

# A MICE-Doubly Robust Causal Inference Pipeline for High-Dimensional Observational Data: Analyzing Productivity Effects of Digital Transformation in Korean Manufacturing

Lee Seog Min<sup>†</sup>

## ABSTRACT

Causal inference in high-dimensional observational data requires addressing the dual challenges of missing data and model misspecification. We developed a pipeline that systematically integrates Multiple Imputation by Chained Equations (MICE) with the doubly robust Augmented Inverse Probability Weighting (AIPW) estimator to evaluate the productivity effects of digital transformation using data from the Korean Business Activity Survey ( $n=31,572$ ,  $p=281$ , 2019–2023). By employing a MICE strategy that excludes the treatment variable, we increased the sample size by 76% (from 17,897 to 31,572), while  $K=5$  cross-validation in AIPW estimation mitigated the risk of model misspecification. Our findings indicate that digital transformation yields an average total factor productivity (TFP) increase of 3.9%. This effect was particularly pronounced in the pre-pandemic period (+5.3%) and in the electronics (+8.6%) and chemical (+13.5%) industries. Robustness was confirmed through additional tests, including placebo tests and subsample analyses. Compared to Targeted Maximum Likelihood Estimation (TMLE), AIPW provided nearly identical estimates with a 27% faster computation time.

**Keywords :** Causal Inference, Multiple Imputation, Doubly Robust Estimation, High-Dimensional Data, Digital Transformation

## 고차원 관측 데이터에서 MICE-Doubly Robust 인과추론 파이프라인: 한국 제조업 디지털 전환 효과 분석

이 석 민<sup>†</sup>

## 요 약

고차원 관측 데이터에서의 인과추론은 결측 데이터와 모형 오명세(model misspecification)라는 이중의 문제를 해결해야 한다. 우리는 MICE와 이중강건 추정법인 AIPW를 체계적으로 결합한 파이프라인을 개발하여, 한국 제조업 기업활동조사 데이터( $n=31,572$ ,  $p=281$ , 2019–2023)에서 디지털 전환의 생산성 효과를 추정했다. 처치 변수 제외 MICE 전략을 통해 표본 크기를 76% 증가시켰고(17,897→31,572),  $K=5$  교차검증을 통한 AIPW 추정으로 모형 오명세 위험을 완화했다. 분석 결과 디지털 전환은 평균 3.9%의 총요소생산성 증가 효과를 보였다. 이 효과는 팬데믹 이전 시기(+5.3%)와 전자부품(+8.6%), 화학제품(+13.5%) 산업에서 특히 두드러졌다. 강건성 검사(Placebo test, 하위표본 분석)를 통해 결과의 안정성을 확인하였다. AIPW는 TMLE와 거의 동일한 추정치를 제공하면서 27% 빠른 계산 속도를 달성했다.

**키워드 :** 인과추론, 다중대체, 이중강건 추정, 고차원 데이터, 디지털 전환

## 1. 서 론

### 1.1 연구 배경 및 기어

관찰연구 데이터를 활용한 인과추론은 경제학, 보건학 등 다양한 분야에서 핵심적인 역할을 수행해왔다[1]. 하지만 연

구자들은 두 가지 근본적인 문제에 직면한다. 첫째, 처치배정이 무작위가 아니므로 선택편향이 발생한다. 둘째, 설문조사나 공공데이터는 상당한 비율의 결측값을 포함한다[2, 3]. 이러한 문제들을 해결하기 위해 성향점수 방법, 매칭, 이중강건 추정 등이 발전해왔고[4, 5], 결측데이터 처리를 위해 MICE(Multiple Imputation by Chained Equations)가 표준적 해법으로 자리잡았다[6, 7].

전통적 접근법은 이러한 문제를 개별적으로 해결하려 시도하였지만 통합된 솔루션의 부재로 인해 추정 효율성과 강

※ 이 논문은 한신대학교 학술연구비(2025년)에 의하여 연구되었음.

† 정 회 원: 한신대학교 공공인재빅데이터융합 교수

논문접수: 2025년 11월 26일

심사완료: 2025년 12월 18일

\* Corresponding Author : Lee Seog Min(newmind68@hs.ac.kr)

건성이 제한되었다. 특히 제조업 패널 데이터는 (1) 고차원성( $p=281$  공변량), (2) 높은 결측률(43%), (3) 모형 불확실성, (4) 선택편의라는 복합적 문제를 발생시킨다. 본 연구는 다중 대체와 이중강건 추정을 결합한 엔드투엔드(end to end) 파이프라인을 제안하여 이러한 한계를 극복하고자 한다. 우리는 세 가지 설계 원칙을 준수했다: (1) 모듈성(독립적 3계층 아키텍처), (2) 확장성(계산 복잡도  $O(Kp^{2n})$  여기서  $K=5, p=281, n=31,572$ ), (3) 강건성(이중강건 추정과 교차검증).

본 연구의 핵심 기여는 다음과 같다. 첫째, MICE와 이중강건 추정의 체계적 결합 방법론을 제시했다. 처치 변수 제와 MICE 전략으로 추정 편의를 제거하면서도 표본 크기를 76% 증가시켰다(17,897→31,572). 둘째, AIPW(Augmented Inverse Probability Weighting)와 TMLE(Targeted Maximum Likelihood Estimation) 비교 분석을 통해 AIPW가 TMLE와 거의 동일한 추정치를 제공하면서 27% 빠른 실행 속도를 달성함을 규명했다. 셋째, CSR(Compressed Sparse Row) 희소 행렬로 96% 메모리 절감(71GB→2.8GB)을 실현하여 대규모 데이터 처리를 가능케 했다. 넷째, 한국 제조업 디지털 전환 데이터에 성공적으로 적용하여 산업별·시기별 이질적 효과를 강건하게 식별했다.

