

Evaluation of Public Policies and Social Programs

General Equilibrium and External Validity

Andrés Barrios Fernández

Wednesday, October 21, 2026

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

1. What changes when a program affects a whole market, not only its participants?
2. What do the Medicare and cash-transfer cases teach us about market feedback?
3. Why might credible effects differ across places, people, implementers, and scale?
4. What can Astoreca's existing schools tell us about a proposed new school?

No in-class calculation or team exercise today; we will use evidence and open discussion to examine the assumptions behind a decision-ready forecast.

Learning Objectives

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

- Distinguish a partial-equilibrium effect on participants from wider market and welfare effects
- Explain how prices, supply, income, and crowding out can change the consequences of a large program
- Distinguish internal validity from the transport of an estimate to another population or setting
- Assess the relevance of Astoreca's existing-school evidence without relying on an arbitrary external-validity discount

From Participants to Markets

A program changes more than its participants' outcomes

A causal effect for participants is not yet the total effect of a policy.

Partial equilibrium

- Holds the surrounding market largely fixed
- Focuses on direct effects for treated people
- Often the right starting point for a small intervention

General equilibrium

- Allows prices, supply, and behavior to respond
- Includes effects on untreated people and institutions
- Becomes more important as reach and market share grow

The boundary between the two is a matter of degree, not a claim that one is always correct.

Four channels of market feedback

1. **Prices:** higher demand can raise wages, rents, or input prices; increased supply can lower prices.
2. **Supply and investment:** providers may enter, expand capacity, change quality, or adopt new technology.
3. **Income and local demand:** program spending can create further transactions and affect local production.
4. **Crowding out and reallocation:** public or subsidized provision can replace other spending, providers, or activity.

A useful first question

Which market is plausibly affected, how large is the intervention relative to it, and how quickly can that market respond?

A 5–10% market-share threshold can be a screening heuristic, not a universal cutoff or a substitute for a causal argument.

Two cases: why the market boundary matters

Health insurance and Medicare

Finkelstein (2007) studies the aggregate effects of Medicare's introduction. Expanding insurance can affect hospital spending and investment beyond the direct value of coverage to newly insured people.

Cash transfers and local markets

Egger et al. (2022) study cash transfers at scale. Transfers affect recipient households directly and can generate broader local economic activity, alongside possible price and supply responses.

- The market response is an empirical object, not something that can be inferred from the participant-level treatment effect alone.
- A multiplier or aggregate change depends on context, the comparison, and the market boundary used.

Finkelstein, A. (2007), "The Aggregate Effects of Health Insurance," *QJE*, 122(1); Egger, D. et al. (2022), "General Equilibrium Effects of Cash Transfers," *Econometrica*.

General Equilibrium in a School Decision

Which markets could a new school affect?

Potential resource-market responses

- Demand for qualified teachers and changes in local wages
- Recruitment from other schools and resulting staffing or quality changes
- Construction, land, transport, and other local input demand

Potential education-market responses

- Enrollment shifts and changes in nearby schools' capacity or composition
- Competition, imitation, or changes in other providers' quality
- Housing choices and capitalization of perceived school quality

These are plausible channels, not a checklist of effects that must all be monetized.

When should a feedback enter the evaluation?

- Is there a credible mechanism linking the school to an affected market?
- Is the effect large relative to direct benefits, resource costs, or the decision margin?
- Would the effect be additional, or would it mainly reallocate activity from another school or employer?
- Can it be measured well enough to improve the decision, rather than add speculative precision?

Interpretation

Model important, decision-relevant feedbacks. Be explicit about uncertain or second-order channels, and do not count the same gain twice.

From a Causal Estimate to a Forecast

A credible effect can still be the wrong forecast

Suppose a study credibly finds that a school model raised achievement by 0.20 standard deviations.

- **Internal validity:** did the program cause that effect for the study population, place, and time?
- **External validity:** what effect should we expect for a different population, location, implementation team, or scale?
- The first question concerns the counterfactual in the study. The second concerns transport from the study setting to the decision setting.

A causal estimate is evidence for a forecast; it is not automatically the forecast itself.

