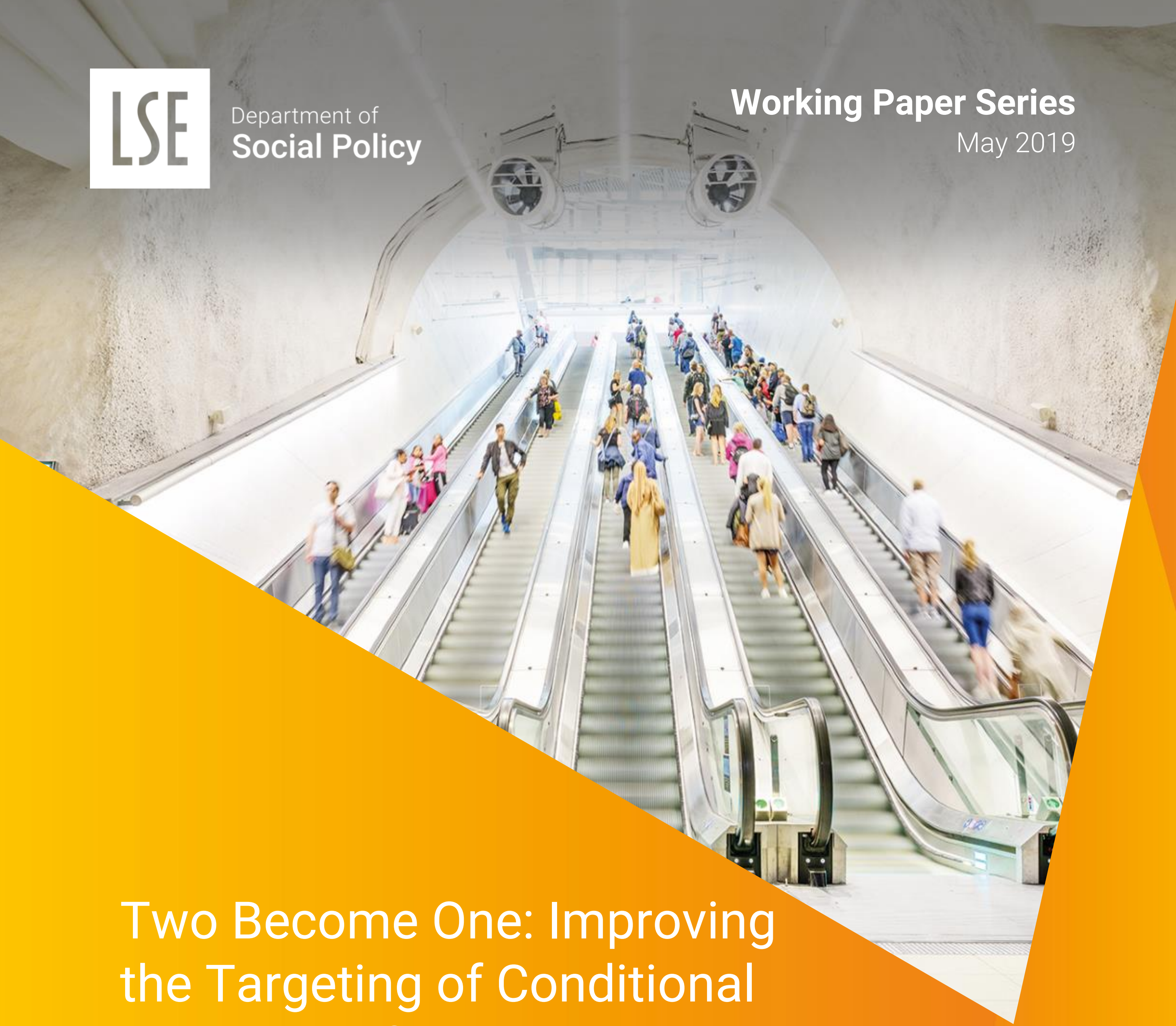
LSE Department of Social Policy

Working Paper Series
May 2019



Two Become One: Improving the Targeting of Conditional Cash Transfers With a Predictive Model of School Dropout

Working Paper 05-19

Cristian Crespo

Social Policy Working Paper 05-19

LSE Department of Social Policy

The Department of Social Policy is an internationally recognised centre of research and teaching in social and public policy. From its foundation in 1912 it has carried out cutting edge research on core social problems, and helped to develop policy solutions.

The Department today is distinguished by its multidisciplinarity, its international and comparative approach, and its particular strengths in *behavioural public policy, criminology, development, economic and social inequality, education, migration, non-governmental organisations* (NGOs) *and population change and the lifecourse.*

The Department of Social Policy multidisciplinary working paper series publishes high quality research papers across the broad field of social policy.

Department of Social Policy
London School of Economics and
Political Science Houghton Street
London WC2A 2AE

Email: socialpolicy.workingpaper@lse.ac.uk
Telephone: +44 (0)20 7955 6001

lse.ac.uk/social-policy

To cite this paper:
Crespo, C. (2019), Cash for Grades of Money for Nothing? Evidence from Regression Discontinuity Designs, Social Policy Working Paper 05-19, London: LSE Department of Social Policy.

Cristian Crespo

Abstract

This paper analyses whether a common targeting mechanism of conditional cash transfers (CCTs), an income-proxy means test (PMT), can identify the poor and future school dropouts effectively. Despite both being key target groups for CCTs, students at risk of dropping out are rarely considered for CCT allocation and in targeting assessments. Using rich administrative datasets from Chile to simulate different targeting mechanisms, I compare the targeting effectiveness of a PMT with other mechanisms based on a predictive model of school dropout. I build this model using machine learning algorithms, one of their first applications for school dropout in a developing country. I show that using the outputs of the predictive model in conjunction with the PMT increases targeting effectiveness except when the social valuation of the poor and future school dropouts differs to a large extent. Public officials that value these two target groups equally may improve CCT targeting by modifying their allocation procedures.

Key words: conditional cash transfers, targeting, school dropout prediction, machine learning, proxy means tests

Author

Cristian is currently a PhD candidate at LSE. His thesis studies questions located at the intersection of conditional cash transfers (CCTs), targeting mechanisms of CCTs, and educational outcomes. Cristian has experience as a program/policy evaluation consultant for universities, governmental agencies, and non-profit organizations. Additionally, he served as the Head of the Research Department in the Ministry of Social Development in Chile.

Introduction

Conditional cash transfers have become a favoured social policy in developing nations. Their rapid expansion, from a few countries in the late 1990s to more than 60 by 2014 (Honorati, Gentilini, & Yemtsov, 2015), demonstrates their popularity. One goal of CCTs is to increase the income of low-income households. Although CCTs have not stated their objectives identically across the globe, authors have pointed out that these schemes seek to reduce the incidence and depth of poverty (Handa & Davis, 2006) and provide a minimum consumption floor to poor households (Fiszbein & Schady, 2009).

Targeting is a crucial element in the design of CCTs. These programmes have intended to allocate their benefits primarily or "rather narrowly" to the poor (Fiszbein & Schady, 2009) (p.7). This is not unplanned, as poor households or individuals are a key target group for CCTs. Targeting is a channel through which to increase a programme's effectiveness within a fixed budget. The more resources that are directed towards the target group (the poor), the more likely a CCT is to achieve its goal of poverty reduction. This explains why evaluations of their targeting (Maluccio, 2009; Robles, Rubio, & Stampini, 2015; Skoufias, Davis, & De la Vega, 2001; Stampini & Tornarolli, 2012) have focused primarily on determining whether CCTs have been given to those who live in poverty.[1]

Different targeting mechanisms exist for CCTs. Proxy means tests are one of the most common in Latin America (Fiszbein & Schady, 2009; Stampini & Tornarolli, 2012). A PMT refers to a system or situation where information correlated with income is used in a formula to proxy income. The formula is obtained through statistical analysis and tends to use data that is easily observable by public officials (Coady, Grosh, & Hoddinott, 2004; Grosh & Baker, 1995).