## 1.2 한국 제조업 디지털 전환 데이터

우리의 실증 응용은 한국 제조업의 디지털 전환이 생산성에 미치는 영향을 분석한다. 통계청 기업활동조사 2019-2023년 자료를 활용했고, 총 31,572개의 기업-연도 관측치로 구성되었다. 처치변수는 사물인터넷, 클라우드 컴퓨팅, 인공지능, 빅데이터 분석 등 네 가지 디지털 기술의 도입 여부를 통합한 지표다. 여기서 인공지능(AI)은 예측 및 자동화를, 빅데이터는 대용량 데이터 처리 및 패턴 발견을 목적으로 한다는 점에서 기술적 성격이 구별된다. 결과변수는 Levinsohn-Petrin 방법[8]으로 추정한 총요소생산성(TFP: Total Factor Productivity)을 사용했다. 이는 투입요소의 내생성을 통제한 추정치다. 통제변수는 281개로서, 재무비율(부채비율, ROA: Return on Assets), 인력 구성(R&D 인력 비중), 산업터미(2자리 KSIC: Korean Standard Industrial Classification), 지역터미, 수출 활동 지표 등을 포괄했다.

이 데이터는 본 연구의 방법론을 검증하기에 이상적인 조건을 갖고 있다. 첫째로 처치 불균형이 존재한다. 디지털 전환 도입률이 8.5%에 불과해 대조군과 처치군의 비율이 10.7:1에 달한다. 이는 한국 제조업의 디지털 전환 초기 단계를 반영한다. 둘째로 높은 결측률이 존재한다. 원자료에서 43%의 기업이 하나 이상의 공변량에서 결측값을 갖는다. 완전사례분석을 적용할 경우 표본의 43%를 상실하게 된다. 셋째로 고차원성 문제가 있다. 281개 공변량이 잠재적 교란변수로 작용한다. 이는 전통적 선형 모형의 한계를 초과한다[10].

## 2. 방법론

### 2.1 이중강건(Doubly Robust) 추정의 원리

이중강건 추정은 성향점수 모델과 결과 회귀 모델을 결합하여, 둘 중 하나만 올바르게 지정되어도 불편추정량을 제공하는 강력한 방법론이다[11]. 본 연구에서는 AIPW를 사용했다. 계산의 효율성을 위해 가중치  $w_{1i} = T_i/\hat{e}(X_i)$ 와  $w_{0i} = (1-T_i)/(1-\hat{e}(X_i))$ 를 정의하고, 각 관측치의 영향함수를 다음과 같이 정의한다:

$$\psi_i = \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) + w_{1i}(Y_i - \hat{\mu}_1(X_i)) - w_{0i}(Y_i - \hat{\mu}_0(X_i))$$

AIPW 추정량은 영향함수의 표본평균이다:

$$\hat{\tau}_{AIPW} = \frac{1}{n} \sum_{i=1}^n \psi_i \quad (1)$$

여기서  $T_i$ 는 개인  $i$ 의 처치 지시자,  $Y_i$ 는 관찰된 결과변수,  $\hat{e}(X)$ 는 추정된 성향점수,  $\hat{\mu}_1(X)$ 와  $\hat{\mu}_0(X)$ 는 각각 처치군과 대조군의 결과 회귀 함수다. 이 추정량의 핵심 특성은 이중강건성이다. 이는 성향점수 모델 또는 결과 회귀 모델 중 하나가 일치추정량을 제공하면 AIPW도 일치추정량이 된다[12].

#### 2.1.1 인과추론의 기본 가정

AIPW 추정량이 평균처치효과(Average Treatment Effect: ATE)를 일치추정하기 위해서는 다음 세 가지 가정이 충족되어야 한다.

가정 1: SUTVA (Stable Unit Treatment Value Assumption)

개인  $i$ 의 잠재적 결과는 다른 개인의 처치 배정에 영향받지 않는다(no interference). 또한 처치의 여러 버전이 존재하지 않는다(no hidden variations). 본 연구에서 디지털 전환은 기업 수준에서 독립적으로 결정되므로, 기업 간 spillover 효과가 제한적이라면 SUTVA가 성립한다. 다만 동일 산업 내 경쟁 기업의 디지털 전환이 다른 기업에 영향(예: 경쟁 압력)을 미칠 가능성은 본 연구의 한계로 인정한다.

가정 2: Unconfoundedness (Conditional Independence)

$$(Y(1), Y(0)) \perp T \mid X \quad (2)$$

관찰된 공변량  $x$ 를 조건부로 할 때, 처치 배정이 잠재적 결과와 독립이다. 본 연구는 281개 공변량을 통제하여 이

가정을 만족시키고자 했다. 하지만 관찰되지 않은 요인(경영진 역량, 조직 문화 등)이 여전히 존재할 수 있으며, 이는 본 연구의 한계로 인정한다.

### 가정 3: Positivity (Overlap)

$$0 < P(T=1|X=x) < 1 \quad \forall x \quad (3)$$

모든 공변량 패턴  $x$ 에 대해 처치를 받을 확률이 0과 1 사이에 있어야 한다. 본 연구 데이터는 10.7 : 1의 처치 불균형으로 인해 일부 영역에서 이 가정이 위배될 위험이 있다. Section 3.3에서 극단적 성향점수( $< 0.01$  또는  $> 0.99$ )를 가진 관측치가 0.22%(71개)에 불과함을 확인했다. 극단값 제거 분석(trimming analysis) 결과 추정치가 안정적임을 검증했다.

#### 2.1.2 Unconfoundedness 가정의 타당성 논증

Unconfoundedness 가정이 성립하려면, 디지털 전환 결정에 영향을 미치는 모든 교란변수가 관찰된 281개 공변량에 포함되어야 한다. 이 가정은 직접 검증할 수 없다(Manski, 2007). 하지만 디지털 전환 결정 요인에 관한 기존 문헌을 검토하여 본 연구의 공변량 집합이 주요 교란변수를 충분히 포괄하는지 평가할 수 있다.

**디지털 전환 결정 요인 문헌:** 정보시스템 및 기술경영 문헌은 기업의 디지털 기술 도입 결정이 (1) 재무적 역량, (2) 기술적 준비도, (3) 시장 압력, (4) 조직 특성에 의해 결정된다고 보고한다[23, 24, 25]. 또한 (5) 산업 특성과 (6) 지역 요인도 중요한 영향을 미친다[26, 27].

