Pedagogical Report: Teaching Causal Inference Through Startup Hiring Analysis

Author: Anusha Prakash
Course: INFO 7390 - Advances in Data Sciences and Architecture
Date: December 2025
Project: Causal Inference for Startup Hiring: Does Funding Cause Hiring?

Github: https://github.com/AnushaPrakash03/causal-inference-80Days-to-Stay

Video link: https://youtu.be/w-EiBWECMQ4

Executive Summary

This pedagogical report documents the design, implementation, and assessment strategy for a comprehensive educational module on causal inference methods in data science. The module teaches Propensity Score Matching (PSM) and Difference-in-Differences (DiD) through a real-world application: determining whether venture funding causally increases startup hiring. The project bridges theoretical econometrics with practical data science implementation, targeting graduate students in data science, business analytics, and related fields. By grounding abstract causal inference concepts in a tangible problem—one that directly impacts international students seeking employment—the module achieves high engagement while maintaining technical rigor.

---

1. Teaching Philosophy (Context & Approach)

1.1 Target Audience

Primary Audience: Graduate students in data science, analytics, or related quantitative programs with:

- **Prerequisites:** Basic statistics (hypothesis testing, regression), Python programming, pandas/numpy proficiency
- **No prior knowledge required:** Econometrics, causal inference, panel data methods
- **Typical background:** Industry professionals or recent graduates seeking to strengthen causal reasoning skills

Secondary Audience: Data scientists in industry who need to move beyond predictive modeling to answer "what if" questions about interventions, policy changes, or business decisions.

1.2 Pedagogical Philosophy: Problem-First Learning

This module employs a **problem-first pedagogical approach** rather than theory-first instruction. This decision is grounded in constructivist learning theory and adult learning principles:

Why Problem-First Works:

1. **Immediate motivation:** Students understand WHY they need causal inference before diving into technical details
2. **Concrete anchoring:** Abstract concepts (counterfactuals, selection bias) become tangible when tied to real decisions
3. **Cognitive scaffolding:** Students build from familiar concepts (correlation, regression) to new ones (propensity scores, parallel trends)
4. **Transfer learning:** Real-world context enables students to apply methods to their own domains

The Central Problem:

"Does receiving $5M+ venture funding causally increase a startup's hiring over the next 12 months?"

This question is powerful pedagogically because:

- **Real stakes:** International students depend on accurate predictions for visa sponsorship
- **Intuitive but wrong:** Correlation is obvious (funded companies DO hire more), but causation is questionable
- **Requires rigor:** Simple comparison is biased; proper causal inference is essential
- **Generalizable:** The framework applies to any treatment effect question (policy evaluation, A/B testing, medical trials)

1.3 Learning Objectives (Bloom's Taxonomy)

By completing this module, students will be able to:

Remember & Understand (Knowledge):

- Define the fundamental problem of causal inference vs. correlation
- Explain the potential outcomes framework (Rubin Causal Model)
- Describe the key assumptions underlying PSM and DiD

Apply (Skills):

- Implement Propensity Score Matching from scratch in Python
- Apply Difference-in-Differences to panel data
- Generate and interpret diagnostic plots (parallel trends, covariate balance)

Analyze (Critical Thinking):

- Assess whether causal assumptions hold in a given dataset
- Diagnose violations of parallel trends or unconfoundedness
- Compare strengths/weaknesses of different causal methods

Evaluate (Judgment):

- Critique observational studies claiming causal effects
- Determine which causal method is appropriate for a research question
- Conduct sensitivity analyses for hidden bias

Create (Synthesis):

- Design causal analyses for novel business or policy questions
- Implement robustness checks and validate results
- Communicate causal findings to non-technical stakeholders

1.4 Instructional Design Principles

1. Worked Examples with Variation: The module provides complete implementations (PSM, DiD) as worked examples, then asks students to apply the same framework to variations (different industries, different treatment definitions).

2. Productive Failure: Students first attempt a "naive comparison" (correlation) and see it fail—this cognitive dissonance motivates learning proper causal methods.

