

Evaluation of Public Policies and Social Programs

Session 1: Course Overview & the Private–Social Distinction

Andrés Barrios Fernández

Monday, August 3, 2026

MA in Economics, Universidad de los Andes, Chile — Second Semester 2026

Today

1. Course overview and logistics
2. Private evaluation vs. social evaluation — and where it comes from theoretically
3. Why causal inference is the foundation of evaluation
4. Exercise and discussion: cash vs. in-kind transfers
5. What we learned today

Learning Objectives

By the end of today's session, you will be able to:

- State, precisely, what separates a private evaluation from a social evaluation of the same program
- Ground that distinction in Harberger's postulates for applied welfare economics
- Explain, using the potential-outcomes framework, why every welfare estimate presupposes a credible counterfactual
- Apply the selection-bias decomposition to the specific empirical challenge the course project poses
- Identify the paternalism, externality, and screening rationales that can make an in-kind transfer socially preferable to cash transfer.

Course Overview

Who I am, and how this course works

- **Instructor:** Andrés Barrios Fernández (abarriosf@uandes.cl)
- **Office hours:** Monday–Wednesday, 15:30–16:00 — open drop-in, no topic or appointment needed
- **Format:** Monday/Wednesday, 90 minutes, advanced course — no tutorials, no teaching assistants
- **Everything is built around one project:** evaluating whether a government agency or donor should fund a new school opened by Fundación Astoreca

What this course is for

You already know how to estimate a causal effect.

This course asks: given a credible causal estimate, should the program be pursued?

That requires tools you have *not* yet built: pricing non-market goods, incorporating externalities, deciding whose dollar counts more, and choosing the right discount rate.

Component	Weight
Paper Presentation 1	10%
Paper Presentation 2	10%
Project: First Presentation	10%
Project: Second Presentation	15%
Project: Final Presentation	20%
Project: Final Report	25%
Participation	10%

The class (8 students) is divided into 2 teams of 4, randomly assigned by the instructor. Paper presentations are done in pairs: a 15-minute presentation closes 8 sessions (4 in Sessions 3–8, 4 in Sessions 9–14); each student presents twice, with a different partner each time.

The project timeline

- **Sessions 2–3:** Traditional CBA toolkit (private appraisal, consumer surplus)
- **Session 4 (Wed, Aug 12):** Fundación Astoreca guest session, plus project/report expectations
- **Sept 14:** First Presentation (10%)
- **Oct 26:** Second Presentation (15%)
- **Nov 18:** Final Presentations (20%)
- **Final Report (25%)**

Plus two paper presentations (10% each, done in pairs) along the way.

Project teams

The class of 9 is divided into 2 teams (one of 4, one of 5):

Team 1

Raúl Lagomarsino

Pedro Vergara

Victoria Rauseo

Emilia Perez

Team 2

Javiera Hasbun

M. Emilia Neut

Emilia Welkner

Felix Gimeno

Paolo Vivanco

All four project milestones are graded at the team level.

Paper presentations: who, what, when (1 of 2)

Paper Presentation 1 — Sessions 5, 6, 7, 10

Date	Presenters	Paper
Mon, Aug 17 (S5)	Pedro Vergara & Felix Gimeno	Greenstone & Gallagher (2008), "Does Hazardous Waste Matter?" <i>QJE</i>
Mon, Aug 24 (S6)	Emilia Perez & R. Lago-marsino	Bailey & Goodman-Bacon (2015), "The War on Poverty's Experiment in Public Medicine," <i>AER</i>
Wed, Aug 26 (S7)	M. Emilia Neut & Victoria Rauseo	Black (1999), "Do Better Schools Matter?" <i>QJE</i>
Wed, Sept 16 (S10)	Emilia Welkner, Javiera Hasbun & Paolo Vivanco	Finkelstein, Hendren & Shepard (2019), "Subsidizing Health Insurance for Low-Income Adults," <i>AER</i>

Paper presentations: who, what, when (2 of 2)