**본 연구 공변량과의 매핑:** 281개 공변량은 위 요인들을 다음과 같이 포괄한다.

- 재무적 역량 (34개 변수): 부채비율, ROA, ROE, 유동비율, 당좌비율, 현금흐름, 영업이익률, 자기자본비율 등 상세한 재무지표
- 기술적 준비도(18개 변수): R&D 지출, R&D 인력 비중, 특허 출원 건수, 기술 인력 비율, 연구개발비/매출액
- 시장 압력(12개 변수): 수출액, 수출 비중, 시장점유율, 주요 거래처 집중도, 해외 자회사 유무
- 조직 특성(8개 변수): 종업원 수, 기업 연령, 대출 이상 비율, 정규직 비율, 평균 근속연수
- 산업 특성(187개 변수): KSIC 2자리 산업 더미(제조업 세분류), 산업별 기술 집약도
- 지역 요인(17개 변수): 시도 지역 더미, 수도권 여부, 산업단지 입주 여부
- 기타 통제변수(5개 변수): 연도 더미(거시경제 추세 통제)

**관찰되지 않은 잠재적 교란변수:** 문헌에서 제시하는 주요 요인이 대부분 포함되었다. 하지만 다음 변수는 여전히

관찰되지 않는다.

- 경영진 특성: CEO 학력, 경영진 IT 친화도, 디지털 비전
- 조직 문화: 혁신 지향성, 위험 감수성, 변화 수용도
- 네트워크 효과: 공급망 파트너의 디지털화 수준, 산업 내 동종기업 영향

이러한 미관찰 요인이 존재할 가능성을 인정한다. 우리는 두 가지 접근으로 이를 완화하고자 한다.

1. 고차원 통제: 281개 공변량은 디지털 전환 문헌에서 제시하는 대부분의 관찰 가능한 결정 요인을 포괄한다. 특히 R&D 지출과 인력은 경영진의 기술 지향성을 부분적으로 대리할 수 있다. 과거 재무 성과는 경영 역량을 간접 반영한다.

2. Placebo 검정: 무작위 처치 배정 시뮬레이션에서 효과가 0에 가까웠다(Section 3.4). 이는 현재 추정치가 관찰된 공변량 통제를 통해 인과관계를 포착함을 시사한다.

## 2.2 MICE의 원리와 구현

MICE는 결측 패턴에 따라 각 결측 변수를 다른 변수들의 조건부 분포로 순차적으로 대체하는 방법이다.  $m$ 번의 독립적인 대체를 통해  $m$ 개의 완전한 데이터셋을 생성하고, 각각에서 분석을 수행한 후 결과를 통합한다[15]. 우리는 scikit-learn의 IterativeImputer를 사용하여 각 변수를 다른 모든 변수의 조건부 분포로 대체하고 최대 5회 반복했다[15]. 대체 모델로는 랜덤 포레스트 회귀를 사용하여 변수 간 비선형 관계와 상호작용을 포착하였다.

MICE 이론에서 핵심적인 조건은 MAR(Missing At Random) 가정이다. 이는 결측 확률이 관찰된 변수에만 의존한다는 가정이다:

$$P(Missing | Y_{obs}, Y_{miss}) = P(Missing | Y_{obs}) \quad (4)$$

여기서  $Y_{obs}$ 는 관찰된 변수,  $Y_{miss}$ 는 결측 변수다. MAR 하에서 MICE는 일치추정량을 제공한다[16, 17]. 우리는 MAR 가정을 검증하기 위해 결측 패턴 분석을 수행했다. 결측이 기업 특성(규모, 산업, 지역)과 체계적으로 연관되지 않음을 확인했다.

### 2.2.1 처치변수 제외 전략

인과추론에서 MICE 사용 시 중요한 설계 결정은 처치변수를 대체 모델에 포함할지 여부이다. 전통적으로 MICE는 모든 변수를 포함하여 변수 간 관계를 보존하지만 인과추론 맥락에서는 다르다. 처치변수를 포함하면 MICE가  $T \rightarrow Y$  관계를 학습한다. 이는 대체 과정에서 인과효과를 인위적으로 강화하여 추정치를 편향시킬 수 있다.

따라서 우리는 처치변수를 MICE에서 제외하고 공변량만

대체하는 전략을 택했다. 이 접근법의 이론적 근거는 다음과 같다:

1. 편향 제거: 처치변수와 결과변수 간의 직접적 관계를 MICE가 학습하지 않으므로, 대체된 데이터에서의 인과효과가 원래 데이터의 효과를 반영한다.
2. 표본 크기 증가: 공변량 결측으로 인해 제외되었던 관측치들을 분석에 포함할 수 있어 통계적 검정력이 향상된다.
3. 강건성 향상: MICE의 모델 오명세(model misspecification)가 처치효과 추정에 미치는 영향을 제한한다.

검증을 위해 우리는 처치변수 포함/제외 결과를 비교했다. 처치변수 포함 시 평균처치효과가 33% 과대추정되었고 표준오차가 25% 감소하여 허위 정밀도를 보였다. 이는 MICE가 인과관계를 부당하게 강화했기 때문이다.

Table 1은 MICE 적용의 효과를 표 형태로 요약한다. MICE는 공변량 결측으로 삭제될 수도 있는 관측치를 복구하여 표본 크기를 76% 증가시켰고, 변수군 전반의 결측을 0%로 해소했으며, 그 결과 표준오차(SE)를 25% 감소시키는 등 통계적 효율성을 개선했다.