3. Interleaved Practice: Rather than mastering PSM completely before moving to DiD, students alternate between methods, comparing their assumptions and results. This prevents rote memorization and promotes deeper understanding.

4. Authentic Assessment: Exercises use realistic (though synthetic) data with real-world context. Students aren't just running code—they're making decisions that affect people's lives.

5. Metacognitive Reflection: Throughout the tutorial, students are prompted to reflect: "Why did this assumption matter?" "What happens if it's violated?" "How confident should we be in this result?"

---

2. Concept Deep Dive (Technical Foundations)

2.1 The Fundamental Problem of Causal Inference

The Core Challenge: We want to know the causal effect of treatment (funding) on outcome (hiring) for a given unit (company). In an ideal world, we would observe both:

- Y_{1i}: Outcome for company i WITH funding
- Y_{0i}: Outcome for company i WITHOUT funding

The **individual treatment effect** is: $\tau_i = Y_{1i} - Y_{0i}$

The Problem: We can never observe both potential outcomes for the same unit at the same time. This is the **fundamental problem of causal inference**—it's a missing data problem, not a statistical estimation problem.

Pedagogical Emphasis: This is taught through a concrete example: "Imagine Company A received funding and now has 50 employees. How many would they have WITHOUT funding? We'll never know—that universe doesn't exist." This tangibility helps students internalize why causal inference is hard.

2.2 Selection Bias: Why Correlation ≠ Causation

The Confounding Problem: In observational data, treatment assignment is not random. Better companies are MORE likely to:

1. Receive funding (they're more attractive to VCs)
2. Grow faster (they have better products, teams, markets)

This creates **selection bias**:

Naive Difference = $E[Y \mid Treated] - E[Y \mid Control]$

$= E[Y_1 \mid Treated] - E[Y_0 \mid Control]$

$= ATT + Selection\ Bias$

Where:

- ATT = Average Treatment Effect on Treated (what we want)

- Selection Bias = $E[Y_0 \mid Treated] - E[Y_0 \mid Control]$ (confounding)

Teaching Strategy: Students first calculate the naive difference (25.5% in our data) and compare it to the true ATE (20%). The 5.5 percentage point gap is selection bias—they can see it, quantify it, and understand why it matters.

2.3 Method 1: Propensity Score Matching (PSM)

Core Idea: Create "synthetic twins" by matching each treated unit to a similar control unit based on observable characteristics.

Mathematical Framework:

1. **Propensity Score:** Probability of receiving treatment given covariates
2. $e(X) = P(T = 1 \mid X)$

Estimated via logistic regression: $e(X) = 1 / (1 + \exp(-X\beta))$

3. **Matching:** For each treated unit i, find control unit j with closest propensity score
4. $j(i) = \underset{j \in Control}{\operatorname{argmin}} |e(X_i) - e(X_j)|$
5. **ATT Estimation:** Average difference in outcomes between matched pairs
6. $ATT = (1/N\_treated) \Sigma [Y_i - Y\_j(i)]$

Key Assumption: Unconfoundedness (CIA)

$(Y_1, Y_0) \perp T \mid X$

"Conditional on observables X, treatment assignment is as good as random"

Why This Assumption Matters: If there are unobserved confounders (e.g., founder quality, market timing), PSM will still be biased. This is why sensitivity analysis is crucial.

Pedagogical Implementation:

- Students see the propensity score distribution for treated vs. control (common support check)
- They examine covariate balance before/after matching (standardized mean differences)
- They conduct placebo tests and sensitivity analyses

2.4 Method 2: Difference-in-Differences (DiD)

Core Idea: Compare the CHANGE in outcomes over time between treated and control groups.

