Robust estimation method for the economic subsidies to educational institutions in Chile

Valentina Ugalde, Sebastián Dávila-Gálvez, Óscar C. Vásquez

University of Santiago of Chile (USACH), Faculty of Engineering, Program for the Development of Sustainable Production Systems (PDSPS), Chile, University of Santiago of Chile (USACH), Faculty of Engineering, Industrial Engineering Department, Chile.

Abstract

In Chile, the Ministry of Education of Chile (MINEDUC) provides resources to the various educational institutions in the country based on a budget estimation model, where the "attendance of students" is the main component to be considered by annual economic subsidies. In the process, this estimation must be computed in the middle of the year, and therefore, the reduction of the impact of error in the estimation face to uncertainty is a desirable goal. In this work, we propose an easy-to-implemented robust estimation method, which is based on the formulation, implementation, and integration of a set of mathematical models that target the minimization of the maximum estimation error. The obtained results show that the proposal obtains a better estimation than the current MINEDUC estimation model, providing an efficient alternative of robust estimation method for the decision makers in the process, exposing insights to be considered in the future.

Keywords: School Subsidies; attendance, mathematical models; robust method; estimation.

1. Introduction

Education is the fundamental basis for people's opportunities throughout their lives and constitutes the essential underpinning of a country's productivity and development (Marcel & Tokman, 2005). Therefore, States must finance school education considering its compulsory nature and guarantee access and equal participation of all members of the community in national life. In this context, the Ministry of Education of Chile (MINEDUC) is responsible for ensuring an inclusive and quality education system that contributes to the comprehensive and lifelong learning of people and the development of the country, through the formulation and implementation of policies, standards, and regulations, from kindergarten to higher education (Gobierno de Chile, 2022). For this purpose, MINEDUC grants municipal and subsidized private education and local education services the concept of school subsidy.

The schooling subsidy granted by MINEDUC finances an average of 10.595 educational institutions in Chile, which translates into an average of 3.269.831 students in the country. Moreover, concerning the distribution of financing resources for the educational system, in 2021, the Chilean government allocated 6.396.670.457 MM CLP, of which 3.670.242.048 MM CLP are destined for the schooling subsidy (MINEDUC, 2021), representing more than 50% of the annual budget ministry.

In this context, MINEDUC has a budget estimation model that is conducted in collaboration with the Ministry of Finance and is composed of different variables and parameters. The main assumptions are enrollment and attendance percentage, which are determined by analyzing the behavior of recent years. For example, the methodology used in the Ministry of Education's 2023 subsidy estimation considered the most recent enrollment at the time of the estimation (May enrollment) and a 2,48% increase in the number of PIE students (students with permanent or temporary special educational needs) according to the growth reported by the General Education Division (School Integration Program Coordination). To make the estimation, the percentage of occupied attendance corresponding to the average of the last four years was used.

However, the attendance estimation for determining the budget to be requested for the school subsidy delivery needs to be more explicit due to various situations, including social mobilizations and Covid-19, among others. On the other hand, MINEDUC must annually make the budget estimation for the following year in May. That is, with the information from January to May 2023, they must estimate the attendance for all the months of 2024. That generates uncertainty about the value of estimated attendance, causing a margin of error in the estimation, either underestimating or overestimating the actual budget. If the calculated

budget is lower than that needed in the subsidized educational institutions, MINEDUC must again request an amount to subsidize the educational institutions that should have been covered.

1.1. Our contribution

This paper deals specifically with the problem of uncertainty in the number of students attending educational institutions to determine the amount of school subsidy to be received. The contribution of this research is based on a robust estimation method, which involves the formulation, development, and implementation of a set of mathematical models that target the estimation error. This method allows determining a more efficient estimation of the number of students attending educational institutions per month, and per teaching modality, complying with the MINEDUC's request, that is, that the estimation values have at most a 2,5% error. This result will, in turn, be used as a parameter in the financing formula of the Chilean educational system and will reduce the margin of error in estimating of the total budget for educational financing. The data corresponding to Chile's municipal and private educational institutions and local education services were considered for the above.

2. Material and methods

2.1. Statement of the problem

In Chile, schools are characterized according to their administration, for which we consider a set of educational institutions D. Each educational institution can offer a set of days for holding classes, so we consider a set of days J for educational institutions. In turn, each educational institution has different teaching modalities that we denote by a set P that depends on the type of day J offered by that institution. Then, we define $P_c \subset P$ as the subset of teaching modalities for educational institutions with a full school day and a subset $P_s \subseteq P$ as the subset without a full school day. Student attendance is recorded from January to December for each year. We define a set of months M and a set of years A that depends on the collected data that considers information from educational institutions from 2017 to 2021. As MINEDUC annually must make the budget estimation in May for the attendance estimation, we consequently consider a subset $M_t \subset M$ representing the training months to be used for the estimation.