Targeting low-income households or individuals makes sense not only for CCTs but for a wide range of social programmes. Correspondingly, assessing targeting mechanisms of social policies in terms of their ability to find this target group is a widespread practice. For example, Coady et al. (2004) evaluate the pro-poor targeting performance of 122 social programmes from 48 countries. Similarly, Grosh and Baker (1995) assess whether PMTs provide useful information on income to target social programmes in three countries in Latin America. Social policy targeting has been implicitly associated with finding the poor and poverty alleviation.

Since many CCTs are provided only if children or adolescents are enrolled in school, an additional purpose of most CCTs is to increase school enrolment (Handa & Davis, 2006). Hence, to maximise the likelihood of achieving this goal, CCTs need to be delivered to a differently-defined target group, namely students with the highest risk of dropping out of primary or secondary school. However, CCT

[1] Stampini and Tornarolli (2012) provide targeting assessments for 13 countries in Latin America. The authors show that the expansion of CCTs on the continent led to increased inclusion of the poor. For example, by the year 2010 the three largest programmes (in Colombia, Mexico and Brazil) had achieved poor coverage rates near 50%. However, this was accompanied by growing levels of non-poor leakage (the proportion of CCT recipients who are not poor). On average, leakage increased by 0.46 percentage points for each additional point in poor coverage.

targeting has been focused more on the income dimension. Thus, CCTs have rarely been assessed regarding their ability to direct their resources to those who are more likely to drop out of school. My paper addresses this gap in the CCT literature.[2]

Assessing the capacity of CCTs to reach potential school dropouts is very important. Without this knowledge the allocation process of CCTs could be sub-optimal from the human capital accumulation perspective. If the targeting mechanism used by a CCT is not an accurate predictor of school dropout, then some students will be given the CCT despite the fact that they would have finished their secondary education without any intervention. Conversely, other students who are at risk of leaving school will never have been the subject of the CCT. In both cases a problem of misidentification exists and its consequence is an ineffective use of resources.

Not considering potential dropouts when targeting CCTs would be less of a cause for concern if school dropout were a negligible problem. But in Latin America the graduation rate (among those one year older than the school finishing age but younger than 27) only reached 0.54 in the late 2000s (Bassi, Busso, & Muñoz, 2015). Similarly, dismissing potential dropouts in CCT targeting would be less of a problem in contexts where a high degree of overlap exists between the latter group and those living in poverty. However, this is not guaranteed. For example, in Chile in 2013, only 16.1% of young school dropouts (aged 15 to 19 years old) lived in a poor household while only 12.4% of poor adolescents had dropped out of school (Opazo, Ormazabal, & Crespo, 2015).

Targeting CCTs exclusively according to the likelihood of dropping out of school would lead to a different problem. In this case, the ability of CCTs to find the poor would be weakened. The problem of targeting both groups for a CCT is well addressed by Maluccio (2009), who states: "It is clear that not all non-poor children attend school. Such children, then, would be missed under a pure poverty-based targeting scheme, but possibly not under a targeting scheme which focused on enrolment. Conversely, many poor children already attend school. While there certainly would be overlap among the beneficiary households selected under various possible approaches they almost certainly would not yield identical groups of beneficiaries" (p.5).

This paper analyses whether a proxy means test can identify the poor and future school dropouts effectively. I evaluate the capacity of a PMT to jointly identify these two target groups relative to alternative targeting mechanisms available to public officers. I use rich administrative datasets from Chile to simulate different targeting mechanisms. The core of the alternative mechanisms I test is a predictive model of school dropout. Therefore, the paper has two complementary parts. From its first part, I derive the predictive model using a range of machine learning algorithms (MLA). In the

[2] Analysing the targeting effectiveness of CCTs in terms of reaching students at risk of dropping out of school is different from assessing the impact of CCTs on school dropout. The former evaluation assesses whether the target group is (or would be) reached by a programme. The latter assessment focuses on the (potential) effect of the programme after implementation. The literature on the impact of CCTs on school enrolment is vast, especially in Latin America. For example, positive effects of CCTs on school enrolment have been found in Colombia (Attanasio et al., 2010; Barrera-Osorio, Bertrand, Linden, & Perez-Calle, 2011), Ecuador (Schady & Araujo, 2008), Honduras (Galiani & McEwan, 2013; Glewwe & Olinto, 2004), Mexico (Schultz, 2004) and Nicaragua (Maluccio & Flores, 2005).

second part, I assess the targeting effectiveness of the PMT, the predictive model and mechanisms combining both sources of information. The paper seeks to advise about the merit of using other targeting mechanisms instead of PMTs for CCTs.

My paper contributes to the body of knowledge on school dropout. The literature to date has mostly focused on the question of why students drop out rather than who will drop out. The work in the latter area is primarily confined to North America, where Bowers, Sprott, and Taff (2013) provide a comprehensive summary. Additionally, machine learning applications of dropout in educational contexts have been more salient for higher education with few papers focusing on primary and secondary schools (Knowles, 2015; Sara, Halland, Igel, & Alstrup, 2015; Sorensen, 2018). Furthermore, the few existing papers that predict school dropout in developing countries (Adelman, Haimovich, Ham, & Vazquez, 2017) have not used MLA.

My paper additionally contributes to the literature on the targeting of CCTs. There have been few attempts to assess CCT targeting that consider more dimensions than just income. A notable exception is Azevedo and Robles' (2013) evaluation in Mexico. To assess the targeting performance of a CCT, their paper presents indicators separately for each dimension. My paper offers not only unidimensional indicators but also two indicators that combine information from the target groups, which facilitates making comparisons between targeting mechanisms. My paper provides a headcount index but also a measure of social welfare to assess targeting.[3]

I compare the MLA using receiver operating characteristic (ROC) curves.[4] The most effective algorithm leads to an area under the curve (AUC) of 0.866. I observe this result for the model predicting school dropout at any point within two years. The most effective algorithm for predicting school dropout within one year produces an AUC of 0.893. My results are better than the ones obtained in Guatemala and Honduras (Adelman et al., 2017) and are in line with the best models tested in the United States (Knowles, 2015; Sorensen, 2018).

In my targeting assessment, there is a trade-off between using the PMT relative to using the MLA-based predictive model. When I use the PMT to target a hypothetical CCT, the targeting indicators associated with the poor improve, but the indicators related to dropouts worsen. The opposite also holds. For different fixed budgets, total leakage (the fraction of students receiving the CCT who are neither poor nor a future dropout) is minimised when I use both instruments in conjunction with each other. In other words, it is more effective to combine the predictive model and the PMT than to use them independently. However, this is not true when the social valuation of the two target groups differs to a large extent. If allocating the CCT to a poor student is four times more valuable relative

[3] Some other differences exist between Azevedo and Robles' (2013) paper and mine**.** Their paper focuses on three dimensions (income, health and education) at the household level while I focus on two dimensions (income and education) for individuals in a specific age range for which these dimensions are critical. Their paper uses few variables or predictors to identify deprivation or risk in the educational dimension. Additionally, they use normative criteria (arbitrary selection of thresholds and weights to combine the predictors) for this purpose. My paper uses a larger pool of variables, which allows for predicting empirically which adolescents will drop out of school.

[4] An ROC curve presents the false positive rate and the true positive rate of all the possible results of a predictive model simultaneously. Its area under the curve measures the overall predictive performance of a model. A model that makes no predictive mistakes has an area under the curve of 1, while a model that predicts at random should achieve an area under the curve near 0.5. Further information about ROC curves is available in the next section.

to allocating it to a future dropout, or vice-versa, the likelier optimal approach is to use solely the mechanism designed to find the target group that is valued the most.