Mathematical Framework:

Standard DiD regression:

$Y_{it} = \beta_0 + \beta_1 \cdot Treated_i + \beta_2 \cdot Post_t + \beta_3 \cdot (Treated_i \times Post_t) + \varepsilon_{it}$

Where:

- β_3 = DiD estimator (causal effect)

- $Treated_i$ = 1 if unit i is treated

- $Post_t$ = 1 if time t is post-treatment

Intuition:

DiD = [E[Y | Treated, Post] - E[Y | Treated, Pre]]

- [E[Y | Control, Post] - E[Y | Control, Pre]]

= "Difference in differences"

Key Assumption: Parallel Trends

$E[Y_{0t} \mid Treated] - E[Y_{0t-1} \mid Treated] = E[Y_{0t} \mid Control] - E[Y_{0t-1} \mid Control]$

"In the absence of treatment, treated and control groups would follow parallel trends"

Why This Assumption Matters: If treated companies were already growing faster BEFORE treatment, DiD will overestimate the effect. This is why pre-trend testing is essential.

Pedagogical Implementation:

- Students visualize parallel trends graphically (most intuitive check)
- They run pre-trend regression tests (statistical validation)
- They conduct event studies to see dynamic treatment effects
- They perform placebo tests (fake treatment in pre-period)

2.5 Connecting PSM and DiD

When to Use Each Method:

Criterion	**PSM**	**DiD**
Data requirement	Cross-sectional	Panel (multiple time periods)
Assumption	Unconfoundedness	Parallel trends
Controls for	Observable confounders	Time-invariant unobservables
Strengths	No time series needed	Controls for selection on unobservables
Weaknesses	Assumes no hidden bias	Requires pre-treatment data

Key Pedagogical Point: These methods are complementary, not competing. If results agree across methods, we have stronger evidence of causality. If they disagree, we need to investigate why (likely assumption violation).

2.6 Connection to Course Themes

GIGO (Garbage In, Garbage Out): Causal inference is the ultimate example of GIGO. If your identifying assumptions are wrong (e.g., parallel trends violated), your causal estimates are meaningless, no matter how sophisticated your code.

Data Quality over Quantity: 10,000 observations with severe selection bias are less useful than 100 observations from a well-designed experiment. Causal inference teaches students to prioritize research design over sample size.

Botspeak/AI Context: As LLMs become more prevalent, the ability to distinguish correlation from causation becomes critical. AI can identify patterns, but only humans can design causal analyses. This module prepares students for a future where causal reasoning is a core data science skill.

---

3. Implementation Analysis (Design Decisions)

3.1 Technology Stack Rationale

Core Libraries:

1. **pandas/numpy:** Industry standard for data manipulation; students already familiar
2. **statsmodels:** Econometric focus; provides proper causal inference tools (clustered SEs, OLS with formulas)
3. **scikit-learn:** For propensity score estimation (LogisticRegression); familiar API
4. **matplotlib/seaborn:** Publication-quality visualizations; highly customizable

Why NOT Other Options:

- **EconML (Microsoft):** Too complex for beginners; abstracts away important details
- **DoWhy (Microsoft):** Excellent for causal graphs, but overkill for this intro module
- **R (Matching, plm packages):** Students in Python-first programs; translation friction
- **Stata:** Proprietary; not accessible to all students

Design Philosophy: Use widely-adopted tools that students can apply immediately in industry, while maintaining transparency about what's happening "under the hood."

3.2 Key Implementation Decisions

3.2.1 Propensity Score Matching

Decision 1: Nearest Neighbor Matching (vs. Kernel, Stratification)

- **Why:** Simplest to explain; intuitive "find your twin" concept
- **Tradeoff:** Less efficient than kernel matching, but easier to debug and validate
- **Educational value:** Students can inspect individual matches and understand the process

Decision 2: Caliper = 0.01 (1% of PS distribution)

- **Why:** Ensures high-quality matches; prevents bad pairs
- **Tradeoff:** Some treated units go unmatched (reduces sample size)
- **Pedagogical point:** Trade-off between bias (bad matches) and variance (smaller sample)

Decision 3: Matching Without Replacement

- **Why:** Each control unit used once; avoids over-reliance on few "good" controls
- **Tradeoff:** Fewer matches possible
- **Teaching moment:** Discuss when replacement is appropriate (e.g., if control group is much larger)

Code Design:

```
class PropensityScoreMatcher:
  def fit(self, data):
    # Estimate propensity scores

  def match(self, data, caliper=0.01):
    # Perform matching

  def check_balance(self, matches):
    # Validate balance

  def estimate_att(self, matches):
    # Calculate treatment effect
```

Pedagogical rationale: Object-oriented design mirrors real-world data science workflows. Students can instantiate, fit, and inspect—familiar pattern from scikit-learn.