To estimate attendance, we rely on estimation errors. On the one hand, the mean absolute percentage error (MAPE), which relates the estimation error to the level of demand, is helpful to place the estimation performance in its correct perspective, where the objective is to minimize this error to obtain the lowest possible margin of error in the estimation (Krajewski, Ritzman, Malhotra, & Krajewski, 2008). On the other hand, the maximization of the

minimum of the errors (MINMAX) aims to obtain the minimum of the errors, that is, the most negative, to represent an overestimation of the real value of the attendance to make this overestimation the minimum possible subsequently. This last error is considered because, in conversations with MINEDUC, the latter stated that it wanted to obtain an overestimation in the estimations.

2.2. Mathematical model

To address the problem above, we propose a robust estimation method based on formulating two mathematical models that target the estimation error: the MAPE and the MINMAX. Specifically, we define the decision variables $\hat{x}_{d,j,p,a}$ and $e_{d,j,p,a}$ that express the estimation of the average attendance and the estimation error of the educational institutions of dependency $d \in D$ with a type of day $j \in J$ with teaching modality $p \in P$ in the year $a \in A$. A parameter α that expresses an estimation error of 2,5%. A non-negative decision variable $\lambda_{m,d,j,p} \in [0,1]$ that expresses the relevance of the training month $m \in M_t$ of the educational institutions of dependency $d \in D$ with the type of school day $j \in J$ with teaching modality $p \in P$. The parameters $x_{d,j,p,m,a}$ and $\bar{x}_{d,j,p,a}$ that express the real attendance and the real average attendance, respectively, of the educational institutions of dependency $d \in D$ with the type of day $j \in J$ with teaching modality $p \in P$ in the month $m \in M$ in the year $a \in A$. For both models, let us consider the following set of restrictions $\forall d \in D, \forall j \in J, \forall p \in P$:

$$\hat{x}_{d,j,p,a} = \sum_{m \in M_t} \lambda_{m,d,j,p} x_{d,j,p,m,a} \qquad \forall a \in A \quad (2) \qquad \sum_{m \in M_t} \lambda_{m,d,j,p} = 1 \qquad (3)$$

$$e_{d,j,p,a} = \hat{x}_{d,j,p,a} - \bar{x}_{d,j,p,a} \qquad \forall a \in A \quad (4) \qquad -\alpha \leq \frac{e_{d,j,p,a}}{\bar{x}_{d,j,p,a}} \leq \alpha \qquad \forall a \in A \quad (5)$$

where, the set of constraints (2) and (3) define the convex combination for the attendance estimation that corresponds to the weighted sum of the actual attendance value of the educational institutions in the training month $m \in M_t$. Constraint (4) defines the estimation error corresponding to the difference between the estimation and the actual average attendance. Constraint (5) defines the estimation error tolerance as 2,5%.

Then, the objective of the MAPE model is to minimize the mean absolute percentage error for which we consider an auxiliary variable $Y_{d,j,p,a}$ that expresses the linearization of the absolute value function of the estimation error defined by Expression (6) $\forall d \in D, \forall j \in J, \forall p \in P$:

$$[MIN] \sum_{a \in A} Y_{d,j,p,a} \qquad (6)$$

Subject to the following restrictions (2) to (5):

$$Y_{d,j,p,a} \geq e_{d,j,p,a}/\bar{x}_{d,j,p,a} \,, \qquad Y_{d,j,p,a} \geq -e_{d,j,p,a}/\bar{x}_{d,j,p,a} \qquad \qquad \forall a \in A \qquad (7)$$

where, the set of constraints in (7) defines the linearization of the absolute value function of the estimation error corresponding to the mean absolute percentage error. On the other hand, the objective of the MINMAX model is to obtain the most positive error to represent an overestimation of the real value of attendance, to subsequently make that overestimation the minimum possible for which we consider an auxiliary variable $Z_{d,j,p}$, defined by Expression (8) $\forall d \in D, \forall j \in J, \forall p \in P$:

$$[MIN]\ Z_{d,j,p} \qquad \qquad (8)$$

Subject to the following restrictions (2) to (5):

$$Z_{d,j,p} \geq e_{d,j,p,a} \,, \qquad \qquad e_{d,j,p,a} \geq 0 \qquad \qquad \forall a \in A \qquad (9)$$

where, the set of restrictions in (9) defines the auxiliary variable that rescues the most considerable overestimation of the estimation, and the estimation errors must be positive to represent an overestimation in the attendance estimation.

