A diagram of a cause and effect

AI-generated content may be incorrect.

Here is a **full case study explanation** using **DoWhy**, based on your uploaded documentation, designed to help you answer questions during your **Amazon Applied Scientist II – Causal Learning, Pricing and Promotions** interview.

**🎯 Case Study: Does a $10 Echo Dot Discount Increase Purchase Rate?**

**💼 Business Context (Amazon)**

Amazon is running a **$10 promotional discount** on Echo Dots. You want to **estimate the causal effect** of this discount on the **likelihood of purchase**, accounting for confounding factors such as customer income, browsing history, and seasonal effects (e.g., Prime Day).

**🚀 Step-by-Step Causal Inference Process with DoWhy**

**Step 1: Model – Define the Causal Assumptions**

DoWhy requires you to **explicitly write down your assumptions** in the form of a **causal graph** (DAG). This helps formalize the problem and avoid hidden bias.

**Assumed Graph:**

season income browsing\_time

↘ ↘ ↘

discount ---------> purchase

* **discount** is our **treatment**
* **purchase** is our **outcome**
* **income**, **season**, **browsing\_time** are **observed confounders** — they influence both treatment and outcome.

**Why this matters:**

“We assume these confounders influence whether someone receives or responds to a discount, and also whether they purchase—so we need to control for them.”

**Code:**

model = CausalModel(

data=df,

treatment="discount",

outcome="purchase",

common\_causes=["income", "season", "browsing\_time"]

)

**Step 2: Identify – Check if Causal Effect is Identifiable**

DoWhy uses **causal logic (do-calculus)** to verify whether it is possible to estimate the causal effect of the treatment on the outcome, given the graph.

DoWhy applies the **backdoor criterion**:

* If we block all backdoor paths from treatment to outcome using observed variables, then the effect is identifiable.

**Intuition:**

“If we observe all relevant confounders, we can simulate a randomized experiment by conditioning on them.”

**Output:**

identified\_estimand = model.identify\_effect()

print(identified\_estimand)

You’ll get a statement like:

“Backdoor identified: discount ⟶ purchase conditioned on [income, season, browsing\_time]”

**Step 3: Estimate – Quantify the Effect**

With identification confirmed, we estimate the effect using a **statistical or ML method**.

**Example Methods:**

* **Linear regression**: Interpretable, fast.
* **Propensity score matching**: Nonparametric, better for overlap issues.
* **EconML DML**: When using complex ML models and high-dimensional covariates.

**Option A: Simple Linear Estimation**

estimate = model.estimate\_effect(

identified\_estimand,

method\_name="backdoor.linear\_regression"

)

**Interpretation**:

"Controlling for income, browsing time, and season, offering a $10 discount increases purchase probability by 8 percentage points."

**Option B: ML-based Estimation (EconML)**

estimate = model.estimate\_effect(

identified\_estimand,

method\_name="backdoor.econml.dml.DML"

)

This uses **double machine learning** to model:

* the treatment assignment (discount ~ income + season + browsing)
* the outcome (purchase ~ income + season + browsing)

Then, it estimates the causal effect **net of biases from these two stages**.

**Step 4: Refute – Validate the Estimate**

This is DoWhy’s **most unique feature**: it encourages you to **question your findings**.

**Purpose:**

* Simulate what would happen if your assumptions were wrong.
* Ensure robustness of your estimate.

**Common Refuters:**

| **Method** | **Logic** |
| --- | --- |
| **Placebo** | Replace the treatment with a random variable. If you still see an effect, something’s wrong. |
| **Random common cause** | Add a synthetic confounder. If the estimate changes, your model is fragile. |
| **Subset testing** | Run on a subgroup (e.g., Prime vs non-Prime). Should yield similar estimates. |
| **Bootstrap** | Run many times on different subsamples to assess stability. |

**Code Example:**

refutation = model.refute\_estimate(

identified\_estimand,

estimate,

method\_name="placebo\_treatment\_refuter"

)

print(refutation)

**Output Interpretation:**

“The placebo test shows no significant effect when using a fake treatment. This supports the robustness of our causal claim.”

**📊 Final Summary for Amazon Interview**

| **Step** | **What You Did** | **Value for Amazon** |
| --- | --- | --- |
| **Model** | Defined causal relationships using DAG | Transparent assumptions avoid bias |
| **Identify** | Applied backdoor criterion | Ensured valid estimation strategy |
| **Estimate** | Used regression / ML to get effect size | Quantified effect: 8% increase from $10 discount |
| **Refute** | Ran placebo and subset tests | Showed result is robust, trustworthy |

**🧠 Interview Response Example**

"In a recent project, I wanted to evaluate whether a $10 Echo Dot discount led to higher purchase rates. I used DoWhy to draw out the causal graph, identifying income, browsing time, and season as key confounders. I confirmed identifiability using the backdoor rule. Then, I estimated the effect using linear regression and validated the results with placebo refuters. The causal effect was about an 8% lift in purchases, and robustness tests suggested the effect was reliable. This process mirrors how I’d approach pricing experiments at Amazon—ensuring we don’t just find correlations, but causal evidence to support business decisions."