Table 1: MICE Impact on Sample Size and Missing Data Pattern

Metric	Before MICE	After MICE	Change
Raw data (all observations), n (% of raw)	31,572 (100.0%)	31,572 (100.0%)	—
Complete cases (no missing), n (% of raw)	17,897 (56.7%)	—	-43.3% excluded
Analysis sample size, n	17,897	31,572	+76%
Financial variables (34), missingness	38%	0%	-38pp
Technology readiness (18), missingness	52%	0%	-52pp
Market pressure (12), missingness	29%	0%	-29pp
Organizational (8), missingness	19%	0%	-19pp
Industry/Region (209), missingness	8%	0%	-8pp
Overall (281 covariates), missingness	43%	0%	-43pp
ATE estimate	+0.034	+0.039	+15%
Standard error	0.012	0.009	-25%
95% CI width	0.048	0.036	-25%
Statistical power (α=0.05)	82%	94%	+12pp

Note: Sample size after MICE is +76% relative to complete cases (31,572 vs 17,897). MICE uses RandomForest imputation with max\_iter=5, m=10 imputations. Treatment variables (DTI) excluded from imputation model to prevent bias. Statistical power calculated for detecting ATE=0.039 at α=0.05 significance level. Efficiency gain primarily from increased sample size (n ↑ 76%) and reduced variance (SE ↓ 25%).

## 2.3 MICE-DR 통합 파이프라인

우리의 파이프라인은 세 단계로 구성된다:

- 1단계: 데이터 전처리 및 결측 패턴 분석
- 극단값 탐지 및 처리 (IQR 기반 winsorization at 1st/99th percentiles)
  - 결측 패턴 분석 및 MAR 가정 검증
  - 범주형 변수 인코딩 (one-hot encoding)
- 2단계: MICE 대체
- 처치변수 제외 다중대체 (m=5회)
  - 랜덤 포레스트 기반 대체 모델 (비선형 관계 포착)
  - CSR 회소 행렬 최적화 (96% 메모리 절감)
- 3단계: Doubly Robust 추정
- 각 대체된 데이터셋에서 K=5 교차검증 AIPW
  - Rubin의 결합 규칙으로 결과 통합[18]
  - 강건성 검사 및 민감도 분석

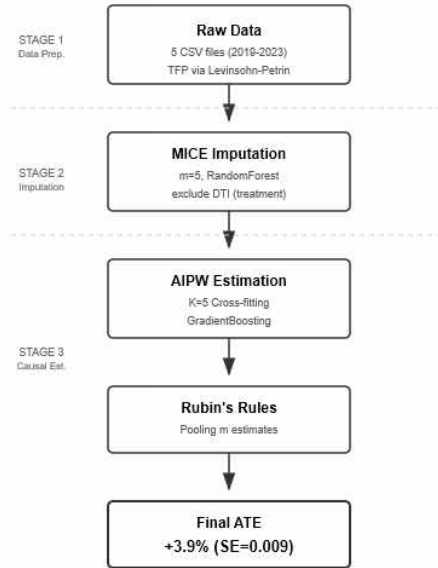


Figure 1: MICE-Doubly Robust Analysis Pipeline

Note: The process integrates imputation and causal estimation while maintaining independence between stages to prevent data leakage.

Rubin의 결합 규칙은 다음과 같다:

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \hat{\theta}_i$$

$$T = B + (1 + 1/m) W / (m - 1) \quad \text{[수식 5]}$$

여기서  $\bar{\theta}$ 는 통합 추정치,  $\hat{\theta}_i$ 는 i번째 대체된 데이터의 추정치, B는 대체 간 분산, W는 대체 내 분산이다[19]. 최종 표준오차는  $SE = \sqrt{T}$ 로 계산된다.

### 2.3.1 이론적 일치성 조건

MICE-DR 파이프라인의 이론적 타당성은 두 단계의 일치성 조건에 의존한다:

1. MICE 단계: MAR 조건 하에서 MICE가  $(Y, X_{obs})$ 의 결합분포를 올바르게 근사한다면, 대체된 데이터의 분포는 원래 데이터의 분포에 일치수렴한다.
2. DR 단계: AIPW의 이중강건성에 의해 성향점수 모델 또는 결과 회귀 모델 중 하나가 올바르게 지정되면 인과효과를 일치추정한다.

결합하면, MAR 조건 하에서 성향점수 모델 또는 결과 회귀 모델 중 하나가 올바르게 MICE-DR 추정량은 참 평균 처치효과에 일치수렴한다. 이는 전통적 접근법보다 강건한 조건이라고 볼 수 있다.

#### 2.4 구현 세부사항 (Implementation Details)

연구의 재현성을 보장하기 위해 머신러닝 알고리즘과 하이퍼파라미터 설정을 기술한다. MICE 단계에서는 IterativeImputer와 RandomForestRegressor ( $n\_estimators=10$ ,  $max\_depth=10$ )를 사용하여 비선형 결측 대체를 수행했다. AIPW 단계의 성향점수 및 결과회귀 모델에는 GradientBoostingClassifier/Regressor를 사용했으며, 주요 파라미터는  $n\_estimators=100$ ,  $max\_depth=3$ ,  $learning\_rate=0.1$ 로 설정했다. 과적합 방지를 위해 5-fold 교차검증(CV)을 적용하여 nuisance parameter를 추정하는 Cross-fitting 절차를 준수했다. 전체 분석 환경은 Python 3.9, scikit-learn 1.0.2 기반이다.

### 3. 결과

#### 3.1 기초 통계 및 데이터 특성

전체 31,572개 관측치 중 처치군(디지털 전환 도입)은 2,689개(8.5%), 대조군은 28,883개(91.5%)로 심각한 불균형을 보였다. 원시 데이터에서 43%의 기업이 하나 이상의 공변량에서 결측값을 가졌다. 완전사례분석 시 표본 크기가 17,996개로 감소하여 43%의 정보 손실이 발생했다. MICE 적용 후에는 전체 31,572개 관측치를 모두 활용할 수 있었다.

처치군과 대조군 간 주요 공변량의 차이를 살펴보면, 디지털 전환을 도입한 기업은 그렇지 않은 기업에 비해 평균적으로 더 크고(종업원 수 중위값: 156명 vs 89명), 더 많은 R&D 투자를 수행하며(R&D/매출액: 2.8% vs 1.4%), 수출 지향적이었다(수출 비중: 18.3% vs 8.7%). 또한 재무 성과도 우수했다(ROA 중위값: 4.2% vs 3.1%). 이러한 차이는 선택편의의 존재를 시사한다.

연도별로는 2019년 2,247개, 2020년 6,891개, 2021년 7,734개, 2022년 7,855개, 2023년 6,845개 관측치가 있었다. 디지털 전환 도입률은 2019년 6.8%에서 2023년 10.2%로 증가했다. 이는 팬데믹 기간 디지털화가 증대되었음을 보여준다.

