment	0.403	0.311	0.380	0.255

각 셀의 95% 표본 부트스트랩 신뢰구간은 다음과 같다. F-augment 분기에서 $\bar{b}=5$ 는 [0.100, 0.221], $\bar{b}=10$ 은 [0.155, 0.312], $\bar{b}=15$ 는 [0.156, 0.254], $\bar{b}=20$ 은 [0.059, 0.169]이며, W-augment 분기에서 $\bar{b}=5$ 는 [0.333, 0.473], $\bar{b}=10$ 은 [0.254, 0.371], $\bar{b}=15$ 는 [0.316, 0.444], $\bar{b}=20$ 은 [0.185, 0.330]이다.

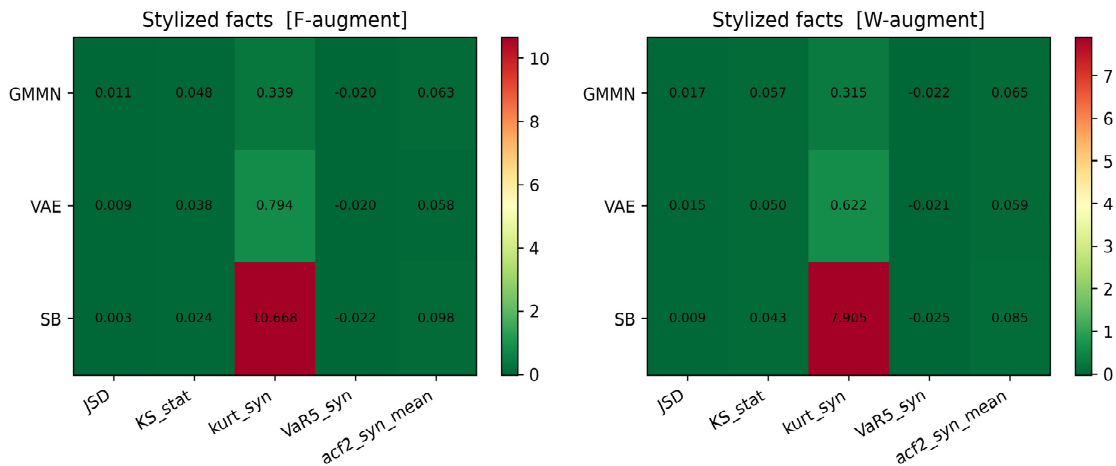


그림 2. 양식화 사실 보존도 히트맵 (좌: F-augment 분기, 우: W-augment 분기).

Fig. 2. Heatmap of stylized facts preservation (Left: F-augment regime, Right: W-augment regime).

SB 블록 길이에 따른 PBO 변동 폭은 F-augment 0.122포인트, W-augment 0.148포인트로 결코 무시할 수 없는 수준이며, 단순한 단조 패턴을 보이지 않는다. SB의 결과가 블록 길이 선택에 민감함을 의미하며, 본 연구의 기본 매개변수($\bar{b}=10$)는 Politis and Romano[20] 및 Politis and White[21]의 권고에 따른 일관된 선택이다.

V. 논 의

5.1 양식화 사실과 PBO 완화의 상충관계

본 연구에서 흥미로운 점은 양식화 사실 보존이 가장 우수한 방법론이 PBO 완화 효과에서는 가장 우수한 방법론이 아닐 수 있다는 사실이다. 구체적으로, SB는 첨도와 자산 간 상관구조 보존에서 두 신경망 방법론보다 한 자리 수 이상 우월하나, W-augment 분기의 PBO 완화 효과에서는 GMMN보다 0.057포인트 열위에 위치한다.

이러한 비단조성은 다음과 같이 해석된다. SB는 실제 자료의 시점 단위 재배열을 통해 합성 자료를 생성하므로 표본의 정보적 다양성이 본질적으로 제한된다. 즉 SB의 합성 표본은 원자료의 동일한 정보 핵을 반복적으로 재구성한 것에 가깝다. 반면 GMMN은 MMD 최소화를 통해 모멘트 일치 추구를 하면서도 잠재변수의 표현력을 통해 원자료에 직접 관측되지 않은 결합 패턴을 생성한다. 사양 탐색 위험을 평가하는 PBO 관점에서, 이러한 새로운 결합 패턴의 다양성은 양식화 사실의 충실한 재구성보다 더 큰 가치를 가질 수 있다.