Paper Presentation 2 — Sessions 11, 12, 13, 14

Date	Presenters	Paper
Mon, Sept 21 (S11)	M. Emilia Neut & Felix Gimeno	Angelucci & De Giorgi (2009), "Indirect Effects of an Aid Program," <i>AER</i>
Wed, Sept 23 (S12)	Emilia Perez & Javiera Hasbun	Chay & Greenstone (2003), "The Impact of Air Pollution on Infant Mortality," <i>QJE</i>
Mon, Sept 28 (S13)	Emilia Welkner & R. Lago-marsino	Deming (2011), "Better Schools, Less Crime?" <i>QJE</i>
Wed, Sept 30 (S14)	Pedro Vergara, Victoria Rauseo & Paolo Vivanco	Bobonis & Finan (2009), "Neighborhood Peer Effects in Secondary School Enrollment Decisions," <i>REStat</i>

No student presents with the same partner twice.

Today's other logistics

- Astoreca case packet distributed today — start reading it; we return to it before the guest session in Session 4
- Session 4 (Wed, Aug 12): Fundación Astoreca guest presentation, plus a detailed look at what the project and final report require
- Paper presentation assignments and pairings distributed today
- Problem Set 1 begins today (self-study; solutions posted; not collected)

Private vs. Social Evaluation

A motivating example

A firm decides whether to open a factory.

Its own evaluation asks: will revenue exceed cost?

That evaluation is, on its own terms, entirely correct — and entirely silent on:

- the pollution imposed on neighboring households
- the value of jobs created for previously unemployed workers
- the tax revenue the government collects from the new wages

None of this is a mistake in the firm's arithmetic. It is a difference in *whose* costs and benefits are being counted.

Does the policy work? Is it worth it?
For whom, and according to whom?

Two very different evaluations of the same program

Private evaluation

- Cash flows to *one* decision-maker
- Discounted at a *private* rate (that decision-maker's cost of capital)
- Example: does this pay off for Fundación Astoreca's own budget, or for a donor's narrow ROI?

Social evaluation

- Cash flows *and* non-market values to *everyone* affected
- Externalities, distributional weights
- Discounted at the *social* discount rate

Private vs. social evaluation, visually

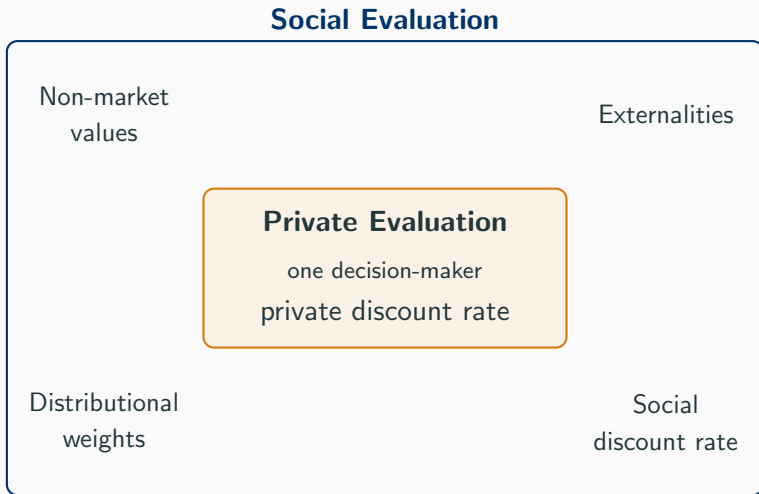


Diagram: author's construction, illustrating Harberger's (1971) postulate 2.

Grounding the distinction: Harberger's three postulates (1971)

Harberger, A.C. (1971), "Three Basic Postulates for Applied Welfare Economics," JEL

1. The competitive market price of a good measures its value at the margin, to both suppliers and demanders — surplus changes read off observed demand and supply curves are valid welfare measures
2. In evaluating the net benefits of a project, sum the gains and losses to *all* parties affected, with no regard to who gains and who loses
3. Costs and benefits are measured in a common numeraire (money), in present-value terms, appropriately discounted

Why these postulates matter for this course

- **Postulate 1** justifies almost everything we do with demand curves, WTP, and hedonic prices (Sessions 3–6)
- **Postulate 2** is exactly the assumption we will interrogate and relax once we introduce distributional weights (Session 8) — and it is the assumption a *private* evaluation does not even satisfy, because it never sums across all affected parties in the first place
- **Postulate 3** raises the question we spend an entire session on: *which* discount rate? (Session 12)

A social evaluation is, in this sense, an attempt to actually satisfy Harberger's postulates. A private evaluation, by construction, does not try to.