3.2.2 Difference-in-Differences

Decision 1: Two-Way Fixed Effects (TWFE) Specification

- **Why:** Standard in economics; well-understood properties
- **Limitation:** Recent research (Goodman-Bacon 2021) shows issues with staggered treatment
- **Mitigation:** Our data has single treatment time (no staggering), so TWFE is appropriate

- **Future extension:** Could add Callaway-Sant'Anna (2021) estimator for advanced students

Decision 2: Clustered Standard Errors by Unit

- **Why:** Panel data has serial correlation within units; OLS SEs are too small
- **Teaching moment:** Explains why "just adding time dummies" isn't enough
- **Implementation:** cov_type='cluster', cov_kwds={'groups': unit_id}

Decision 3: Event Study with Separate Regressions

- **Why:** Avoids patsy formula parsing issues with negative time indicators
- **Tradeoff:** Slower (multiple regressions), but more transparent
- **Educational value:** Students see exactly what each coefficient represents

Code Design:

```
class DifferenceInDifferences:
  def plot_parallel_trends(self):
    # Visual check

  def estimate(self, cluster_se=True):
    # Main DiD regression

  def event_study(self):
    # Dynamic effects

  def placebo_test(self):
    # Robustness check
```

Pedagogical rationale: Sequential workflow guides students through the full DiD checklist: (1) check assumptions, (2) estimate, (3) validate.

3.3 Data Generation Strategy

Synthetic Data Advantages:

1. **Ground truth known:** Can validate methods against true ATE
2. **Controlled confounding:** Can dial up/down selection bias
3. **No privacy concerns:** Can share openly on GitHub
4. **Reproducible:** Seed ensures everyone gets same results

Realism Built In:

- Propensity scores based on realistic covariates (size, industry, location, prior funding)
- Treatment effect heterogeneity (varies by baseline characteristics)
- Measurement noise and missing data challenges
- Panel structure mimics quarterly reporting

Educational Tradeoff: Real data would be more engaging, but synthetic data allows:

- Clean demonstration of concepts without confusing real-world messiness
- Validation against known truth (critical for learning)
- Legal/ethical simplicity (no IRB, no data agreements)

3.4 Performance Considerations

Computational Efficiency:

- PSM: $O(N^2)$ for naive nearest neighbor; optimized with KD-trees (not implemented to maintain simplicity)
- DiD: Linear in panel length; clustered SEs add overhead
- Expected runtime: <10 seconds for 1000 companies (PSM), <5 seconds for 4000 observations (DiD)

Scalability:

- Current implementation handles datasets up to ~10K observations comfortably
- For production use (100K+ observations), would need:
    - Approximate nearest neighbors (Annoy, FAISS)
    - Sparse matrix operations for DiD
    - Parallel processing for bootstrapping

Educational Consideration: We prioritize clarity over speed. Students should understand the algorithm before optimizing it.

3.5 Edge Cases and Limitations

PSM Edge Cases:

1. **No common support:** Some treated units have PS > max(control PS)
    - **Handling:** Warning message; those units excluded from analysis
    - **Teaching point:** External validity concerns
2. **Perfect separation:** Some covariate perfectly predicts treatment
    - **Handling:** Logistic regression fails; use penalized regression (L2)
    - **Teaching point:** When to worry about multicollinearity
3. **Ties in matching:** Multiple controls equidistant from treated unit
    - **Handling:** Break ties randomly (with seed for reproducibility)
    - **Teaching point:** Sensitivity to small perturbations

DiD Edge Cases:

1. **Unbalanced panel:** Some units missing time periods
    - **Handling:** Code checks for balance; warns if gaps detected
    - **Teaching point:** Attrition bias
2. **Treatment reversal:** Some units "un-treat"
    - **Handling:** Not implemented (assumes permanent treatment)
    - **Teaching point:** Limitations of standard DiD
3. **Staggered treatment:** Units treated at different times
    - **Handling:** Current code assumes single treatment time
    - **Extension opportunity:** Implement stacked DiD

Documentation Strategy: Every limitation is explicitly documented with suggestions for when alternative methods are needed.