이러한 비단조적 양상의 정량적 근거는 다음과 같다. §4.2에서 보고된 자산 간 상관구조 프로베니우스 거리는 SB가 0.53 - 0.58, GMMN이 6.76 - 7.12로 한 자리 수 이상의 격차를 보이는데, 이는 SB가 원자료의 시점 단위 재배열을 통해 자산 간 동시적 상관구조를 본질적으로 보존하는 반면 GMMN의 다변량 학습은 한계 분포 일치(JSD 0.011 - 0.017)에 비해 결합 분포 학습이 어렵다는 잘 알려진 특성을 반영한다[7]. 그럼에도 PBO 완화에서 GMMN이 SB를 앞서는 결과는 다음의 정보적 가설로 정합적으

로 해석된다. SB의 합성 표본은 시점 인덱스 재배열에 한정되므로 $K=20$ 추출 전체의 결합 분포가 원자료의 정보 핵에 본질적으로 묶여 있는 반면, GMMN은 잠재변수의 표현력을 통해 원자료에 직접 관측되지 않은 결합 패턴을 일정 범위 내에서 생성한다. 사양 탐색 위험을 평가하는 PBO 관점에서, 이러한 새로운 결합 패턴의 다양성은 평가 자료의 정보적 풍부함을 높여 사양 탐색이 야기하는 우연한 IS-OOS 상관성을 약화시키는 데 기여할 수 있다.

다만 본 해석은 결합 분포 커버리지를 직접 정량화한 결과가 아니라 양식화 사실 보존도와 PBO 완화도의 비단조적 관계에서 도출된 간접적 추론이다. 잠재 공간 표현 분석, 결합 분포 커버리지의 정보 이론적 측정, GMMN과 SB가 생성하는 합성 표본의 결합 모멘트 비교 등의 직접적 정량 분석은 본 연구의 가용 자원 제약상 후속 연구의 과제로 남는다.

이 해석은 López de Prado[2]의 사양 탐색 위험관리 관점과 정합적이다. 동일한 자료에 다수의 사양을 적용하는 과정의 핵심 위험은 사양 탐색이 우연한 표본 외 우위를 만들어내는 것이며, 이를 완화하기 위해서는 평가 자료가 원자료와 통계적으로 충분히 가까우면서도 원자료에 없던 다양성을 보유해야 한다. SB는 전자에 강하지만 후자에서 제한적이며, GMMN은 두 측면 사이의 균형이 다르다. 이는 White[14]의 현실성 검증 절차가 단일 부트스트랩 분포에 의존하는 데서 오는 한계를 다중 패러다임 비교로 보완한 본 연구의 경험적 시사점이다.

5.2 시간 분할의 결정적 중요성

F-augment와 W-augment 분기의 비교는 합성 자료의 평가 환경 설정이 분석의 결론에 상당한 영향을 미침을 보여준다. F-augment 분기에서는 세 방법론 사이의 PBO 격차가 0.04포인트 이내로 좁아 어떤 방법을 선택하더라도 결과의 차이가 무의미하게 보일 수 있다. 그러나 W-augment 분기에서는 GMMN과 다른 두 방법론 사이의 격차가 통계적으로 의미 있는 수준으로 확장된다.

IS-OOS 분리 문제는 단순한 형식적 우려가 아니라 실증적 결론에 영향을 미칠 수 있는 변수이다.

F-augment 분기에서 합성 모형이 평가 구간을 사전에 관찰한 비현실적 정보 누출은 신경망 모형의 우위를 가리는 효과를 가지며, SB와 같은 비모수 모형에 상대적으로 유리하게 작용한다. W-augment의 엄격한 분할은 이러한 누출을 차단하여 모형 본연의 일반화 능력을 노출시킨다. Andrews[28]이 제시한 표본 외 안정성 검정의 정신이 본 연구의 W-augment 설계와 본질적으로 동일한 관점에 서 있다.