What a social evaluation adds, concretely

- Non-market values (no price tag on student achievement)
- Externalities and spillovers (effects on siblings, peers, the neighborhood)
- Distributional weights (does a dollar to a poor family count more?)
- The *social* discount rate, not a private cost of capital

The welfare metrics we'll build, over the semester

- Willingness to pay (WTP)
- Consumer surplus
- Compensating variation / equivalent variation (CV/EV)
- Marginal Value of Public Funds (MVPF)

You don't need to know what these mean yet — that's the semester.

Why Causal Inference Is the Foundation of Evaluation

Every welfare number is a comparison

WTP. A benefit. An MVPF. An earnings gain.

Every one of them is a comparison between **what happened** and **what would have happened otherwise**.

Without a credible counterfactual, there is no welfare estimate — only a description.

Formalizing the counterfactual (Rubin, 1974)

Rubin, D.B. (1974), “Estimating Causal Effects of Treatments...,” J. Educational Psychology

For each student i , define two *potential outcomes*:

- $Y_i(1)$: the outcome if student i attends Astoreca
- $Y_i(0)$: the outcome if student i does *not* attend Astoreca

The individual causal effect is $Y_i(1) - Y_i(0)$.

The fundamental problem of causal inference (Holland, 1986)

We observe only one potential outcome per student:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0), \quad D_i \in \{0, 1\} \text{ is Astoreca attendance}$$

We can never observe $Y_i(1)$ and $Y_i(0)$ for the *same* student at the *same* time.

So we replace the impossible individual comparison with a feasible comparison across *groups*.

The missing-data problem, visually

	$Y_i(1)$: attends Astoreca	$Y_i(0)$: does not attend
$D_i = 1$: attends Astoreca	Observed	Missing (counterfactual)
$D_i = 0$: does not attend	Missing (counterfactual)	Observed

Two of the four cells are, by definition, never observed for any student. Causal inference is a strategy for filling them in credibly.

From individuals to averages: the selection-bias decomposition

$$\underbrace{E[Y_i | D_i=1] - E[Y_i | D_i=0]}_{\text{naive comparison of means}} = \underbrace{E[Y_i(1) - Y_i(0) | D_i=1]}_{\text{ATT: causal effect on the treated}} + \underbrace{E[Y_i(0) | D_i=1] - E[Y_i(0) | D_i=0]}_{\text{selection bias}}$$

The naive comparison equals the causal effect *only if* the selection-bias term is zero.

Applying the decomposition to Astoreca

- Naive comparison: mean achievement of Astoreca students vs. the national average
- Selection bias term: $E[Y_i(0) \mid D_i=1] - E[Y_i(0) \mid D_i=0]$ — how would Astoreca's *actual* students have done elsewhere, relative to the average student elsewhere?
- Families who seek out and gain admission to Astoreca plausibly differ in motivation, resources, or baseline ability

Likely direction

The selection-bias term is probably **positive** — so the naive comparison likely *overstates* the school's causal effect. This is exactly why a credible research design matters.

The credibility revolution (Angrist & Pischke, 2010)

Angrist, J.D. & Pischke, J.S. (2010), JEP

Applied microeconomics shifted decisively toward research designs that make the selection-bias term small, known, or directly estimable — rather than toward more complicated statistical or structural machinery layered on top of an incredible comparison.

The toolkit: randomized experiments, and quasi-experiments exploiting instrumental variables, regression discontinuities, and differences-in-differences.

The toolkit you already have

Mastering 'Metrics (Angrist & Pischke) — not taught in this course, but assumed:

- Randomized experiments
- Regression / matching on observables
- Instrumental variables
- Regression discontinuity (e.g., an admissions or oversubscription cutoff)
- Differences-in-differences

Your project will require applying the right one of these to real data on Astoreca's students.

If your team simply compared the test scores of Astoreca's students to the national average, what would bias that comparison — and in which direction?