What the cross-site evidence cautions us about

Vivalt (2020), “How Much Can We Generalize from Impact Evaluations?”

Impact estimates vary across contexts. A study's effect is not a universal constant that can simply be copied into every new setting.

- Observed variation can reflect real differences in populations, institutions, and implementation, as well as sampling and measurement uncertainty.
- Publication and reporting processes can make the available evidence unrepresentative of all attempted evaluations.
- Differences across implementer types are a reason to investigate who delivered the program and under what conditions, not a formula for shrinking every estimate.

Vivalt, E. (2020), *Journal of Development Economics*, 147, 102485.

How effects can change as programs scale

1. **General-equilibrium effects:** expansion changes wages, prices, competition, or the availability of inputs.
2. **Implementation quality:** pilot attention, expert staff, and close monitoring may not survive routine delivery at scale.
3. **Population heterogeneity:** later participants may have different needs, constraints, or alternatives.
4. **Political-economy effects:** expansion can alter incentives, targeting, budgets, accountability, or support.

The central question

Does expansion merely deliver the same intervention to more people, or does it change the environment and the people the intervention reaches?

Muralidharan, K. & Niehaus, P. (2017), "Experimentation at Scale," *Journal of Economic Perspectives*, 31(4), 103–124.

What might travel from Astoreca's existing schools?

Potentially stable features

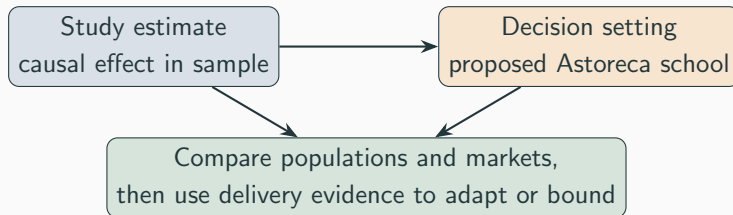
- Educational model, curriculum, and instructional routines
- Teacher training and accumulated organizational know-how
- Similar student ages and outcome measures

Potentially changing features

- Students' baseline achievement, needs, and alternatives
- Neighborhood, peer environment, and nearby schools
- Leadership, staff recruitment, start-up capacity, and scale

Similarity matters only insofar as it bears on the mechanisms that produce outcomes.

A transportability map



- Transfer directly only when relevant features and mechanisms are sufficiently similar.
- Adjust when differences and credible effect moderators provide evidence about direction and magnitude.
- When they do not, report a range and make the assumptions behind it visible.

Discussion and Implications

Questions for a decision-maker

- Which evidence from Astoreca's existing schools would most change your confidence in a new school's projected effect?
- What would make a difference in student composition consequential rather than merely descriptive?
- If teacher wages rise because the school recruits scarce teachers, is that a project cost, a transfer, or both under the chosen perspective?
- Which apparent benefit might actually be a gain shifted from another school, and what evidence would distinguish the two?

The point is not to produce one mechanically adjusted number, but to expose the assumptions on which the recommendation rests.

From evidence to a defensible recommendation

- **Keep the estimands distinct:** participant impact, market-wide response, and the forecast for this decision are related but not interchangeable.
- **Use evidence to adjust:** avoid an arbitrary “external-validity discount” or speculative multiplier.
- **Represent uncertainty honestly:** use a range when transport assumptions are consequential but weakly identified.
- **Learn over time:** phased implementation and early cohorts can update a forecast, while not prematurely claiming definitive long-run evidence.

A good evaluation explains why the evidence travels, where it may fail, and what would change the decision.

What We Learned Today

- General-equilibrium analysis adds market responses and effects on people beyond direct participants.
- Prices, supply, income, and reallocation can change the size and distribution of a program's consequences.
- Internal validity establishes a causal effect in a study setting; external validity concerns its relevance to a new setting.
- Astoreca's track record is valuable evidence, but its relevance depends on the people, place, delivery, scale, and market responses involved.
- Adjust only when evidence supports the adjustment; otherwise show a transparent range and identify the uncertainty that matters.

- Finalize your analysis and connect it to the complete private-versus-social evaluation.
- Bring your headline estimate, key assumptions, and largest remaining uncertainty.