5.3 선행연구와의 비교

본 연구의 결과는 다음과 같이 선행연구와 대조된다. Wiese et al.[4]과 Lin et al.[6]은 적대적 학습 기반 모형이 양식화 사실을 부분적으로 보존함을 보고하였으나, 본 연구는 비적대적 모형(GMMN, VAE)이 동일한 평가 기준 하에서 한계 분포 일치(JSD, KS)에서는 SB와 비슷한 수준이지만 첨도와 변동성 군집 보존에서는 명확한 열위를 보임을 밝힌다. 비적대적 학습의 안정성이 시계열 충실성과 자동적으로 양립하지는 않음을 의미한다.

PBO 완화 측면에서 본 연구는 Bailey et al.[3]의 CSCV가 합성 자료 증강과 결합되었을 때 어떤 변화를 겪는지를 체계적으로 살펴본 시도 중 하나이다. López de Prado and Bailey[16]이 제안한 디플레이션된 샤프 비율이 다중 검정의 사후 보정에 머물렀다면, 본 연구는 평가 자료 자체의 사전적 확장을 통해 사양 탐색 위험을 완화하는 새로운 경로를 제시한다.

5.4 한계 및 후속 연구

본 연구의 한계는 다음과 같다.

첫째, 신경망 모형(GMMN, VAE)의 윈도우 단위 학습은 윈도우 경계에서의 시간 의존성 단절을 야기한다. 본 연구는 윈도우 길이 $W=60$ (약 3개월)에서 단기 자기상관과 일반화 성능 사이의 균형을 채택하였으나, 보다 긴 시간 의존성을 보존하는 합성 모형의 개발은 후속 연구의 과제이다.

둘째, VAE의 가우시안 디코더는 조건부 이분산성을 직접 학습하지 않으므로 변동성 군집 보존이

GMMN보다 약하다. 디코더에 GARCH (Engle[29], Bollerslev[30]) 형태의 자기회귀 조건부 분산 구조를 도입한 변형 모형의 평가가 필요하다.

셋째, 본 연구는 시드 수준 변동성 평가를 11개 종목에 대한 표본 부트스트랩으로 대체하였다. 이는 통계적 검정력 측면에서 충분하나, 시드별 학습 결과의 분산을 직접 정량화하는 것은 가용 전산 자원이 확장된 후속 연구의 과제이다.

넷째, 분석 대상이 미국 GICS 11개 산업 ETF에 한정되어 있다. 다른 자산군(개별 종목, 채권, 외환, 원자재)으로의 일반화 가능성은 별도 연구를 통해 평가되어야 한다.

VI. 결 론

본 연구는 적대적 학습을 사용하지 않는 두 다변량 생성모형(GMMN, VAE)과 비모수 대표본추출 기준선(정상 부트스트랩)을 동일한 표본 수와 동일한 평가 윈도우 위에서 비교하여, 합성 시계열 자료를 활용한 증강이 Bailey et al.[3]의 백테스팅 과적합 확률(PBO)을 어느 정도 완화하는지를 실증적으로 규명하였다. 미국 GICS 11개 산업 ETF의 2020 - 2025년 일별 자료를 분석한 결과, W-augment 분기에서 GMMN의 PBO 평균은 실제 자료만 사용한 기준값 대비 최대 약 45%(가중치 $w=0.1$ 기준) 하락하였으며, 이는 동일 조건의 VAE 및 SB보다 통계적으로 유의하게 낮은 개선이다. 비적대적 학습 패러다임 가운데 특히 GMMN의 채택이 학습 안정성과 모멘트 일치의 동시 확보를 통해 합성 자료의 운영적 가치를 높였음을 시사한다.

본 연구의 핵심 학술적 기여는 양식화 사실 보존과 PBO 완화 효과 사이의 상충관계를 정량화한 것이다. 첨도와 자산 간 상관구조 보존에서 우월한 SB가 PBO 완화 효과에서는 GMMN보다 열위에 위치한다는 비단조적 결과는, 합성 자료의 평가 기준이 단일 분포 일치 지표로 환원될 수 없음을 의미한다. 또한 시간 분할의 엄격성에 따라 합성 방법론 사이의 상대적 우위가 변화한다는 결과는, IS-OOS 분리가 단순한 형식적 절차가 아니라 실증적 결론에 영향을 미칠 수 있는 변수임을 보여준다.