All these results have important policy implications. Firstly, the paper shows that appropriate predictive models of school dropout using administrative datasets can be available for public administrators. These models can prove useful not only for CCTs but also for further policies, such as Early Warning Systems, whose purpose is to prevent school dropout. Secondly, in contexts where public officials value finding the poor and future school dropouts equally, the paper demonstrates that the targeting of a CCT can be improved when other dimensions beyond income are considered in the design. This result highlights the importance of: i) avoiding misalignment between the policy goals, its target groups and the selection of the targeting mechanisms, and ii) developing targeting assessments that are in line with all these definitions.

In summary, the paper provides novel contributions to the social policy targeting field. Overall, the paper's findings are not only relevant for the specific Chilean case but for all developing countries that either have CCTs, wish to develop predictive models of school dropout using administrative records or want to strengthen the targeting effectiveness of their social policies.

The paper unfolds as follows. The second section introduces the data and describes the methods I use in the development of the predictive model of school dropout and the targeting assessment. The third section presents the results of the MLA predicting school dropout. The fourth section shows the findings for the targeting assessment. The concluding section discusses the paper's main findings and comments on its main contributions and implications.

Data and Methods

This section describes in detail the data and methods. The first subsection introduces the data. The second part presents the methodological approach of the predictive model of school dropout. The third subsection elaborates on the procedures and indicators of the targeting assessment. Finally, the fourth part explains how I structure the dataset for the analysis.

Data

Most of the datasets I use were provided by the Ministry of Social Development (MSD) at my request. I combine the datasets using the individual ID number provided by the Chilean State. For privacy purposes the ID numbers were changed by the MSD using an algorithm that is unknown to me but enabled me to merge the datasets. The two most important sources of information in this research are the Ministry of Education (ME) Performance Dataset and the Social Protection File (SPF) Dataset.

Ministry of Education Performance Dataset

This dataset contains information for every student who finishes an academic year in primary and secondary education in Chile. The dataset only excludes students in differential education and flexible adult education. Each yearly dataset has approximately 2,950,000 observations (one per student). I requested eight datasets (from 2009 until 2016) for this paper.

The variables available in this dataset are: i) school ID (9,500 unique values), ii) type of school (with categories such as traditional primary education, scientific-humanistic or technical-professional secondary education), iii) grade (first to twelfth), iv) academic performance, v) percentage of attendance, vi) academic end of year classification, and vii) student ID.

With this information I create additional variables such as school dropout. A full explanation of this indicator is available in the next subsection. Other variables I create directly from this dataset are: i) school size, ii) relative academic performance, iii) relative attendance, iv) school mobility, v) historical dropout rates by school, and vi) academic cohort size.[5]

More educational information at the school level is available from public sources. Using the variable school ID, as a key to merge, I obtain the schools': i) administrative dependency (such as public or private subsidised), ii) geographic location (region), iii) urban or rural status, iv) average performance in SIMCE (the national standardised test), and v) management indicators.

<u>Social Protection File Dataset</u>

This dataset contains information for Chilean households and all their members. The dataset has a two-level structure. Each observation represents an individual (adult or child) who lives in a household. No individual can belong to more than one household. Each household has a unique ID number that allows for identifying all the individuals who belong to it.

Households voluntarily requested the SPF at the local government level. Having an SPF was essential to be eligible for multiple social policies. From January 2010, the dataset had 10,782,270 individuals (Comité de Expertos Ficha de Protección Social, 2010), approximately 63.5% of Chile's population. I use four of these datasets (from 2011 to 2014) in this research.

Some of the variables I access are: income, date of birth, proxy means test score, gender, race, head of household, schooling and employment. With this information I can generate additional variables for each individual such as poverty status (explained in detail in the third subsection) and number of children less than six years old within the household.

I combine the information from the Social Protection File with the Ministry of Education Performance Dataset (at the individual level) to build variables for each academic cohort of students. Some of these variables are: average household income per capita, average schooling of the head of the household and proportion of students with a proxy means test score.

[5] Three variables define an academic cohort: i) the school, ii) the type of education received within that school (for example traditional or adult education, scientific-humanistic or technical-professional), and iii) the grade in which the students were enrolled. Students belonging to the same cohort have these characteristics in common.

Methods: Predictive Model of School Dropout

This subsection explains in detail the methodological approach I take to build the predictive model of school dropout. The first part focuses on the predictors. Secondly, the subsection describes the outcome. The third part elaborates on the characteristics of the functions I use for the predictions. The subsection concludes with the criteria I use to assess the predictions.

In general terms, the problem I address in this part of the paper is to find the best function to predict future school dropout given the information I observe from the past. Most formally:

$$Y_{it+k} = f(X'_{it}, X'_{it-1}, \dots \dots, X'_{it-\mathrm{j}}, Z'_i),$$

I need to find a function f that, given the vectors of variables X'_t (where t is year), X'_{t-1},....., $X'_{t-\mathrm{j}}$ and Z' available for each individual *i*, produces on average the most accurate prediction of the outcome Y in t+k. Given that the outcome, school dropout, is a dichotomous variable, this is a statistical classification problem and f is known as a classifier.

<u>The Predictors</u>

I include two types of predictors in the model. The first group of predictors are contained in vectors $X'_{it}, X'_{it-1}, \dots \dots, X'_{it-\mathrm{j}}$. Specifically, X'_i is a vector of variables that change through time for student *i* (such as academic performance, grade repetition, attendance and mobility). The second group of predictors are embedded within Z'_i, a vector of variables for student *i* that do not vary through time (such as race) or where I can only use one observation (such as age).

The selection of variables I include in the model is motivated by the literature on determinants of school dropout and bounded by the availability of administrative records.

Rumberger and Lim (2008) summarise 203 studies for the United States over 25 years to identify statistically significant predictors of school dropout. Some individual characteristics of students that are relevant predictors are: i) educational performance (for example academic achievement, mobility, grade promotion and age or difference between age and expected age for the grade), ii) behaviours (such as absenteeism, deviance and employment), iii) attitudes (like goals and self-perceptions), and iv) background (for example demographics and health).

Their review also identifies institutional characteristics of students' families, schools, and communities. For example, the structure, practices, financial and human resources of students' families are singled out as predictors. Additionally, the student composition, structural characteristics, resources, processes and practices of schools are highlighted in their research.

Hunt (2008) reviews the literature on factors associated with school dropout in developing countries. The author identifies similar predictors to Rumberger and Lim (2008), but also adds other intrinsic challenges that these nations face such as migration, conflict and limited school supply.

The complete list of predictors I use is available in Appendix A. There are 50 variables in total that aim to cover all the dimensions highlighted by Rumberger and Lim (2008). Given the nature of the sources, the information is richer regarding educational performance and the characteristics of students' families, relative to other predictors such as students' attitudes towards education.

A complementary source for predictor selection is Lamote et al. (2013), who argue that predictive models of school dropout need to account for the longitudinal and hierarchical structure of the datasets. This makes perfect sense due to the relevance of educational performance, which is a time-variant variable, and schools and communities as predictors of future dropout. Accordingly, where this is feasible, all my models use three years of historical information $(X'_t, X'_{t-1}, X'_{t-2})$ and include many variables that are at a higher level than the students.[6]

The dataset I assemble is appropriate for the task as it includes multiple strong predictors of school dropout. The dataset possesses multiple years of information on academic attainment, mobility, attendance in conjunction with information at the household level (such as years of schooling of its members and income per capita) and the school and academic cohort levels.