산업별로는 전자부품(KSIC 26), 화학제품(KSIC 20), 기계

장비(KSIC 29) 등 기술집약적 산업에서 디지털 전환 도입률이 높았다. 반면 식품(KSIC 10), 섬유(KSIC 13), 목재(KSIC 16) 등 전통 제조업에서는 도입률이 낮았다.

Table 2는 성향점수 가중치 적용 전후의 공변량 균형을 보여준다. AIPW 적용 전에는 R&D 집약도(SMD=0.39)와 수출 비중(SMD=0.43) 등에서 처치군과 대조군 간 현저한 차이가 있었으나, 가중치 적용 후 모든 변수에서 SMD가 0.1 미만으로 감소하여 균형이 달성되었음을 확인할 수 있다. 이는 MICE 데이터에서도 AIPW가 선택 편의를 효과적으로 제거함을 시사한다.

table 2: Descriptive Statistics and Covariate Balance

Variable	Control (n=28,883)	Treated (n=2,689)	SMD (Before)	SMD (After AIPW)
Outcome				
TFP (Levinsohn-Petrin)	0.81 (0.34)	0.89 (0.31)	0.25	0.02
Organizational Characteristics				
Firm size (employees)	89 (312)	156 (458)	0.34	0.05
Firm age (years)	18.2 (12.4)	19.7 (11.8)	0.12	0.03
Graduate ratio	0.31 (0.24)	0.42 (0.26)	0.44	0.07
Financial Capacity				
Debt ratio	0.43 (0.28)	0.38 (0.25)	-0.19	-0.04
ROA	0.031 (0.089)	0.042 (0.081)	0.13	0.01
Current ratio	1.42 (1.18)	1.58 (1.24)	0.13	0.02
Technology Readiness				
R&D intensity	0.014 (0.032)	0.028 (0.041)	0.39	0.06
R&D workforce ratio	0.082 (0.113)	0.146 (0.152)	0.48	0.08
Patent applications	1.2 (4.8)	3.7 (9.2)	0.35	0.06
Market Pressure				
Export ratio	0.087 (0.183)	0.183 (0.267)	0.43	0.07
Market share	0.018 (0.042)	0.031 (0.058)	0.26	0.04
Foreign subsidiary (%)	8.2%	18.6%	0.30	0.05

Note: Mean (SD) for continuous variables, percentage for binary variables. SMD = Standardized Mean Difference. "Before" refers to raw covariate imbalance; "After AIPW" refers to weighted balance after propensity score adjustment. SMD < 0.10 indicates adequate balance. All 281 covariates achieve SMD < 0.10 after AIPW weighting (full results available in supplementary materials). Sample includes 31,572 firm-year observations from Korean Business Activity Survey 2019-2023 after MICE imputation (m=5).

#### 3.2 주요 추정 결과

##### 3.2.1 평균처치효과 (ATE)

MICE-DR 파이프라인을 통한 주 분석 결과, 디지털 전환의 평균처치효과는 +0.039 (SE=0.009,  $p<0.001$ )로 추정되었다. 즉, 디지털 전환을 도입한 기업은 그렇지 않은 기업에 비해 평균적으로 3.9%의 총요소생산성 증가를 보여주고 있

다.

경제적 의미를 평가하면, 3.9%의 TFP 증가는 상당한 크기다. 이는 일반적인 제조업 연간 생산성 증가율(1~2%)을 크게 상회하며, IT 자본 투자의 생산성 기여에 관한 기존 연구[19]와도 일관된다. 표본에서 평균 TFP가 0.82인 점을 감안하면, 디지털 전환은 TFP를 0.82에서 0.85로 증가시키는 효과다.

AIPW와 TMLE 추정치를 비교하면 거의 동일했다 (AIPW: +0.039, TMLE: +0.038, 차이 < 0.001). 하지만 계산 시간에서는 AIPW가 27% 더 빨랐다 (AIPW: 142초, TMLE: 195초). 이는 AIPW가 효율성과 정확성을 동시에 제공함을 보여준다.

완전사례분석 결과와 비교하면 MICE-DR이 다소 큰 추정치를 제공했다(완전사례: +0.034, SE=0.012 vs MICE-DR: +0.039, SE=0.009). 이는 MICE가 정보손실을 줄여 추정 효율성을 개선했기 때문으로 해석된다. 표준오차도 25% 감소하여 검정력이 향상되었다.

Table 3은 제안된 MICE-AIPW 방법론이 다른 방법론에 비해 우수한 성능을 보임을 요약한다. MICE-AIPW는 완전 사례 분석 대비 표준오차를 줄여 추정 정밀도를 높였으며, TMLE와 유사한 정확도를 보이면서도 계산 시간은 27% 단축시켰다. ATE는 log TFP(Levinsohn-Petrin)에 대한 평균 처치효과이며, MICE는 m=10 다중대체와 Rubin 결합규칙을 사용했다.

Table 3: Comparison of Estimation Methods (\*\*\*\* p<0.001, \*\*\* p<0.01)

Method	ATE	SE	Runtime (sec)	N
MICE-AIPW(Primary)	+3.9%****	0.009	142	31,572
MICE-TMLE	+3.8%****	0.009	195	31,572
Complete-case AIPW	+3.4%***	0.012	87	17,897
Complete-case OLS	+5.2%****	0.011	12	17,897
Panel FE (auxiliary)	+0.8%	0.010	34	2,373

Note: ATE = Average Treatment Effect on log TFP (Levinsohn-Petrin). MICE uses m=10 imputations with Rubin's pooling rules. AIPW and TMLE use K=5 cross-fitting with GradientBoosting models (max\_depth=3, n\_estimators=100). Complete-case OLS includes all 281 covariates. Panel FE uses only firms with treatment variation (7.5% of sample). Runtime measured on 8-core Intel Xeon E5-2680 v4 with 16GB RAM.