- What does that comparison implicitly assume about the selection-bias term?
- What comparison group would you look for instead — and what would make it credible?
- Is there a plausible source of quasi-random variation in who attends (a lottery, a capacity cutoff, a phased rollout)?

Once you have a credible **causal estimate** of the school's effect...

the rest of the semester is about what to do with it:

how to value it, whose perspective to value it from, and what discount rate to bring it back to the present with.

Exercise: Cash vs. In-Kind Transfer

The scenario

The government has a \$1000/person budget. Two options:

1. Give \$1000 cash to each recipient
2. Provide \$1000 worth of food vouchers (can only buy food)

Task (small groups, 15 minutes)

- Which option generates more benefit for recipients, and why?
- Draw the indifference-curve diagram showing the difference
- What if some recipients would have spent *less* than \$1000 on food anyway?
- When might in-kind be better?

Groups present: what did you find?

New vocabulary from your answers: **revealed preference**, **consumer surplus**, **paternalism**, **externalities**.

Note

A purely private (recipient-only) evaluation would miss the paternalism and externality arguments entirely.

Deep discussion: what does the theory — and the data — actually say?

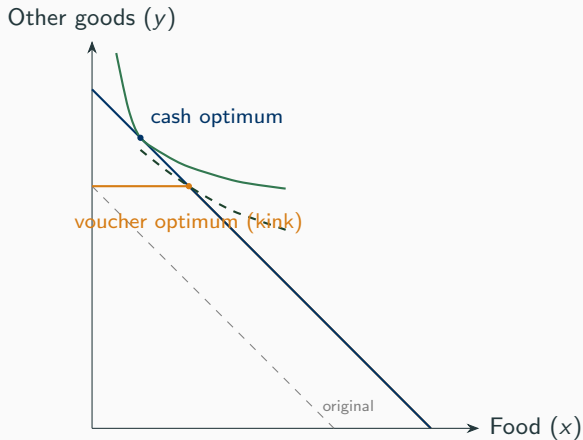
Currie, J. & Gahvari, F. (2008), JEL

Cash weakly dominates an equal-value in-kind transfer *when the recipient is unconstrained* by it — i.e., when they would have bought at least that much of the good anyway.

Three things can flip this result in favor of in-kind:

- **Paternalism** — the planner values the recipient's consumption of the good more than the recipient's own revealed preference (e.g., children's nutrition or schooling)
- **Externalities** — others benefit from the recipient's consumption of the good
- **Screening** — the requirement to consume the good self-selects the truly eligible population

The classic diagram: cash vs. voucher



A binding voucher forces a strictly lower indifference curve than equal-value cash.

Following Southworth (1945); surveyed in Currie & Gahvari (2008).

The data don't fully agree with the simple theory

Currie & Gahvari's survey of the empirical literature: recipients of in-kind transfers (e.g., US Food Stamps) do *not* always behave as if the transfer were fully fungible with cash.

- Consumption of the targeted good shifts *more* than the unconstrained model predicts
- Candidate explanations: binding liquidity constraints, mental accounting / reference dependence, or the in-kind transfer program itself changing information or norms

Theory gives us a benchmark; it does not settle the empirical question.

Is a school more like “food stamps,” or more like “cash with strings attached”?

- Which of paternalism, externalities, and screening apply here — and how strongly?
- Would an equivalent-value cash transfer to these families buy comparable educational gains?
- Your private-vs-social comparison should explicitly address *why in-kind, rather than cash*

What We Learned Today

- A private evaluation and a social evaluation of the same program can reach opposite conclusions — not because either is wrong on its own terms, but because they count different things
- Harberger's postulates give the theoretical foundation for (unweighted) social CBA — and tell us exactly which assumption distributional weighting (Session 8) will later relax
- Every welfare number presupposes a credible counterfactual; the potential-outcomes framework makes precise what “credible” means, and why naive comparisons are generally biased
- Cash weakly dominates in-kind transfers *unless* paternalism, externalities, or screening are present — and the empirical evidence is more complicated than the simple theory

Looking Ahead

Before Session 2

- Begin reading the Astoreca case packet (you'll return to it before the guest session in Session 4)
- Read Boardman et al., Chapters 3–4
- Begin Problem Set 1 (self-study)

See you Wednesday.