The Outcome

I use the Ministry of Education Performance Dataset to identify students who dropped out of school. The process involves merging different years of this dataset, and linking observations by the student ID. More precisely, I link each student in primary and secondary education who concluded their academic year *t* and did not graduate from their secondary studies with itself in years *t*+1 and/or *t*+2. Using this procedure, I identify the students that dropped out of school after year *t*.[7] Student dropout can be measured in multiple ways. I use three different measures of school dropout to verify the consistency of the results. These measures are:

- dropout_t1: The student finished the academic year *t* and then failed to enrol in *t*+1 or enrolled but withdrew before the end of the academic year *t*+1.
- dropout_t2: The student finished the academic year *t* and (disregarding what happened in *t*+1) then failed to enrol in *t*+2 or enrolled but withdrew before the end of the academic year *t*+2.
- dropout_t12: The student finished the academic year *t* and then failed to enrol in *t*+1 or enrolled but withdrew before the end of the academic year *t*+1 or failed to enrol in *t*+2 or enrolled but withdrew before the end of the academic year *t*+2.

Dropout_t12 takes a value of one if any of dropout_t1 or dropout_t2 takes a value of one. Along these lines, dropout_t12 can be interpreted as dropping out of school at any point within two years of

[6] A trade-off exists concerning how many years of historical information to use. Adding a year can improve the prediction of school dropout but reduces the sample size. I decide to use three years. Information on *t*−2 has some, albeit limited, predictive power (this will be shown in the next section) and this decision allows me to pool four different cohorts (I give complete details on this topic in the fourth subsection).

[7] Given how I measure school dropout, the sample only includes students that finished the academic year *t*.

completing an academic year. Likewise, dropout_t1 can be interpreted as leaving the school at any point within one year of finishing an academic year.

The Classifier and Machine Learning Algorithms

I determine f using supervised MLA. MLA have expanded their range of users, from computer science to social sciences, such as economics. MLA are a powerful and flexible provider of quality predictions (Mullainathan & Spiess, 2017) and a helpful tool for prediction policy problems (Kleinberg, Ludwig, Mullainathan, & Obermeyer, 2015).[8] Research on topics such as recidivism, teachers' hiring and identification of vulnerable groups has used MLA.

Mullainathan and Spiess (2017) identify three essential characteristics of MLA. Firstly, these algorithms find functions that predict well out of sample or that do not overfit the data. Secondly, MLA can discover a complex structure that is not specified in advance. Finally, a subset of MLA allows researchers to manage high dimensional settings, the cases where the number of variables is larger than the observations (James, Witten, Hastie, & Tibshirani, 2013).

MLA are suitable in my case for three reasons. Firstly, in theory, an approach that maximises the predictions of an outcome outside of the sample is preferred for a prediction policy problem (such as determining which students will drop out) relative to an approach that maximises predictions within the sample. Secondly, *a priori* I ignore the structure of the function (for example the number of variables to include) or the form that achieves the best prediction of school dropout. Using MLA expands the likelihood of finding the best model because some MLA consider interactions and polynomials while other MLA directly address the challenge of variable selection. Finally, with MLA I can better manage the number of parameters to include in the dataset. Although I do not face a high-dimensionality problem, reducing the numbers of predictors (by not directly including higher order terms) facilitates the calculations.

To obtain predictions that work well out of sample, machine learning uses a training dataset and a test dataset. The models must be estimated in the former dataset and assessed with the latter. MLA aim to avoid overfitting, in other words, MLA seek to optimise their predictions in the test dataset (out of sample) rather than in the training dataset (in-sample). To do so, each algorithm first tries to determine its optimal level of complexity in the training dataset. The specific indicators of model complexity vary by algorithm, but in general terms these are called regularisers. The less regularisation there is, the better the model will predict in-sample (Mullainathan & Spiess, 2017). These parameters of model complexity can be viewed as variables that can be tuned with the purpose of producing optimal predictions in the test dataset (Varian, 2014).

The last process is known as empirical tuning. It consists of fitting the algorithm in one part of the training dataset and then determining the optimal value of the regulariser by assessing its prediction

[8] The machine learning literature has focused mainly on the problem of prediction and not on capturing the relationship between the predictors and the outcome. Initially, MLA were not designed to obtain deep structural parameters or for causal inference (Nichols, 2018). However, there is emerging literature that connects MLA with causal inference for policy (Abadie, Athey, Imbens, & Wooldridge, 2014; Athey & Imbens, 2015a, 2015b).

performance in another part of the training dataset (Mullainathan & Spiess, 2017). Van der Vaart, Dudoit, and Van der Laan (2006) show that the effectiveness of the procedure is increased if the training dataset is subdivided into multiple subsamples or folds. This is known as cross-validation, with 5 or 10-fold being the most adopted practices (Mullainathan & Spiess, 2017). In this type of cross-validation, the regulariser with the best average performance is chosen.

Machine learning algorithms vary regarding the flexibility they can offer to find the best f. Shrinkage methods such as lasso and elastic nets are the most restrictive as they can only generate linear functions (no interactions between the predictors or other higher order terms). These algorithms are less flexible than ordinary least squares as there is a penalty for every regression coefficient that is different from zero. This penalty leads to the coefficients of the linear regression being shrunken towards zero relative to least squares (James et al., 2013). Generalised additive models (GAMs) expand the range of shapes to estimate f from linear to more complex approaches, for example some non-linear relationships (James et al., 2013). In practice, GAMs fit a non-linear function separately for each predictor and then add all these functions. As the model is additive, interactions between the predictors are not considered.

Approaches based on trees admit interactions by stratifying the predictor space into some regions (McBride & Nichols, 2016). For example, if only two predictors of school dropout are available (age and attendance), a classification tree algorithm can be as follows: a dropout is predicted only if a student is older than 17 years old and has an attendance of lower than 70%. Methods such as random forest and boosting are the result of the combination of multiple trees.

Finally, a highly flexible approach is support vector machines. In broad terms, in a classification problem this algorithm aims to find a hyperplane separating the two classes. If this hyperplane cannot be found a kernel trick (Theodoridis & Koutroumbas, 2009) is applied. The feature space of the problem is expanded, and a new hyperplane is fitted in this transformed space. This process may produce non-linear class boundaries in the original predictors' space.

James et al. (2013) claim that no single algorithm is superior to all the others in every possible context. Thus, I try multiple MLA. For simplicity, the paper presents results for only six of them: i) elastic nets (glmnet), ii) generalised additive models (gam), iii) gradient boosting (gbm), iv) lasso, v) support vector machines (svm), and vi) random forest (rf). These six MLA use the same inputs, which are the 50 predictors I describe in Appendix A.

I implement the MLA in the software R using the Caret Package. Kuhn (2008) is a precious source for this purpose. I utilise 10-fold cross-validation. I use two test datasets. The first test dataset is useful to conduct out of sample validation while I use the second test dataset to assess the quality of the predictions over time. I explain in detail the design of the training dataset and the two test datasets in the fourth subsection. In respect to the treatment of the predictors, I convert categorical variables into dummies and the Caret Package carries out standardisation on all the predictors before executing each machine learning algorithm.

The Criterion to Select the Best f

Statistical classification problems have only four possible outcomes for dropout prediction. A model either: i) correctly predicts a dropout, ii) incorrectly predicts a dropout, iii) fails to predict a dropout, or iv) correctly predicts a non-dropout. More generally, these processes result in four categories that are labelled true positives, false positives, false negatives and true negatives. This can be summed up in a confusion matrix like the one I present in Table 1.

Table 1: Confusion Matrix or Contingency Table of a Classifier of Student Dropout

True Class	Predicted Class		Total
	Not a Dropout	Dropout	
Not a Dropout	True Negatives (TN)	False Positives (FP)	Non-Dropouts (ND)
Dropout	False Negatives (FN)	True Positives (TP)	Dropouts (D)
Total	Negatives (N)	Positives (P)	All Population (T)

Multiple indicators derived from combinations of the nine shaded cells of Table 1 have been used to report the quality of predictions. Within studies on dropout prediction there is no standard metric that facilitates comparisons (Bowers et al., 2013). Following these authors, I provide true positive rates, false positive rates and accuracy. Table 2 describes these three indicators in detail. The right-hand column of Table 2 uses information from six shaded cells of Table 1. An exhaustive list of this family of indicators is available in Appendix B.

