

EGG-TIMER SESSION · ENVIRONMENTAL VALUATION AND CHOICE EXPERIMENTS

Accounting for Substitutes in Single-Site Travel Cost Models with Limited Data

Limited data doesn't mean ignoring substitution!

Katarzyna Skrzypek · Mikołaj Czajkowski · Marek Giergiczny

University of Warsaw

WCERE 2026

This research is supported by the National Science Centre in Poland (2024/53/N/HS4/04303; 2021/43/B/HS4/03371; 2018/30/E/HS4/00388)

THE PROBLEM

Substitute treatment — a quiet choice that's easy to get wrong

Single-site travel cost models are the workhorse of applied recreation valuation. But recreation sites have **substitutes** — and ignoring them **overstates welfare**.

The usual excuse: *“we don’t have multi-site data.”*

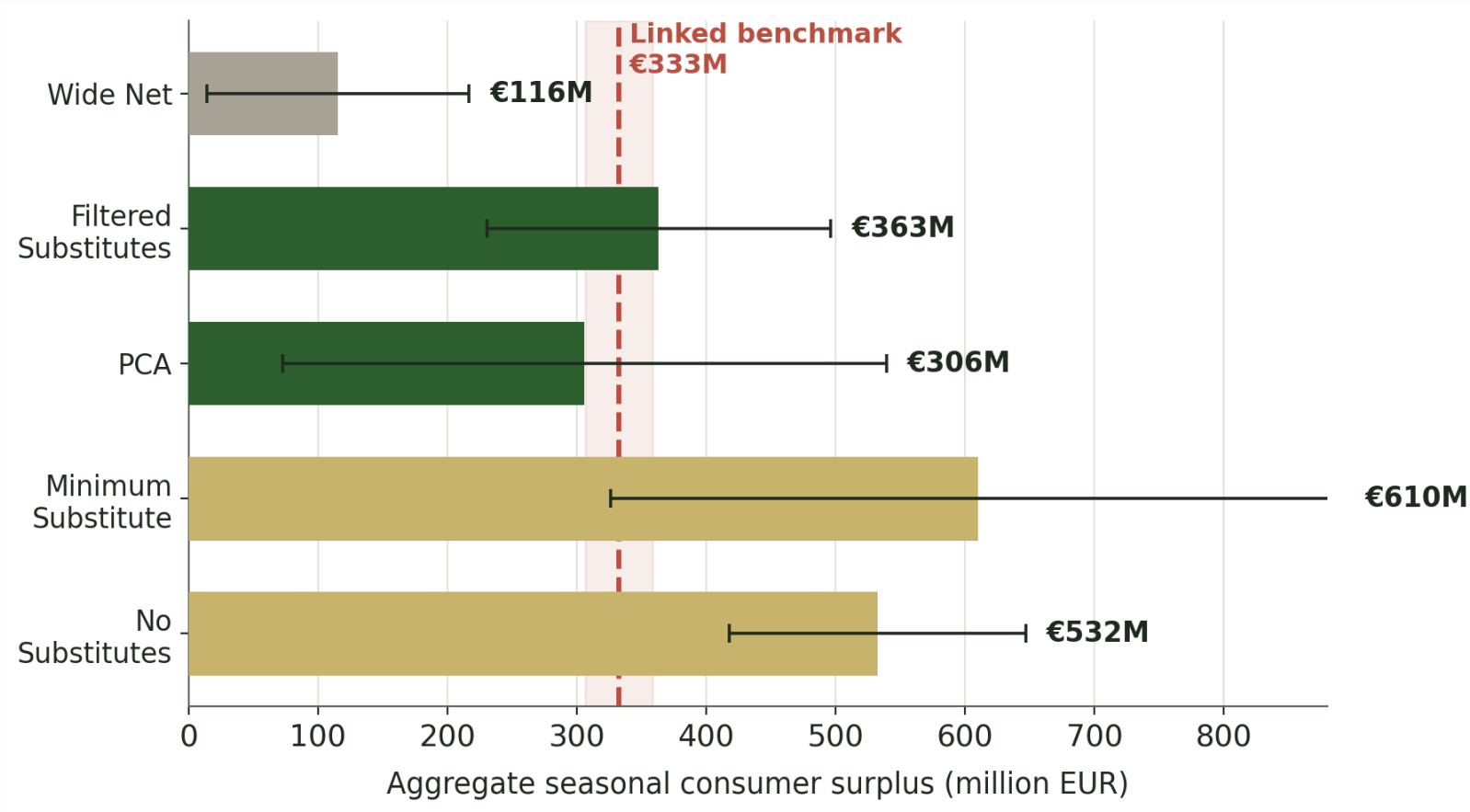
So substitution is either **ignored** or patched with *ad hoc* controls that are rarely tested.

OUR CLAIM

You don’t need multi-site data to handle substitution credibly.

Two simple, constrained controls recover the welfare estimate of a full utility-consistent model — using only single-site data plus travel costs you can build after the fact.

Same park, same data — welfare swings 5×



THE HEADLINE

€116M → €610M

for the recreational value of one national park — driven only by how substitutes are handled.

Omit / nearest-only → too high

Wide net → unstable, too low

Our controls → on benchmark

Benchmark (full model): €333M

Both common fixes fail — in opposite directions

↑ overstates

Ignore substitutes

Displaced trips look like lost trips. Welfare biased upward.
Numerically stable — which gives false confidence.

× breaks down

Cast a wide net

Substitutes cluster in space → travel costs are highly collinear → unstable estimates. Low welfare is an artefact, not caution.

The challenge isn't whether to include substitutes — it's how to include them without breaking identification.

Two constrained controls keep the signal, drop the noise

FS

Filtered substitutes

Screen out travel-cost variables that are highly correlated (>0.9) with the focal site or each other. What remains still maps to real, named parks — so it stays interpretable.

PCA

Principal-component indices

Replace the whole correlated bundle of substitute costs with a few orthogonal indices (97.5% of variance). Collinearity gone; maximum statistical stability.

Both reduce dimensionality while preserving welfare-relevant substitution. Build the substitute costs ex post — Google Distance Matrix / openrouteservice.

Validated twice — on real data and in simulation

EMPIRICAL · TATRA NATIONAL PARK

Benchmarked against a full linked site-choice/count model over all 23 Polish parks (n = 5,855).

Benchmark **€333M**
 Filtered subs. **€363M**
 PCA indices **€306M**

Constrained controls land on the benchmark; omit & nearest-only overshoot.

MONTE CARLO · 1,000 REPLICATIONS

Data generated from a known substitution structure — so the true welfare is known.

	WN	FS	PCA	MS	NS
Closest to benchmark	12.9%	25.7%	52.5%	8.9%	0%
Convergence rate	13.7%	100%	98.8%	100%	100%
Mean error	-75%	+32%	+10%	+64%	+344%

PCA most accurate and reliably stable; common fixes are biased or won't converge.

TAKEAWAY

If you run single-site travel cost models, diagnose your substitute correlations — and use a constrained control.

PCA for stability · Filtered substitutes for interpretability · Both beat omitting or wide-net

Want the details? Full paper, simulation design, and code-ready guidance — happy to share or collaborate.

Katarzyna Skrzypek
k.skrzypek5@uw.edu.pl

University of Warsaw