---

4. Assessment & Effectiveness (Measuring Learning)

4.1 Assessment Design Philosophy

Assessment in this module follows **authentic assessment principles**: students demonstrate mastery by performing tasks that mirror real-world data science practice, not by recalling definitions or formulas.

Three-Tier Assessment Structure:

1. **Formative Assessment (During Learning):**
    - Interactive code checks (does their propensity score estimation work?)
    - Diagnostic plot interpretation (can they spot violated parallel trends?)
    - Immediate feedback loops
2. **Summative Assessment (After Learning):**
    - Complete analysis on new dataset
    - Written interpretation of results
    - Critique of a flawed causal study
3. **Transfer Assessment (Application to New Domains):**
    - Design causal analysis for their own research question
    - Peer review of classmates' analyses

4.2 Concrete Assessment Methods

4.2.1 Knowledge Checks (Formative)

Example Quiz Questions:

Question 1 (Conceptual Understanding):

"Company A received funding and grew 30%. Company B did not receive funding and grew 10%. Can we conclude funding caused 20% growth? Why or why not?"

Expected Answer: No, because we're comparing different companies (selection bias). We need to estimate what Company A's growth would have been WITHOUT funding (counterfactual).

Question 2 (Assumption Identification):

"You want to use PSM to estimate the effect of MBA education on salary. What assumption must hold? Give an example of a potential violation."

Expected Answer: Unconfoundedness—all confounders must be observed. Violation example: innate ability (unmeasured) affects both MBA enrollment and salary.

Question 3 (Method Selection):

"You have data on 100 companies, measured once, some of which received funding. Should you use PSM or DiD?"

Expected Answer: PSM (cross-sectional data, no time dimension for DiD).

4.2.2 Coding Exercises (Formative)

Exercise 1: PSM Implementation (Guided)

```
# Dataset: healthcare_intervention.csv

# Treatment: Received intensive care management (binary)

# Outcome: Hospital readmission within 30 days (binary)

# Covariates: age, prior_conditions, insurance_type, hospital_quality


# TODO 1: Estimate propensity scores

# Hint: Use LogisticRegression with covariates


# TODO 2: Match with caliper=0.05

# Hint: Use the PropensityScoreMatcher class


# TODO 3: Check balance

# Expected: All SMDs < 0.1 after matching


# TODO 4: Estimate ATT

# Expected: ATT ≈ -0.08 (8% reduction in readmissions)
```

Grading Rubric:

- Correct PS estimation (20%)
- Appropriate caliper choice and justification (20%)

- Balance diagnostic plots (20%)
- ATT calculation (20%)
- Interpretation in context (20%)

Exercise 2: DiD Analysis (Independent)

Dataset: minimum_wage_panel.csv

Treatment: State increased minimum wage

Outcome: Teen employment rate

Panel: 50 states × 10 years

Your task: Complete DiD analysis

1. Plot parallel trends

2. Test pre-trends statistically

3. Estimate DiD effect

4. Conduct event study

5. Interpret: Is the effect causal?

Success Criteria:

- Identifies parallel trends violation (if present)
- Correctly interprets DiD coefficient
- Discusses policy implications

4.2.3 Case Analysis (Summative)

Assignment: Critique a Flawed Study

Students receive a published blog post claiming:

"Our analysis shows that remote work CAUSES 15% productivity increase. We compared 500 remote workers to 500 office workers and found remote workers completed 15% more tasks."