Table 2: Indicators Used in the Predictive Model of School Dropout

Name	Formula
True Positive Rate or Sensitivity	True Positives (TP) / Dropouts (D)
False Positive Rate or 1–Specificity	False Positives (FP) / Non-Dropouts (ND)
Accuracy	[True Negatives(TN)+True Positives(TP)] / Total(T)

A perfect classifier would achieve a true positive rate of one, with all dropouts predicted as such, and a false positive rate of zero, or no incorrect predictions of dropouts. No classifier achieves this performance though. In practice, dropout prediction models intend to maximise the true positive rate (or sensitivity) and minimise the false positive rate (or 1–specificity). Nonetheless, there is a trade-off between these two indicators. As a predictive model classifies more observations as dropouts, the true positive rate increases but so does the false positive rate.

Receiver operating characteristics curves summarise this trade-off. An ROC curve simultaneously displays the false positive rate and the true positive rate given by a classifier (James et al., 2013). While the former indicator goes on the horizontal axis, the latter is located on the vertical axis. All possible outputs or scenarios provided by a classifier are represented in an ROC curve. Given this feature, the area under the ROC curve provides a measure of the overall predictive performance of a classifier. The AUC scale ranges from zero to one as the true positive rate and the false positive rate. The better the classifier is, the closer its AUC will be to 1. Conversely, a classifier making predictions at random has an expected AUC of 0.5.

The AUC is a useful indicator to compare the overall performance of multiple predictive models. Models with a higher AUC are, on average, better in statistical classification relative to models with a lower AUC. A model with an ROC curve that is on top of other curves all the way along the horizontal axis is unambiguously a better classifier in every possible scenario.

I use the AUC estimates to select the best f. I calculate these in the first test dataset (for out of sample validation) and for the three measures of school dropout introduced at the beginning of this subsection. The advantages of using the AUC are twofold. In the first instance, the performance of the MLA predicting school dropout can be compared graphically. Secondly, the AUC integrates in one value all the potential classification outputs of each algorithm. This feature frees me to select an arbitrary threshold to assess the performance of the classifiers (such as to choose the algorithm with the highest true positive rate when the false positive rate reaches 0.20).

Additionally, I provide true positive rates, false positive rates and accuracy for two specific scenarios. I force the MLA to classify 10% and 30% of students as future dropouts. These indicators help to establish comparisons with the outputs obtained by other scholars.

Methods: Targeting Assessment

After identifying the best performing algorithm, I can use two indexes to target a hypothetical CCT. The first is the income-proxy means test score from the Social Protection File. I derive the second from the outputs of the best machine learning algorithm (f). Each of these outputs represents the probability that the model is observing a future school dropout.

This subsection describes the methods related to the targeting assessment of a hypothetical CCT on the poor and future school dropouts. The first part explains how I construct the poverty variable. The following two parts elaborate on the indicators I use to assess targeting: total leakage and leaked welfare. The final part discusses the policy alternatives I utilise.

Poverty

Poverty status is not directly available from the Social Protection File dataset. However, it is possible to build an estimate of poverty status by using household structure and income (in the Social Protection File the most relevant sources of income are labour and pensions). There are many approaches to constructing this variable. I use total household income over its members. Using income per capita is consistent with the "traditional" methodology used in Chile to measure poverty. I define a student as poor if he or she is part of the poorest fifth regarding income per capita in the sample. I choose this poverty line considering the poverty rate was approximately 20% for the population analysed in my paper in one year of the assessment. This threshold is higher than the official poverty line because the sample in my study is not representative of the whole population. This is explained in more detail in the next subsection.

Total Leakage

The poverty targeting literature has offered multiple indicators of targeting effectiveness. For example, the AUC in ROC analyses are available (Baulch, 2002; Wodon, 1997). Nonetheless, one of the more common approaches to assessing the targeting effectiveness of transfers consists of providing undercoverage and leakage rates (Coady et al., 2004). The undercoverage rate is the proportion of poor households or individuals not receiving the programme. The leakage rate is the fraction of the non-poor among those who are receiving the programme.

Two common limitations are associated with using these two rates (Coady & Skoufias, 2004). The first is that they disregard distributional information; for example giving a transfer to someone in the highest 1% of income counts the same as giving it to someone marginally over the poverty line. The second shortcoming is that the size of the transfer is irrelevant. It does not make a difference whether a poor household receives a minuscule transfer or an amount that lifts it over the poverty line. One of the preferred ways to address this limitation has been to assess targeting based on the impact on poverty (Grosh & Baker, 1995; Skoufias et al., 2001).

On the one hand, using leakage and undercoverage rates restricts the depth of the analysis in the dimension of poverty. On the other hand, it facilitates establishing comparisons between targeting indicators of poverty and school dropout. Furthermore, it facilitates combining future school dropouts and the poor into one indicator. Thus, despite the limitations of leakage and undercoverage rates, I opt to use these types of indicators to assess the performance of targeting mechanisms. Five of the indicators I use in the paper are presented in Table 3.

Table 3: Indicators Used in the Targeting Assessment

Name	Formula
Poor Undercoverage	# poor not receiving CCT / # poor
Non-Poor Leakage	# non-poor receiving CCT / # receiving CCT
Dropout Undercoverage	# future dropouts not receiving CCT / # future dropouts
Non-Dropout Leakage	# non-dropouts receiving CCT / # receiving CCT
Total Leakage	# non-dropouts & non-poor with CCT / # receiving CCT

Total leakage, defined as the proportion of non-dropouts and non-poor receiving the CCT after the simulation, can be interpreted as the inclusion error. This can be better appreciated from Table 4. Students (potential recipients of a CCT) can be part of one of four classes. Either they are poor and will drop out of school, they are poor but will not drop out, they are not poor but will drop out of school, or they are not poor and will not drop out. Targeting is unsuccessful when a CCT is given to the fourth type of students because no target group is reached.

Table 4: Successful Targeting and Targeting Errors in a Context of Two Target Groups

True Class	Hypothetical CCT Recipient	
	No	Yes
Non-Poor & Non-Dropout	Successful Targeting	Inclusion Error
Non-Poor & Dropout	Exclusion Error	Successful Targeting
Poor & Non-Dropout	Exclusion Error	Successful Targeting
Poor & Dropout	Exclusion Error	Successful Targeting

The selection of total leakage as the first main indicator of my analysis is also justified on theoretical grounds. One minus leakage can be equivalent to the distributional characteristic (DC), a benefit-cost statistic used to compare the welfare impact of transfers with a common budget (Coady & Skoufias, 2004). The authors show that the DC λ for any given scheme *j* is:

$$\lambda_j = \sum_h \beta^h * \theta^h,$$

where β^h is the social valuation (welfare weight) of extra income to household *h* and θ^h represents the share of the total programme budget received by household *h*.

An advantage of the DC is that welfare weights are made explicit and it generalises from simpler to more complex cases (Coady et al., 2004). When the size of the transfer is identical for each household and the social valuation of extra income is equal to one for a poor household and zero otherwise the DC indicator is equivalent to one minus the leakage rate, as shown:

$$\lambda_j = \sum_h \beta^h * \theta = \theta * \sum_h \beta^h = \frac{\sum_h \beta^h}{\#\ Recipients} = \frac{\#\ Poor\ Recipients}{\#\ Recipients}$$
$$= 1 - \frac{\#\ Non\ Poor\ Recipients}{\#\ Recipients} = 1 - Leakage$$

Under some additional assumptions, when the size of a CCT is identical for each individual and when the social valuation of income is equal to one for any CCT recipient who is either poor or a future dropout and zero otherwise, the DC indicator is equivalent to one minus total leakage.