Then, to determine the attendance estimations of the educational institutions with their respective attributes, we formulated a compromise mathematical model that integrates the two models formulated above. Specifically, we defined two weights $W \in [0,1]$ to give weight to each mathematical model of estimation error. A FO_{MAPE} parameter that expresses the value of the objective function of the MAPE model. And a FO_{MINMAX} parameter that expresses the value of the objective function of the MINMAX model. The above, as shown in Expression (10):

$$[MIN]\ \frac{W * \sum_{d \in D} \sum_{j \in J} \sum_{p \in P} \sum_{a \in A} Y_{d,j,p,a}}{FO_{MAPE}} + \frac{(1-W) * \sum_{d \in D} \sum_{j \in J} \sum_{p \in P} Z_{d,j,p}}{FO_{MINMAX}} \qquad (10)$$

subject to the following expressions (2) to (9) and $\forall d \in D, \forall j \in J, \forall p \in P$:

$$Z_{d,j,p} \geq 0 \qquad \qquad (11)$$

where, constraint (11) indicates which auxiliary variable $Z_{d,j,p}$ represents the minimum overestimation, such that if there is no overestimation $\forall a \in A$, its value is zero.

3. Results

3.1. Statement of the problem

To solve the problem above, we use the actual attendance data provided by MINEDUC of municipal, private subsidized, and local education services educational institutions from 2017 to 2021, corresponding to January to December. For educational institutions whose dependence is "Municipal" (Mun) or "Private Subsidized" (PSub), the corresponding

teaching modalities must have at least three years of historical data. Moreover, the teaching modalities must have at least two years of historical data for those with "Servicios Locales de Educación" (SLE) dependency. Attendance estimations will be made by teaching modality for the year 2021 to be able to compare our results with the results obtained by the MINEDUC model for the same year. Finally, the scope of this work involves the attendance estimations for the year 2021 of the educational modalities whose annual average attendance covers at least 50% of the global average.

3.2. Mathematical model

To solve the instance above, we implemented the mathematical models previously formulated using the Python programming language, considering the following considerations: Pyomo to formulate optimization models, Gurobi as a solver on a PC with an AMD Ryzen 9 5900HS with Radeon Graphics processor at 3.30 GHz and with 32 GB of RAM. Table 1 shows the attendance estimations for the year 2021 for educational institutions of dependency $d \epsilon D$ with a type of day $j \epsilon J$ with teaching modality $p \epsilon P$ obtained by the MAPE model, MINMAX, Compromise Solution with two weightings W equal to 0,5 to integrate the MAPE and MINMAX models and the MINEDUC budget estimation model equally. And a comparison of the actual attendance data for the same year, where the acronyms "C.J" and "S.J" refer to educational institutions with and without a full school day, respectively, while "B" and "M.H.C" correspond to the modalities of primary and secondary scientific and humanistic education, respectively.

4. Discussions and Conclusion

In this paper, we address the problem of uncertainty in the number of students attending educational institutions to allocate the school subsidy in Chile. A robust estimation method is provided by developing a mathematical model that targets the estimation error, the MAPE, and the MINMAX and integrating these through a compromise solution mathematical model. The results obtained are illustrated considering the actual attendance data of the subsidized educational institutions in Chile from 2017 to 2021 delivered by MINEDUC. For this research, the selection of the MINMAX estimation error is because MINEDUC requested that the attendance estimations represent an overestimation of the real data in order not to have a budget deficit in the school subsidy that affects the administrative management of MINEDUC and the quality of the student's education. Now, the attendance estimations of the educational institutions were made at a disaggregated level according to the dependency $d \epsilon D$, type of school day, $j \epsilon J$ and type of education $p \epsilon P$ because MINEDUC's budget estimation model requires it for the school subsidy financing formula. This implies having a carry-over error in each estimation, which could mean a significant error at the aggregate level, i.e., in the final amount of the school subsidy. Therefore, the overestimation could be

massive if we only use the MINMAX error for the estimations. As a result, we will obtain a surplus in the school subsidy budget, which generates inefficiency in the distribution of resources, on the one hand, in the reallocation of resources to other items within the subsidy program or, on the other hand, causing another MINEDUC program not to be executed. Therefore, the MAPE estimation error selection has a regulatory impact at the aggregate level of the school subsidy budget, i.e., the final amount is associated with a smaller confidence level or estimation threshold since some estimations will be overestimated. Others underestimated, where a negative value (underestimated) can save a lot in the overestimation of the estimation.