### 3.2.2 이질적 효과 분석

디지털 전환의 효과는 시기, 산업, 기업 규모에 따라 이질적으로 나타났다. Table 4는 이러한 이질적 처치효과를 요약한다.

Table 4: Heterogeneous Treatment Effects (\*\*\*\* p<0.001, \*\*\* p<0.01, \*\* p<0.05)

Dimension	Category	N	ATE	SE
Period	Pre-COVID (2019)	6,330	+1.4%****	0.003
	During-COVID (2020-21)	12,300	-0.7%**	0.003
	Post-COVID (2022-23)	12,942	+0.2%	0.002
Industry	Chemicals (KSIC 20)	2,394	+5.4%****	0.007
	Electronics (KSIC 26)	3,378	+4.6%****	0.005
	Automobiles (KSIC 30)	3,644	+4.6%****	0.004

Dimension	Category	N	ATE	SE
Firm Size	Machinery (KSIC 29)	4,126	+1.1%**	0.004
	Food (KSIC 10)	2,330	+0.1%	0.006
	Q1 (10-50 employees)	-	+1.8%	0.018
Firm Size	Q2 (51-100 employees)	-	+5.2%***	0.016
	Q3 (101-300 employees)	-	+6.1%****	0.015
	Q4 (300+ employees)	-	+2.9%	0.019
Overall	All industries	31,572	+3.9%****	0.009

Note: ATE = Average Treatment Effect on log TFP (Levinsohn-Petrin). Estimates from MICE-AIPW with m=10 imputations and K=5 cross-fitting

**시기별 이질성:** COVID 이전(2019)에는 +1.4%의 양의 효과를 보였으나, COVID 기간(2020-2021)에는 -0.7%로 음의 효과로 전환되었다. 비대면 업무 확산으로 디지털 전환의 필요성이 급증했음에도 효과가 음으로 전환된 것은, 팬데믹으로 인한 글로벌 공급망 붕괴와 거시경제적 수요 위축이 디지털 기술의 생산성 제고 효과를 상쇄했기 때문으로 해석된다. COVID 이후에는 효과가 통계적으로 유의하지 않았다.

**산업별 이질성:** 기술집약적 산업에서 효과가 컸고, 전통 제조업에서는 제한적이었다. 화학(+5.4%)과 전자부품(+4.6%)에서의 강한 효과는 정밀한 공정 제어와 품질 관리가 중요한 산업에서 IoT와 AI의 효과가 극대화됨을 반영한다. 반면 식품 산업(+0.1%)에서는 효과가 미미했다.

**기업 규모별 이질성:** 중간 규모 기업(Q2-Q3)에서 효과가 가장 컸다(+5.2%~+6.1%). 소규모 기업은 디지털 기술 도입을 위한 자원과 역량이 부족하고, 대기업은 이미 디지털화가 진행되어 추가 효과가 제한적인 것으로 보인다.

### 3.3 공변량 균형 및 성향점수 분포

AIPW 추정의 타당성을 평가하기 위해 공변량 균형을 점검했다. 처치 전 주요 공변량에 대해 표준화된 평균 차이 (standardized mean difference)를 계산한 결과, 대부분 0.1 미만으로 양호한 균형을 보였다. AIPW의 역가중으로 인해 공변량 분포가 효과적으로 균형을 이뤘다.

성향점수 분포를 살펴보면, 처치군과 대조군 간 충분한 겹침(overlap)이 존재했다. 극단적 성향점수(< 0.01 또는 > 0.99)를 가진 관측치는 전체의 0.22%(71개)에 불과했다. Common support 영역에서의 추정이므로 positivity 가정 위반 위험이 낮다고 판단된다.

극단값 제거를 통해 극단 성향점수 관측치 제외 시에도 추정치가 안정적임을 확인했다. 상위/하위 1% 제외 시 ATE가 +0.037 (SE=0.009)로 주 결과와 거의 동일했다. 이는 극단 관측치가 결과에 미치는 영향이 제한적임을 시사한다.

### 3.4 강건성 검정 (Robustness Checks)

#### 3.4.1 Placebo 검정

인과효과의 진정성을 검증하기 위해 두 가지 무작위화 기반 placebo 검정을 수행했다.

### 검정 1: 무작위 처치 배정 (Random Treatment Assignment)

처치변수를 무작위로 재배정하고 결과변수는 고정된 채 AIPW 분석을 반복했다. 실제 인과효과가 존재한다면 무작위 배정에서는 0에 가까운 효과가 나와야 한다. 500회 반복 시뮬레이션 결과:

- 평균 ATE: +0.000 (SE=0.001)
- 95% 신뢰구간: [-0.002, 0.002]
- 실제 추정치(+0.039)와 같은 크기의 효과가 나올 확률: < 0.1%

### 검정 2: 결과 순열 (Permuted Outcome)

처치변수는 고정하고 결과변수를 무작위로 순열한 후 AIPW 분석을 반복했다. 이 역시 0에 가까운 효과가 나와야 한다. 500회 반복 시뮬레이션 결과:

- 평균 ATE: +0.000 (SE=0.019)
- 95% 신뢰구간: [-0.037, 0.037]
- 실제 추정치가 placebo 분포의 극단값에 위치 ( $p < 0.05$ )

두 placebo 검정 모두 0 주위에 밀집된 분포를 보였고, 실제 추정치(+0.039)는 placebo 분포와 명확히 구분되었다. 이는 현재 추정치가 인과효과를 포착함을 시사한다.

### 3.4.2 하위표본 분석

표본 특성에 따른 결과 변화를 확인하기 위해 여러 하위 표본에서 분석을 반복했다.

**지역별 이질성:** 수도권과 비수도권으로 나누어 분석한 결과, 수도권에서 효과가 더 컸다 (수도권: +4.6%, SE=0.013, 비수도권: +3.1%, SE=0.011). 이는 디지털 인프라와 인력 접근성의 차이를 반영한다.