Total leakage is the cornerstone indicator that I use in my research to compare the targeting performance of alternative instruments. The indicator has the major advantage of allowing for the integration of two important target groups for CCTs. Additionally, the logic behind this indicator is useful for other parts of the assessment, when I focus on social welfare and targeting costs.

Peyre Dutrey (2007) criticises the use of leakage in targeting assessments because this indicator does not account for individuals who are excluded. In other words, undercoverage is not considered. However, leakage and undercoverage rates are related. If coverage increases (and undercoverage decreases), leakage is likely to increase. Therefore, rather than intending to find the optimal rate of undercoverage and leakage, my targeting assessment is done for different coverage levels of a hypothetical CCT, more precisely for three budget allocations for a CCT. I explain this aspect of the paper in detail at the end of this subsection. Overall, within a fixed budget and coverage rate, the targeting mechanism with the lowest value of total leakage is optimal.

<u>Leaked Welfare</u>

I also analyse whether the findings of the targeting assessment hold when I change the social valuation of the target groups. Up to this point, I have implicitly assumed that successfully targeting a student who is poor is as socially equivalent worthwhile as correctly targeting a student that will drop out of school. I introduce four different scenarios of social valuation across the two target groups. In the first two scenarios each target group is twice as important as the other. Furthermore, in the last two scenarios the difference in valuation increases to four times the other target group. The choice of these scenarios does not have any theoretical justification, it is merely practical. Following the logic of the DC, the welfare impact of a transfer scheme *j*, which provides an equal amount for each individual *i* can be measured by the formula:

$$\lambda_j = \omega * \sum_i \gamma^i = \frac{1}{\#\,Recipients} * \sum_i \gamma^i = \frac{\sum_i \gamma^i}{\#\,Recipients} ,$$

where ω is the share of the total programme budget received by each adolescent who is a CCT recipient and γ^i is the social valuation (welfare weight) of extra income to adolescent *i*.

In order for λ_j to have minimum and maximum values of zero and one, I choose the welfare weights using the following logic. A hypothetical CCT recipient *i* who is neither a future dropout nor poor receives a value of γ of zero. Conversely, each CCT recipient *i* who belongs to the highly valued target group receives a value of γ of one. The social valuations of each class of student in each of the four scenarios I use in the paper are presented in Table 5.

Table 5: Social Valuation (Welfare Weights γ) in Different Scenarios

True Class	Social Valuation Scenarios			
	The Poor are Twice as Important as Dropouts	The Poor are Four Times More Important than Dropouts	Dropouts are Twice as Important as the Poor	Dropouts are Four Times More Important than the Poor
Non-Poor & Non-Dropout	0	0	0	0
Non-Poor & Dropout	0.5	0.25	1	1
Poor & Non-Dropout	1	1	0.5	0.25
Poor & Dropout	1	1	1	1

The targeting mechanism *j* that provides the highest λ_j maximises welfare. Given the weights I use, the last statement can be rephrased as follows: the targeting mechanism *j* that provides the lowest $1-\lambda_j$ maximises welfare (for any given budget). This last indicator is the focus of the welfare assessment. For simplicity, I refer to it as "leaked welfare". More formally:

$$Leaked\ Welfare = 1-\lambda_j = 1 - \frac{\sum_i \gamma^i}{\#\ Recipients}$$

Budget Available, CCT Coverage and Targeting Mechanisms

CCTs are not universal schemes. Stampini and Tornarolli (2012) show the coverage of CCTs varied by year and country in Latin America. Consequently, I repeat my targeting assessment for different levels of coverage for a hypothetical CCT. Given that in my study the transfer size remains unchanged, an increase or decrease in the CCT programme budget only affects its coverage. For this reason, I repeat my targeting assessment for different budget scenarios available for a hypothetical CCT. I assume that there are no administrative costs in the first instance. I analyse three CCT coverage or budget scenarios. In the first case, the budget allows for reaching only 5% of the students in the sample. In the second and third scenarios the budget allows for reaching 20% and 40% of the sample, respectively. These three cases aim to recreate real policy environments: i) a narrowly targeted CCT, ii) a CCT whose coverage is in line with the population living in poverty (such as in Chile), and iii) a broadly targeted CCT.

I begin by looking at the targeting performance separately for each instrument. First, I only use the proxy means test score of the Social Protection File. Second, I use the predictions derived from the best f. The assessment continues with two combined mechanisms. I target a hypothetical CCT assigning the first 25% of the budget available using the PMT score, and then this percentage ascends to 75%. For example, when the budget allows for reaching 20% of the students in the sample with a CCT and I allocate this using a combined approach, such as 75% is assigned first using the SPF, the procedure works as follows. In the first place, I select the 15% with the lowest PMT scores among the sample and I assign them the CCT. Consequently, I choose the remaining 5% of the students by observing the highest likelihood of dropping out among those not selected in the first step.

Sample and Dataset Structure

The sample excludes students below seventh grade in year t, younger than 12 years old by June of year t and older than 21 years old by March of year t+1. I apply these restrictions considering that student dropout in Chile is a cause for concern mainly in secondary school and that 21 years old is the maximum age to stay enrolled in traditional secondary education.[9]

Another crucial characteristic of the sample is that it includes only adolescents that had an entry in the SPF. Thus, this is not a representative sample of the population, as those households with the highest earnings were less likely to request an SPF. These features do not favour making inferences about the whole student body. However, this is not problematic if the findings are linked to a subset of the entire population: those students with an SPF. Given that this subset is more likely to be recipients of social programmes, the findings of this study remain relevant.

To undertake the machine learning algorithms and the targeting assessment I structure the dataset based on four year-cohorts t (t = 2011, 2012, 2013, 2014), using information from t–2, t–1, t, t+1 and t+2 for each individual in the cohort. Hence, each cohort on its own is a panel dataset. I pool these four cohorts to obtain the “full dataset”. As a result, this dataset contains observations from eight years (from 2009 until 2016). This is explained in detail in Table 6.

Table 6: Dataset Structure

Cohort	Academic and School Information (From the Ministry of Education)	SPF Info	Dropout Information (From the Ministry of Education)
2011	2009 (t–2), 2010 (t–1) & 2011 (t)	2011 (t)	2012 (t+1) and/or 2013 (t+2)
2012	2010 (t–2), 2011 (t–1) & 2012 (t)	2012 (t)	2013 (t+1) and/or 2014 (t+2)
2013	2011 (t–2), 2012 (t–1) & 2013 (t)	2013 (t)	2014 (t+1) and/or 2015 (t+2)
2014	2012 (t–2), 2013 (t–1) & 2014 (t)	2014 (t)	2015 (t+1) and/or 2016 (t+2)

I divide the “full dataset” into two parts. The “old” subset contains cohorts 2011, 2012 and 2013. I partition the “old” subset using random assignment into a training dataset and a test dataset. Each observation in the “old” subset has a 0.75 probability of ending up in the training dataset. In this last dataset the MLA are trained. I test the algorithms and implement the targeting assessment in the

[9] In the Chilean educational system, students are in theory expected to graduate from secondary education at the age of 18. However, grade repetition and school dropout can delay graduation from secondary studies.

test dataset. The “new” subset contains the 2014 cohort. I use the “new” subset only to assess the quality of the predictions of school dropout over time. This process is called out of time validation and its results are not shown in the body of the paper but in an appendix.