자산운용 실무의 관점에서, 본 연구는 합성 자료 증강이 사양 탐색 위험관리의 도구로 기능할 수 있음을 실증적으로 확인하며, 모형 선택이 평가 목적과 시간 분할 분기에 의존해야 함을 제시한다. 후속 연구는 보다 긴 시간 의존성을 보존하는 합성 모형의 개발, 다양한 자산군으로의 일반화 가능성 검증, 시드 수준 변동성의 직접 평가 등을 통해 본 연구의 한계를 보완할 수 있을 것이다.

References

- [1] C. R. Harvey and Y. Liu, "Evaluating trading strategies", *Journal of Portfolio Management*, Vol. 40, No. 5, pp. 108-118, Sep. 2014. <https://doi.org/10.3905/jpm.2014.40.5.108>.
- [2] M. López de Prado, *Advances in Financial Machine Learning*, Wiley, 2018.
- [3] D. H. Bailey, J. M. Borwein, M. L. de Prado, and Q. J. Zhu, "The probability of backtest overfitting", *Journal of Computational Finance*, Vol. 20, No. 4, pp. 39-69, Apr. 2017. <https://doi.org/10.21314/JCF.2016.322>.
- [4] M. Wiese, R. Knobloch, R. Korn, and P. Kretschmer, "Quant GANs: Deep generation of financial time series", *Quantitative Finance*, Vol. 20, No. 9, pp. 1419-1440, Sep. 2020. <https://doi.org/10.1080/14697688.2020.1730426>.
- [5] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks", *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver, BC, Canada, Vol. 32, pp. 5508-5518, Dec. 2019.
- [6] Z. Lin, A. Jain, C. Wang, G. Fanti, and V. Sekar, "Using GANs for sharing networked time series data: Challenges, initial promise, and open questions", *Proc. ACM Internet Measurement Conference (IMC)*, Online, pp. 464-483, Oct. 2020. <https://doi.org/10.1145/3419394.3423643>.
- [7] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs", *Proc. Advances in Neural Information Processing Systems (NIPS)*, Barcelona, Spain, Vol. 29, pp. 2226-2234, Dec. 2016. <https://doi.org/10.48550/arXiv.1606.03498>.
- [8] Y. Li, K. Swersky, and R. Zemel, "Generative moment matching networks", *Proc. International Conference on Machine Learning (ICML)*, Lille, France, pp. 1718-1727, Jul. 2015.
- [9] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes", *Proc. International Conference on Learning Representations (ICLR)*, Banff, Canada, pp. 1-14, Apr. 2014. <https://doi.org/10.48550/arXiv.1312.6114>.
- [10] C. R. Harvey, Y. Liu, and H. Zhu, "... and the cross-section of expected returns", *Review of Financial Studies*, Vol. 29, No. 1, pp. 5-68, Jan. 2016. <https://doi.org/10.1093/rfs/hhv059>.
- [11] D. H. Bailey, J. M. Borwein, M. L. de Prado, and Q. J. Zhu, "Pseudo-mathematics and financial charlatanism: The effects of backtest overfitting on out-of-sample performance", *Notices of the American Mathematical Society*, Vol. 61, No. 5, pp. 458-471, May 2014. <https://doi.org/10.1090/noti1105>.
- [12] C. R. Harvey and Y. Liu, "Detecting repeatable performance", *Review of Financial Studies*, Vol. 31, No. 7, pp. 2499-2552, Jul. 2018. <https://doi.org/10.1093/rfs/hhy014>.
- [13] J. P. Romano and M. Wolf, "Stepwise multiple testing as formalized data snooping", *Econometrica*, Vol. 73, No. 4, pp. 1237-1282, Jul. 2005. <https://doi.org/10.1111/j.1468-0262.2005.00615.x>.
- [14] H. White, "A reality check for data snooping", *Econometrica*, Vol. 68, No. 5, pp. 1097-1126, Sep. 2000. <https://doi.org/10.1111/1468-0262.00152>.
- [15] R. Sullivan, A. Timmermann, and H. White, "Data-snooping, technical trading rule performance, and the bootstrap", *Journal of Finance*, Vol. 54, No. 5, pp. 1647-1691, Oct. 1999. <https://doi.org/10.1111/0022-1082.00163>.