Task: Write a 2-page critique identifying:

1. What causal claim is being made?

2. What is the counterfactual?
3. What confounders might exist?
4. What method would you use to estimate the causal effect?
5. What data would you need?

Grading Rubric:

- Identifies selection bias (remote workers may be more motivated)
- Recognizes missing counterfactual (same workers under both conditions)
- Proposes appropriate method (PSM if cross-sectional, DiD if panel)
- Discusses feasibility and limitations

4.3 Common Student Challenges & Mitigation Strategies

Challenge 1: "Why can't I just control for confounders in regression?"

Why this is hard: Students come from ML backgrounds where "add more features" is the solution. They struggle to understand that regression ≠ causal inference.

Mitigation Strategy:

- Show Simpson's Paradox example (confounders reverse sign of effect)
- Demonstrate "bad controls" (colliders, mediators)
- Emphasize that regression assumes no omitted confounders—strong assumption!

Assessment Check: Ask students to draw a causal graph (DAG) and identify which variables to control for.

Challenge 2: "My parallel trends plot looks parallel, but the test fails"

Why this is hard: Visual inspection is subjective; students over-trust their eyes.

Mitigation Strategy:

- Teach both visual and statistical tests
- Show examples where trends "look" parallel but aren't statistically
- Discuss power (small samples may fail to detect violations)

Assessment Check: Give students ambiguous parallel trends plot; ask them to justify their decision with both visual and statistical evidence.

Challenge 3: "I got a significant result, so funding must cause hiring, right?"

Why this is hard: Students conflate statistical significance with causal validity.

Mitigation Strategy:

- Emphasize that significance ≠ causation
- Teach sensitivity analysis (Rosenbaum bounds)
- Discuss researcher degrees of freedom

Assessment Check: Ask students to list 3 reasons a significant PSM result might NOT be causal.

Challenge 4: "How do I know if my assumptions hold?"

Why this is hard: Assumptions are untestable (by definition—if testable, they'd be verifiable facts).

Mitigation Strategy:

- Teach indirect tests (placebo tests, falsification tests)
- Emphasize robustness checks across specifications
- Discuss "sensitivity to unmeasured confounding"

Assessment Check: Students must conduct at least 2 robustness checks in their final project.

4.4 Addressing Different Learning Styles

Visual Learners:

- Extensive use of plots (parallel trends, propensity score distributions, balance plots)
- Causal graphs (DAGs) to show relationships
- Animated GIFs showing matching process (future enhancement)

Hands-On Learners:

- Jupyter notebooks with executable code
- Exercises where students modify parameters and see effects
- "Break it and fix it" challenges (intentionally violate assumptions, then diagnose)

Reading/Writing Learners:

- Comprehensive written tutorial with theory
- Blog post-style explanations of concepts
- Writing assignments for interpretation

Auditory Learners:

- 10-minute video walkthrough (Show-and-Tell)
- Verbal explanations in video
- Encourage study groups for discussion

Analytical Learners:

- Mathematical derivations in appendices
- Links to original papers for deeper dive
- Challenge problems requiring proof-based thinking

4.5 Validation of Learning Outcomes

How I Know Students Learned:

Short-term indicators:

- Exercise completion rate >90%
- Average quiz score >80%
- Ability to debug their own code without instructor help

Medium-term indicators:

- Successful completion of final project with novel dataset
- Peer reviews show understanding of common mistakes
- Students can explain concepts to each other (Feynman technique)

Long-term indicators (ideal, but outside project scope):

- Students apply methods to their thesis/capstone project
- Students cite causal inference in job interviews as a key skill
- Alumni report using these methods in industry

Measurement Tools:

- Pre/post knowledge tests (10 questions on causal concepts)
- Code review rubrics (did they implement correctly?)
- Concept maps (draw relationships between PSM, DiD, assumptions)

4.6 Future Improvements

Based on anticipated student feedback:

