Student Flow Modeling for School Decongestion via Stochastic Gravity Estimation and Constrained Spatial Allocation

Sebastian Felipe R. Bundoc[1,*], Paula Joy B. Martinez[1,*], Sebastian C. Ibañez[1], and Erika Fille T. Legara[1,2]

[1]Center for AI Research, Department of Education, Philippines
[2]Asian Institute of Management, Makati, Philippines
[*]These authors contributed equally to this work.

Keywords: stochastic gravity modeling, urban mobility, spatial interaction networks, constrained resource allocation, computational policy simulation

Extended Abstract

Motivation. School congestion, where enrollment exceeds capacity, affects education quality across lower- and middle-income countries (LMICs). While public-private partnerships (PPPs) subsidize student transfers to private schools, fragmented data systems have historically prevented a unified analysis of spatial mobility patterns that determine program effectiveness. Our work addresses three primary challenges: (1) estimating behavioral elasticities from a heteroskedastic, overdispersed flow data; (2) predicting flows for unobserved origin-destination (OD) pairs where most historical school pairings have zero historical flows; (3) reconciling behavioral choice models with system-wide capacity constraints. We integrate behavioral estimation with a capacity-constrained allocation mechanism, enabling policy simulations that balance demand-side incentives with supply-side limits.

Approach and Methodology. We unified 2.7M student transitions across 8,000 Philippine schools with geolocations, costs, subsidies, road distances (171M pairs), and socioeconomic indicators, yielding 29,224 observed flows. We move beyond assignment algorithms with fixed preferences [1] and model student flow $E(F_{ij})$ using a behavioral estimation framework:

$$E(F_{ij}) = \exp(\beta_0 + \alpha_3 \ln d_{ij} + \alpha_4 \ln c_{ij} + \gamma_i + \delta_j + \beta_1 \ln I_i + \beta_2 \ln I_j + \beta_3 Q_j) \tag{1}$$

where d_{ij} is road distance, c_{ij} is net cost (tuition minus subsidy), γ_i and δ_j are regional fixed effects, I_i and I_j are origin and destination municipality income, and Q_j is school quality rating.

We employ Negative Binomial (NB) Regression to address significant overdispersion ($\alpha = 0.393$) and the high frequency of zero-flow pairs. This specification is preferred over a restrictive equidispersion Poisson baseline or OLS. Using estimated elasticities, we predict flows under five counterfactual subsidy scenarios ranging from ₱1,000 to ₱20,000, populating potential origin-destination pairs via a lexicographic search. These predictions are reconciled with physical constraints using a doubly constrained allocation: flow a_{ij} for pair (i, j) is accepted as $\min(e_i, k_j, \hat{y}_{ij})$, where e_i is the remaining origin pool and k_j is destination capacity.

Results. Distance is the primary determinant of student mobility, with elasticity nearly four times more influential than that of net cost. This implies that doubling travel distance reduces expected enrollment by 27%, whereas doubling net tuition costs results in only a 7.4% decrease. Subsidies alone have limited efficacy for families lacking access to nearby private schools, suggesting that the strategic geographic expansion of participating institutions is a more potent decongestion lever than increasing subsidy amounts.

Table 2 reveals that the ESC system is primarily limited by a supply-side ceiling rather than price sensitivity. While the model identifies a 34.7% gap between observed and potential enrollment, increasing the subsidy from ₱1,000 to ₱20,000 yields a marginal flow increase of only 1.8%. Although 98.1% of untapped capacity is geographically aligned with "congested-feeding" origins, over half (53.9%) of this potential lies in unobserved counterfactual pathways requiring activation. Ultimately, while full slot utilization would absorb ~6% of total congested flow, the findings again indicate that the primary policy lever for decongestion is the strategic expansion and spatial reallocation of school places rather than demand-side cost reductions.

Table 1: Key Behavioral Elasticities (Negative Binomial Regression)

Variable	**Coefficient**	**p-value**
ln(distance)	-0.451	< 0.001
ln(net cost)	-0.100	< 0.001
Observations: 29,224 α: 0.393		

Note: Net Cost = Tuition - Subsidy. Clustered standard errors at origin school level. Distance and cost are log-transformed. Full results include origin/destination income, school quality, and regional fixed effects.

Table 2: Doubly Constrained Simulation Results

Scenario	**Subsidy Increase**	**Predicted Flow**	**Δ from Observed**	**Δ from −1K**
Observed	—	74,232	—	—
−1K	₱1,000	99,992	+34.7%	—
−5K	₱5,000	100,442	+35.3%	+0.4%
−10K	₱10,000	100,944	+36.0%	+1.0%
−15K	₱15,000	101,428	+36.6%	+1.4%
−20K	₱20,000	101,818	+37.2%	+1.8%

Note: Observed baseline is total ESC beneficiaries from student flow records. All scenarios use the same slot capacity K_j and candidate pool E_i.

Conclusions and Outlook. Distance is the primary determinant for school choice, even after accounting for net enrollment costs, institutional quality ratings, and regional socioeconomic factors. Our framework enables national-scale policy simulations of student mobility and decongestion potential. Further work will study dynamic slot reallocation based on current subsidy levels and optimize capacity by congestion severity.

References

[1] Hongzhao Guan, Nabeel Gillani, Tyler Simko, Jasmine Mangat, and Pascal Van Hentenryck. Contextual stochastic optimization for school desegregation policymaking. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.