- [16] M. L. de Prado and D. H. Bailey, "The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting, and non-normality", *Journal of Portfolio Management*, Vol. 40, No. 5, pp. 94-107, Sep. 2014. <https://doi.org/10.3905/jpm.2014.40.5.094>.
- [17] C.-L. Li, W.-C. Chang, Y. Cheng, Y. Yang, and B. Póczos, "MMD GAN: Towards deeper understanding of moment matching network", *Proc. Advances in Neural Information Processing Systems (NIPS)*, Long Beach, California, USA, Vol. 30, pp. 2203-2213, Dec. 2017.
- [18] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional GANs", *arXiv preprint, arXiv:1706.02633*, Jun. 2017. <https://doi.org/10.48550/arXiv.1706.02633>.
- [19] A. Koshiyama, N. Firoozye, and P. Treleaven, "Generative adversarial networks for financial trading strategies fine-tuning and combination", *Quantitative Finance*, Vol. 21, No. 5, pp. 797-813, May 2021. <https://doi.org/10.1080/14697688.2020.1790635>.
- [20] D. N. Politis and J. P. Romano, "The stationary bootstrap", *Journal of the American Statistical Association*, Vol. 89, No. 428, pp. 1303-1313, Dec. 1994. <https://doi.org/10.1080/01621459.1994.10476870>.
- [21] D. N. Politis and H. White, "Automatic block-length selection for the dependent bootstrap", *Econometric Reviews*, Vol. 23, No. 1, pp. 53-70, Dec. 2004. <https://doi.org/10.1081/ETC-120028836>.
- [22] A. Patton, D. N. Politis, and H. White, "Correction to 'Automatic block-length selection for the dependent bootstrap' by D. Politis and H. White", *Econometric Reviews*, Vol. 28, No. 4, pp. 372-375, Jul. 2009. <https://doi.org/10.1080/07474930802459016>.
- [23] D. N. Politis, "Bootstrap methods in time series analysis", *Handbook of Computational Statistics*, Springer, pp. 1071-1111, 2012.
- [24] B. Efron, "Bootstrap methods: Another look at the jackknife", *Annals of Statistics*, Vol. 7, No. 1, pp. 1-26, Jan. 1979. <https://doi.org/10.1214/aos/1176344552>.
- [25] G. K. Dziugaite, D. M. Roy, and Z. Ghahramani, "Training generative neural networks via maximum mean discrepancy optimization", *Proc. Uncertainty in Artificial Intelligence (UAI)*, Amsterdam, Netherlands, pp. 258-267, Jul. 2015. <https://doi.org/10.48550/arXiv.1505.03906>.
- [26] H. Bühler, L. Gonon, J. Teichmann, and B. Wood, "Deep hedging", *Quantitative Finance*, Vol. 19, No. 8, pp. 1271-1291, Aug. 2019. <https://doi.org/10.1080/14697688.2019.1571683>.
- [27] R. Cont, "Empirical properties of asset returns: Stylized facts and statistical issues", *Quantitative Finance*, Vol. 1, No. 2, pp. 223-236, Feb. 2001. <https://doi.org/10.1080/713665670>.
- [28] D. W. K. Andrews, "End-of-sample instability tests", *Econometrica*, Vol. 71, No. 6, pp. 1661-1694, Nov. 2003. <https://doi.org/10.1111/1468-0262.00466>.
- [29] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, Vol. 50, No. 4, pp. 987-1007, Jul. 1982. <https://doi.org/10.2307/1912773>.
- [30] T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics*, Vol. 31, No. 3, pp. 307-327, Apr. 1986. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1).

저자소개

이 유 섭 (Yuseop Lee)



2023년 3월 ~ 현재 :
한국과학기술원
산업및시스템공학과(전산학
복수전공) 학사과정
관심분야 : 데이터 기반 의사결정,
금융공학

최 인 수 (Insu Choi)



2025년 2월: 한국과학기술원
산업및시스템공학과 (공학박사)
2026년 3월 ~ 현재 : 가천대학교
경영대학 금융·빅데이터학부
조교수
관심분야 : 금융 데이터사이언스,
비즈니스 애널리틱스, 데이터
기반 의사결정