This behavior of the estimation errors is supported by the results obtained by each mathematical model separately. When estimating attendance only with the MAPE model, most of these estimations at the disaggregated level are below the actual attendance value. For the MINMAX model, most estimations are above the actual attendance value. Moreover, comparing the total amount of attendance, it is observed that when using only the MINMAX model, the total of the estimations is higher and closer to the actual value of attendance than when using only the MAPE model. Then, by integrating these two models in the compromise solution model, it can be verified, according to the results obtained, how the MAPE model attenuates the overestimates of the MINMAX model, finding a balance between underestimates and overestimates. In fact, at the global level of attendance, it can be seen how the total value of attendance decreases by only using the MINMAX model, but at the same time, how the total value of attendance increases by only using the MAPE model. Finally, the results obtained by our robust estimation method were more accurate than those obtained by the MINEDUC estimation model, which ratifies the effectiveness of our proposal and the improvement that our research could contribute.

Something interesting to note about the results obtained by the MINMAX model is that although these results represent overestimates, they needed to be sufficiently loose. This is because in the database provided by MINEDUC, the behavior of the data was stable; in fact, they were similar within and outside the estimation period. It follows that the dispersion of the missing data, i.e., the data to be estimated, varies a little, so because of applying MINMAX, the overestimates were not so loose. Therefore, the carry-over error needed to be bigger, which meant that the disaggregated estimations were relatively accurate in the final amount of attendance.

Table 1. Results of the attendance estimations for 2021 of the set of mathematical models versus the MINEDUC estimation model.

				MAPE	MINMAX	S.C	MINEDUC
D	J	P	Real attendance	Estimation	Estimation	Estimation	Estimation
Mun	C.J	B. 1°-4°	230.576	230.121	230.313	230.121	230.462
Mun	C.J	B. 5°-6°	141.102	141.033	141.330	141.330	141.070
Mun	C.J	B. 7°-8°	143.919	143.905	143.905	143.905	144.616
Mun	C.J	M.H.C	145.965	144.205	144.241	144.205	148.472
PSub	C.J	B. 1°-4°	304.839	306.319	304.974	306.319	299.127
PSub	S.J	B. 1°-4°	189.784	189.607	189.789	189.607	187.603
PSub	C.J	B. 5°-6°	213.924	214.254	214.077	214.254	210.310
PSub	C.J	B. 7°-8°	205.330	204.433	205.364	204.433	202.180
PSub	C.J	M.H.C	308.533	305.253	308.821	305.253	304.686
		Total	1.883.972	1.879.130	1.882.814	1.879.427	1.868.526

Source: Own elaboration.

In conclusion, based on the results obtained, for future work, we propose the use of other mathematical models that target the estimation error, such as the mean absolute deviation

(MAD), the absolute mean square error (MSE), among others, to see how the integration of these in the compromise solution model alters the results of the attendance estimations, either overestimating or underestimating each of the estimations. Now, although in the compromise solution model, the weight assigned to the W weight was 0,5, it would be interesting to analyze how the values of the attendance estimations change as W changes in the objective function of the compromise solution model and also to study the behavior of the weights to see the importance of each month in the attendance estimations, that is, to analyze if there is any trend in the weights of the weights, for example, that March is the most relevant in the attendance behavior.

Acknowledgements

References

Gobierno de Chile. (2022). Gob.cl. Retrieved February 8, 2023, from https://www.gob.cl/ministerios/ministerio-de-educacion/

Krajewski, L. J., Ritzman, L. P., Malhotra, M. K. (Manoj K., & Krajewski, L. J. (2008). *Administración de operaciones: procesos y cadenas de valor*. Pearson Educación.

Marcel, M., & Tokman, C. (2005). *¿Cómo se Financia la Educación en Chile?* Retrieved from https://www.dipres.gob.cl/598/articles-21669_doc_pdf.pdf

MINEDUC. (2021). *Informe mesa financiamiento del Sistema Escolar*. Santiago. Retrieved from https://www.mineduc.cl/informe-mesa-financiamiento-del-sistema-escolar/