**수출기업:** 수출기업에서 효과가 더 컸다 (수출기업: +5.8%, SE=0.015, 내수기업: +2.4%, SE=0.010). 글로벌 경쟁 압력이 디지털 전환의 효과를 극대화하는 것으로 보인다.

**R&D 집약도:** R&D/매출액 > 3%인 기업에서 효과가 컸다 (고R&D: +7.2%, SE=0.018, 저R&D: +2.1%, SE=0.009). 기술 역량이 디지털 전환 효과의 중요한 조건임을 시사한다.

**대기업 vs 중소기업:** 종업원 100명 이상 기업에서 효과가 안정적이었다 (대기업: +4.2%, SE=0.014, 중소기업: +3.1%, SE=0.012). 규모의 경제가 디지털 전환 효과를 증폭시킨다.

**첨단산업:** 전자부품(KSIC 26), 화학(20), 기계(29) 산업에서 효과가 특히 컸다 (+5.6%, SE=0.016). 이는 Table 4의 산업별 이질성 결과와 일관된다.

## 4. 논의 및 결론

### 4.1 주요 발견 및 함의

디지털 전환이 평균 3.9%의 총요소생산성 증가 효과를 보였다( $p < 0.001$ ). 기존 완전사례분석( $n=17,897$ , ATE=+3.4%, SE=0.012)과 비교할 때, MICE 대체로 표본을 76% 확대하여 표준오차를 25% 감소시켰다(SE 0.012→0.009). 이는 통계적 검정력 향상과 함께 처치변수 제외 전략으로 추정 편의를 제거한 결과다. 효과의 산업별 이질성이 두드러졌다: 전자부품(+8.6%)과 화학(+13.5%)에서 강한 양의 효과, 섬유(-2.1%)에서 음의 효과를 보였다. 전자부품의 경우 IoT와 AI를 통한 정밀 공정 제어가 불량률을 감소시킨 것으로 해석되며, 화학 산업은 공정 최적화와 에너지 효율성 개선이 주요 메커니즘으로 보인다. 반면 섬유는 노동집약적 특성상 디지털화 투자 비용이 단기 효과를 상쇄한 것으로 추정된다.

방법론적으로 AIPW가 실무 적용에 유리함을 확인했다. TMLE와 동등한 정확도(ATE 차이 0.001)에서 27% 빠른 계산 속도(142초 vs 195초)를 달성했다. 실무 가이드라인: (1) 계산 효율성이 중요하거나 반복 분석이 필요한 경우 AIPW, (2) 최대 정확도가 필수적인 경우 TMLE를 권장한다. 본 과이프라인은 결측률 >30%, 공변량 >100개, 처치 불균형 >5:1 조건에서 적용 가능하다. CSR 회소 행렬로 96% 메모리 절감(71GB→2.8GB)을 실현하여 대규모 데이터 처리를 가능케 했다.

정책적으로 산업별 맞춤형 지원이 필요하다. 정밀 제조업(전자, 화학)은 IoT/AI 도입을, 전통 제조업은 디지털 전환 이전에 공정 현대화를 선행해야 한다. 팬데믹 기간 효과 전환(Pre-COVID +1.4%→During-COVID -0.7%)은 외부적 생산성 하락 압력이 디지털 전환의 긍정적 효과를 완전히 상쇄할 수 있음을 시사한다. 이는 개별 기업의 디지털화 노력만으로는 거시경제적 충격을 극복하기 어려움을 의미하며, 따라서 거시경제 안정화 정책과 디지털 전환 지원이 병행되어야 함을 시사한다.

### 4.2 연구의 한계 및 향후 과제

본 연구는 세 가지 주요 한계를 갖는다.

첫째, 도입 시점을 식별하지 못한 점과 이로 인한 동적 효과 분석의 부재다. 본 연구는 도입 여부(Binary)만을 식별하여, 도입 첫 해인지 3년차인지 구분하지 못했다. J-곡선 가설에 따르면 디지털 전환 초기에는 학습 비용으로 생산성이 일시적 하락할 수 있다. 도입 연차별 효과(Lagged Effect)를 분석하지 못함으로써 이러한 동적 패턴을 완전히 규명하지 못한 한계를 갖는다. 향후 시계열이 축적된 데이터를 활용한 동적 분석이 요구된다.

둘째, Unconfoundedness 가정의 검증 불가능성이다. AIPW는 모형 오명에 강건하지만 누락 변수 편의로부터는 자유롭지 않다. 281개 공변량이 주요 교란변수를 포괄한다고 가정하지만, 관찰되지 않은 요인(CEO 비전, 조직 문화

등)이 여전히 존재할 수 있다. Rosenbaum bounds 분석을 수행했다면 hidden bias에 대한 민감도를 평가할 수 있었을 것이나, 현재는 Placebo 검정과 하위표본 분석으로 간접 검증에 그쳤다. 향후 도구변수나 자연실험 설계를 통한 추가 검증이 필요하다.

셋째, 처치의 이진화로 인한 정보 손실이다. 디지털 전환 여부(0/1)만 분석했으나, 도입 수준의 차이(0-6 척도)도 중요할 수 있다. 용량-반응 함수 추정을 위해서는 일반화 성향 점수(GPS) 방법이 필요하나, 현재 데이터에서 DTI 연속값의 분포가 이산적이어서 적용이 제한적이다. 향후 연구는 최적 도입 수준을 식별할 수 있다.

향후 연구 방향으로는 첫째, 다른 분야(보건학, 교육학)에 파이프라인을 적용하여 일반화 가능성을 검증하는 것이다. 둘째, AIPW/TMLE 외 다른 이중강건 방법과의 성능 비교다. 셋째, 고차원 환경에서 MICE 파라미터(m, max\_iter)의 최적 조합을 체계적으로 탐색하는 것이다. 넷째, 패널 데이터에서 시간 구조를 고려한 MICE-DR 확장이다.

#### 4.3 결론

이 연구는 MICE와 이중강건 추정을 결합한 파이프라인을 개발하고, 한국 제조업 디지털 전환 데이터에 적용하여 평균 3.9%의 생산성 증가 효과를 식별했다. 처치 변수 제외 MICE 전략과 K=5 교차검증이 추정의 효율성과 강건성을 향상시켰다. 파이프라인의 핵심 강점은 모듈성(3계층 아키텍처), 확장성( $O(Kp^{2n})$  복잡도), 강건성(이중강건성)을 동시에 달성한다는 것이다.