1. **More real-world examples:** Add case studies from different domains (healthcare, education, marketing)
2. **Interactive visualizations:** Build Streamlit/Dash app where students can adjust parameters and see effects in real-time
3. **Advanced methods:** Add instrumental variables, regression discontinuity, synthetic control
4. **Industry guest speakers:** Invite data scientists who use causal inference to share experiences
5. **Collaborative projects:** Students work in teams to analyze different sectors, then present findings
6. **Assessment analytics:** Track which questions students struggle with most; refine those sections
7. **Personalized learning paths:** Adaptive difficulty based on student performance (basic vs. advanced track)

---

5. Reflection & Conclusion

5.1 What Makes This Module Effective?

Three Core Strengths:

1. **Authentic Context:** The "80 Days to Stay" framing isn't just motivational—it's real. International students' futures depend on finding employment, and this analysis directly informs platform design. Students feel the weight of getting the answer right.
2. **Transparency of Uncertainty:** Unlike typical tutorials that present methods as foolproof, this module emphasizes limitations, assumptions, and sensitivity.

Students learn that causal inference is fundamentally about making transparent, defensible assumptions—not achieving certainty.

3. **Iterative Validation:** By providing both PSM and DiD, students see that causal inference isn't a single test—it's a cumulative case built through multiple methods, robustness checks, and logical argument.

5.2 Alignment with Data Science Pedagogy Best Practices

This module aligns with NSF-funded research on effective data science education:

- **Problem-driven learning** (Adhikari et al., 2021): Start with question, not method
- **Computational literacy** (Wing, 2006): Emphasize algorithmic thinking, not just running code
- **Ethical reasoning** (Saltz et al., 2019): Discuss real-world implications of causal claims
- **Reproducibility** (Peng, 2011): All code, data, and analyses fully reproducible

5.3 Limitations & Future Work

Current Limitations:

1. **Synthetic Data:** While pedagogically useful, lacks messiness of real data (missing values, measurement error, ambiguous treatment definitions)
2. **Single Context:** Startup hiring is one domain; students may struggle to transfer to healthcare, education, or policy contexts
3. **Simplified Assumptions:** Real-world causal inference often requires instrumental variables, or dealing with noncompliance, which this module doesn't cover

Future Enhancements:

1. **Real Data Module:** Add optional "advanced" section using actual SEC filings and LinkedIn data (with appropriate privacy protections)
2. **Causal Discovery:** Introduce DAGs and structural causal models (Pearl, 2009)
3. **Machine Learning Integration:** Show how causal forests, double machine learning can improve upon traditional methods
4. **Bayesian Perspective:** Incorporate uncertainty quantification via posterior distributions

5.4 Contribution to Course Goals

This project directly addresses INFO 7390's focus on **advanced data architecture and responsible AI deployment**:

- **Architecture:** Students learn to structure causal inference pipelines (data → matching → validation → interpretation)
- **Responsible AI:** Emphasizes that AI predictions aren't always actionable—causal inference is needed for decision-making
- **Real-world impact:** Connects technical methods to human outcomes (visa sponsorship, employment)

5.5 Final Reflection

Teaching causal inference is challenging because it requires students to unlearn the "correlation is enough" mindset drilled into them by ML courses. This module succeeds by:

1. Making the stakes tangible (real students, real consequences)
2. Providing transparent, working code (demystifying "black box" econometrics)
3. Emphasizing critical thinking over mechanical application

The ultimate test of this module's success isn't whether students can run propensity_matching.py—it's whether they pause before claiming causation in their future work, ask "what's the counterfactual?", and design analyses that respect the fundamental impossibility of observing parallel universes.

Causal inference is hard because reality is complex. This module doesn't hide that complexity—it embraces it, and in doing so, prepares students for the messy, uncertain, high-stakes world of real data science.

References

Angrist, J. D., & Pischke, J. S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254-277.

Huntington-Klein, N. (2022). *The Effect: An Introduction to Research Design and Causality*. CRC Press.

Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. Wiley.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5), 688-701.

Appendices:

Appendix A: Full code listings with detailed comments
Appendix B: Sample quiz questions with answer keys
Appendix C: Exercise rubrics and grading guidelines