Results: Predictive Model of School Dropout

The first subsection provides summary statistics of school dropout and for multiple variables included in the model. The second part focuses on the results of the MLA predicting school dropout. I provide ROC curves, their AUC, true positive rates, false positive rates and accuracy for three measures of school dropout. This subsection additionally analyses which variables of the model are the ones that mostly explain the variation in school dropout

Summary Statistics

Table 7 provides summary statistics for some individual-level variables. Panel A presents the dropout rates in years t+1 and/or t+2. Panel B describes academic variables for years t and t−1. Panel C introduces variables related to the schools where adolescents were enrolled in year t. Panel D presents the information provided by the SPF in year t. The first four columns of the table summarise the mean values of each variable for each cohort. The last four columns provide the mean, standard deviation, minimum and maximum values for cohorts 2011, 2012 and 2013. These three cohorts comprise the “old” subset, as explained in the last subsection.

The average dropout rates for years t+1 and t+2 are 0.06 and 0.09 respectively for the 2011- 2013 period. Within this time range 11 out of 100 adolescents dropped out either in year t+1 or t+2. All measures of dropout declined annually from 2011 to 2014: i) from 0.07 to 0.05 for adolescents dropping out in year t+1, ii) from 0.10 to 0.07 for adolescents dropping out in year t+2, and iii) from 0.12 to 0.09 for adolescents dropping out either in year t+1 or in year t+2. The last two rows of Panel A illustrate the dynamics of dropout. On average, for 2011, 2012 and 2013 cohorts, 65 out of 100 adolescents who dropped out in year t+1 did not return to school in year t+2. Among those who were dropouts in year t+2, only 48 out of 100 adolescents dropped out in year t+1. Along these lines, 52 out of 100 dropped out exactly in year t+2.

Regarding their academic information in year t, between 2011 and 2013: i) adolescents had an average grade of 5.30, ii) their attendance was 89.7%, iii) 9 out of 10 students were promoted to the next grade, and iv) their mobility (the rate of students switching school between t−1 and t) was 0.24. The average grade and the rate of students promoted marginally increased from 2011 to 2014.

Between 2011 and 2013, 4 out of 10 adolescents attended traditional primary education in year t while 35% and 21% of adolescents were enlisted in traditional secondary education, in scientific-humanistic (SH) and technical-professional (TP) schools respectively. Within this period, 49% and 46% of adolescents were enrolled in private subsidised and public schools.

According to the information provided by the SPF, between 2011 and 2013, on average: i) adolescents were 15.39 years old by the end of the academic year t, ii) half of the students were males, and iii) 9 out of 100 were labelled as from indigenous backgrounds. Concerning the head of their households: i) 45% were females, ii) 59% lived with a partner, iii) 41% were employed and contributing to social security, and iv) their average schooling was 9.59 years. Between 2011 and 2013, the average number of people within each student household reached 4.26 with these divided among 2.15 rooms. The average monthly income per capita reached $61,012 CLP (equivalent to $116.5 USD at the December 30th, 2013 exchange rate).

Table 7: Summary Statistics

Variables	Year-Cohort							
	2011	2012	2013	2014	2011-2013			
	Mean	Mean	Mean	Mean	Mean	Std. Dev.	Min.	Max.
Panel A: Drop Out Information								
Dropout Year t+1	0.07	0.06	0.06	0.05	0.06	0.25	0	1
Dropout Year t+2	0.10	0.09	0.08	0.07	0.09	0.28	0	1
Dropout Year t+1 or t+2	0.12	0.11	0.10	0.09	0.11	0.31	0	1
Dropout Year t+2 (if Dropout Year t+1=1)	0.67	0.64	0.63	0.62	0.65	0.48	0	1
Dropout Year t+1 (if Dropout Year t+2=1)	0.49	0.48	0.47	0.47	0.48	0.50	0	1
Panel B: Academic Information								
Average Grade Year t	5.27	5.31	5.32	5.36	5.30	0.68	1	7
Attendance (%) Year t	88.1	90.9	90.3	90.4	89.7	11.6	1	100
Promoted Year t	0.88	0.91	0.91	0.92	0.90	0.31	0	1
Mobility Year t	0.25	0.24	0.24	0.23	0.24	0.43	0	1
Average Grade Year t−1	5.41	5.37	5.40	5.41	5.39	0.63	1	7
Attendance (%) Year t−1	92.8	90.1	92.1	91.5	91.7	9.3	1	100
Promoted Year t−1	0.94	0.92	0.94	0.94	0.93	0.26	0	1
Mobility Year t−1	0.24	0.24	0.23	0.23	0.23	0.42	0	1

Table 7 (continued): Summary Statistics

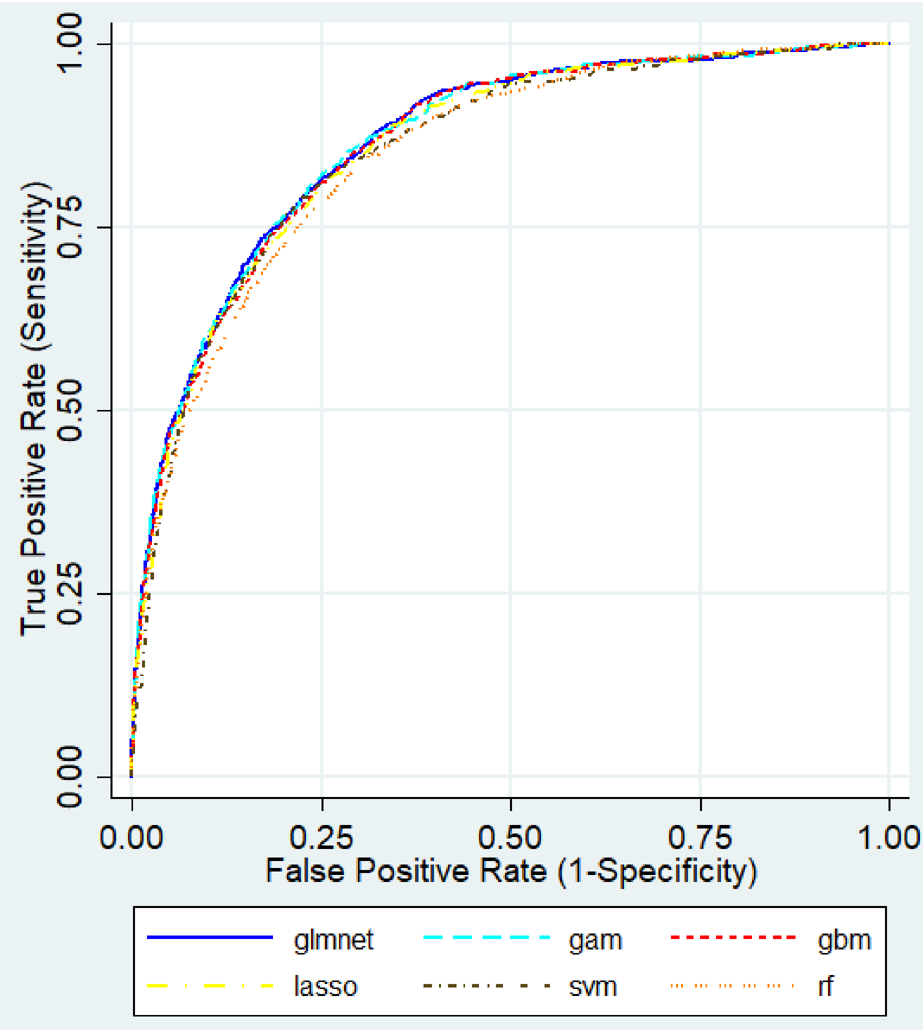
Variables	Year Cohort							
	2011	2012	2013	2014	2011-2013			
	Mean	Mean	Mean	Mean	Mean	Std. Dev.	Min.	Max.
Panel C: School Information (Year t)								
Primary Traditional School	0.40	0.41	0.41	0.41	0.40	0.49	0	1
Secondary Traditional SH School	0.34	0.35	0.36	0.36	0.35	0.48	0	1
Secondary Traditional TP School	0.23	0.21	0.20	0.20	0.21	0.41	0	1
Public School	0.48	0.46	0.45	0.44	0.46	0.50	0	1
Private Subsidised School	0.48	0.50	0.50	0.51	0.49	0.50	0	1
Number of Students in the Academic Cohort	117.1	110.8	108.3	106.1	112.1	103.2	1	1324
Rural School	0.08	0.08	0.07	0.07	0.07	0.26	0	1
Dropout Rate in Previous Academic Cohort	0.07	0.07	0.06	0.06	0.07	0.10	0	1
School Language Score SIMCE	249.8	251.5	251.1	248.7	250.8	26.3	142	345
School Maths Score SIMCE	248.7	249.6	253.7	255.5	250.7	32.7	138	381
Panel D: Social Protection File Information								
Age (Years) at the End of Academic Year *t*	15.39	15.39	15.39	15.41	15.39	1.65	12.50	20.67
Male	0.50	0.50	0.50	0.50	0.50	0.50	0	1
Indigenous Background	0.08	0.09	0.09	0.09	0.09	0.28	0	1
Household Number of Rooms	2.18	2.15	2.11	2.10	2.15	0.93	0	63
Head of Household Lives with a Partner	0.60	0.59	0.58	0.57	0.59	0.49	0	1
Head of Household (HH) is Female	0.44	0.45	0.46	0.48	0.45	0.50	0	1
HH Years of Schooling	9.46	9.62	9.70	9.83	9.59	3.41	0	24
HH Employed and with Social Security	0.41	0.41	0.41	0.41	0.41	0.49	0	1
Household Size	4.30	4.24	4.24	4.21	4.26	1.45	2	34
Household Income per Capita ($CLP)	58,095.0	61,848.8	63,169.2	70,188.6	61,011.9	53,274.3	0	5,533,636
Social Protection File PMT Score	7,537.2	7,384.2	7,193.6	7,025.8	7,373.0	3,733.3	2,072	15,625
Number of Observations	960,514	930,225	937,843	968,584	2,828,582			