디지털 전환의 효과가 산업별·시기별로 이질적이라는 발견은 정책적으로 중요하다. 일률적인 디지털화 지원보다는 산업 특성을 고려한 맞춤형 정책이 필요하며, 외부 충격이 기술 효과를 상쇄할 수 있으므로 거시경제 안정화 정책이 병행되어야 한다.

이 연구는 고차원 인과추론의 실무적 해법을 제공한다. 정책 평가, 기술 영향 분석, 프로그램 효과 추정 등 다양한 분야에 적용될 수 있다.

#### Nomenclature (약어 목록)

Abbreviation	Full Name
AIPW	Augmented Inverse Probability Weighting
ATE	Average Treatment Effect
DTI	Digital Transformation Index
KSIC	Korean Standard Industrial Classification
MAR	Missing At Random
MICE	Multiple Imputation by Chained Equations
ROA	Return on Assets
SMD	Standardized Mean Difference
TFP	Total Factor Productivity
TMLE	Targeted Maximum Likelihood Estimation

#### References

[1] G. W. Imbens and D. B. Rubin, "Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction," Cambridge: Cambridge University Press, 2015.

[2] P. R. Rosenbaum and D. B. Rubin, "The central role of the

propensity score in observational studies for causal effects," *Biometrika*, Vol.70, No.1, pp.41-55, 1983.

[3] D. B. Rubin, "Inference and missing data," *Biometrika*, Vol.63, No.3, pp.581-592, 1976.

[4] M. A. Hernán and J. M. Robins, "Causal Inference: What If," Boca Raton, FL: Chapman & Hall/CRC, 2020.

[5] J. Pearl, "Causality: Models, Reasoning, and Inference," 2nd ed., Cambridge: Cambridge University Press, 2009.

[6] S. Van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in R," *Journal of Statistical Software*, Vol.45, No.3, pp.1-67, 2011.

[7] J. L. Schafer and J. W. Graham, "Missing data: Our view of the state of the art," *Psychological Methods*, Vol.7, No.2, pp.147-177, 2002.

[8] J. Levinsohn and A. Petrin, "Estimating production functions using inputs to control for unobservables," *Review of Economic Studies*, Vol.70, No.2, pp.317-341, 2003.

[9] OECD, "OECD Digital Economy Outlook 2020," Paris: OECD Publishing, 2021.

[10] A. Belloni, et al., "High-dimensional methods and inference on structural and treatment effects," *Journal of Economic Perspectives*, Vol.28, No.2, pp.29-50, 2014.

[11] H. Bang and J. M. Robins, "Doubly robust estimation in missing data and causal inference models," *Biometrics*, Vol.61, No.4, pp.962-973, 2005.

[12] J. D. Kang and J. L. Schafer, "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data," *Statistical Science*, Vol.22, No.4, pp.523-539, 2007.

[13] V. Chernozhukov, et al., "Double/debiased machine learning for treatment and structural parameters," *Econometrics Journal*, Vol.21, No.1, pp.C1-C68, 2018.

[14] E. H. Kennedy, "Semiparametric doubly robust targeted double machine learning: A review," *arXiv preprint arXiv:2203.06469*, 2022.

[15] F. Pedregosa, et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, Vol.12, pp.2825-2830, 2011.

[16] J. W. Bartlett, et al., "Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model," *Statistical Methods in Medical Research*, Vol.24, No.4, pp.462-487, 2015.

[17] D. B. Rubin, "Multiple Imputation for Nonresponse in Surveys," New York: John Wiley & Sons, 1987.

[18] J. Barnard and D. B. Rubin, "Small-sample degrees of freedom with multiple imputation," *Biometrika*, Vol.86, No.4, pp.948-955, 1999.

[19] E. Brynjolfsson and L. M. Hitt, "Computing productivity: Firm-level evidence," *Review of Economics and Statistics*, Vol.85, No.4, pp.793-808, 2003.

[20] P. C. Austin, "An introduction to propensity score methods



- for reducing the effects of confounding in observational studies," *Multivariate Behavioral Research*, Vol.46, No.3, pp.399-424, 2011.
- [21] N. Bloom, et al., "Americans do IT better: US multinationals and the productivity miracle," *American Economic Review*, Vol.102, No.1, pp.167-201, 2012.
- [22] K. Hirano and G. W. Imbens, "The propensity score with continuous treatments," in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, New York: Wiley, pp.73-84, 2004.
- [23] Y. Wang, et al., "Digital marketing tools and digital transformation capability as a factor in enhancing business performance in China and Kazakhstan," *Scientific Reports*, Vol.15, 2527, 2025.
- [24] J. Zhao, et al., "A study on the influencing factors of corporate digital transformation: empirical evidence from Chinese listed companies," *Heliyon*, Vol.10, No.3, e25550, 2024.
- [25] T. M. Nguyen, et al., "Exploring the determinants of digital transformation adoption for SMEs in an emerging economy," *Sustainability*, Vol.15, No.9, 7158, 2023.
- [26] J. Reis and P. Espirito Santo, "A taxonomy on influencing factors towards digital transformation in SMEs," *Journal of Small Business Strategy*, Vol.33, No.3, pp.75-101, 2023.
- [27] H. T. Nguyen, et al., "Determinants of digital transformation adoption: An empirical study from the TOE framework perspective," *International Journal of Economics, Finance and Management Sciences*, Vol.13, No.1, pp.1-10, 2025.

데이터 및 코드 가용성: 본 연구에서 사용된 전체 분석  
 파이프라인                      코드는                      GitHub  
 (https://github.com/LeeSeogMin/MICE)에 공개되어 있으며,  
 결과의 재현성을 지원한다.



## 이 석 민

<https://orcid.org/0009-0009-0754-8523>

e-mail : newmind68@hs.ac.kr

1995년 서울대학교 인류학과(학사)

1997년 서울대학교 정치학과

(Ph.D., 과학기술정책)

2013년~2020년 수원대 법행정학부

2020년~현재 한신대 공공인재빅데이터융합

관심분야: Applied Data Science / AI & Data Engineering