Source: own calculations using administrative datasets, Chilean Ministry of Education and Ministry of Social Development

Results for Models Predicting School Dropout

Figure 1 presents the ROC curves for six MLA predicting dropout_t12. A solid line plots the elastic net algorithm (glmnet). The generalised additive (gam) and boosted trees (gbm) models are plotted by long and short dashes, respectively. Dots and dashes plot the other shrinkage algorithm (lasso) and support vector machines (svm), long dashes in the former case and short dashes for the latter. The random forest model (rf) is plotted by dots only.

Figure 1: ROC Curve for Models Predicting School Dropout in Year *t*+1 or *t*+2



Source: own calculations using administrative datasets, Chilean ME & MSD

According to Figure 1, the curves of the six models are close to each other and no single one is above or below the rest along the whole horizontal axis. This suggests that the six MLA have minor differences in terms of the area under the ROC curve. A closer look at the graph shows that the solid line (representing the glmnet model) has a higher degree of convexity. This curve tends to be above all the other curves in a broad range of false positive rates. Conversely, the random forest curve (comprised exclusively by dots) is below the others in some sections of the graph (for example where the true positive rate lies between 0.50 and 0.75).

Figure 2 and Figure 3 focus on the ROC curves predicting dropout_t1 and dropout_t2. The patterns of the lines follow the same logic as for Figure 1. On the one hand, both figures show that the solid curves (glmnet) are predominantly above the other curves. However, the short-dashed curves (gbm) and the long-dashed curves (gam) closely follow and even overpass the solid curves along some parts of the horizontal axis. On the other hand, the svm algorithm (the short dashed and dotted curves) is unambiguously the worst performer in these assessments.

Table 8 presents the area under the ROC curve for each of the six MLA. The first column of the table focuses on adolescents who dropped out in either of the two years after *t*. The last two columns of the table address the predictions of dropouts in years *t*+1 and *t*+2*,* respectively.

Table 7: Area Under the ROC Curve for Models Predicting School Dropout

Machine Learning Algorithms	School Dropout Measures		
	Dropout in Year *t*+1 or *t*+2	Dropout in Year *t*+1	Dropout in Year *t*+2
glmnet	0.866	0.893	0.857
gam	0.865	0.892	0.854
gbm	0.863	0.891	0.851
lasso	0.858	0.885	0.845
svm	0.853	0.843	0.803
rf	0.849	0.875	0.844

Source: own calculations using administrative datasets, Chilean ME & MSD

The elastic net algorithm has the largest AUC in all three cases. In the test dataset, glmnet reaches 0.866 for dropout_t12, 0.893 for dropout_t1 and 0.857 for dropout_t2. Generalised additive models (gam) reach the second highest area under the curve in all three cases. The third highest AUC out of the sample is provided by the boosted trees algorithm (gbm). More generally, for these three algorithms the AUC is above 0.860 in the classification of dropout within two years, 0.890 in the classification of dropout after one year and over 0.850 in the second-year dropout classification. Conversely, random forest (rf) and support vector machines (svm) algorithms have the two worst performances in all three measures of school dropout.

I derive the confidence intervals of the AUC for some of these algorithms (for dropout_t12). The difference between the glmnet and gbm models is not statistically significant at the 95% level. The AUC of each of these two models is statistically significantly different from lasso.

Table 9 helps the reader to understand how the performance of these models translates into targeting effectiveness. The table presents the true positive rate, the false positive rate and the accuracy of the MLA. I need to set a common threshold to be able to compare the MLA among these three measures. The first three columns provide these indicators for a scenario where I classify as future dropouts 10% of adolescents with the highest probability of being a future dropout. In the last three columns the scenario considered is 30% of adolescents being classified as school dropouts. Appendix C shows the results of Table 8 and Table 9 in the second test dataset, the one that assesses the predictions over time (out of time validation).

The left part of Panel A in Table 9 shows that the glmnet algorithm has the best performance. This model finds future dropouts at a rate of 477 out of 1000 in the scenario where 10% of adolescents are classified as dropouts. Additionally, its false positive rate is 0.053. In other words, non-dropouts are incorrectly classified as dropouts at a rate of 53 cases out of 1000. Finally, this algorithm successfully classifies 89.5% of the students. The second and third best performing models in the first scenario are gam and gbm, consistently with the AUC ranking. The true positive rates in these cases are 0.474 and 0.471, respectively. The false positive rate and the accuracy indicators are the same for both algorithms, reaching 0.054 and 0.894 respectively. The two algorithms with the lower targeting performance are rf and svm. The first of these algorithms finds future dropouts at a rate of 438 out of 1000 and misclassifies non-dropouts at a rate of 57 cases out of 1000. These results are consistent with the ROC curves.

In the right-hand part of Panel A, where 30% of adolescents are classified as dropouts, the best performance belongs to the gam algorithm. In this context, dropouts are found at a rate of 810 out of 1000. However, the false positive rate and the accuracy of the model weakens. The first indicator reaches 0.237 while the second reaches 0.768. The algorithm based on elastic nets (glmnet) has the second-best performance after gam among the true positive rate and accuracy.

From the left part of Panel B, gam has the best performance in the classification of first-year dropouts in the scenario where 10% of adolescents are classified as dropouts. In this model, future dropouts are identified at a rate of 584 out of 1000, while its false positive rate is 0.066. For the glmnet model the true positive rate is 0.567 and the algorithm incorrectly classifies non-dropouts at a rate of 67 out of 1000. The worst performing model (svm) in this exercise has a true positive rate of 0.494 and a false positive rate of 0.072. The last three columns of Panel B present the results for the second scenario, where 30% of adolescents are classified as dropouts after the first year. Boosted trees (gbm), glmnet and gam achieve the three highest values of sensitivity. These models reach the second, third and fourth lowest false positive rates and are among the four highest-ranked accuracies.

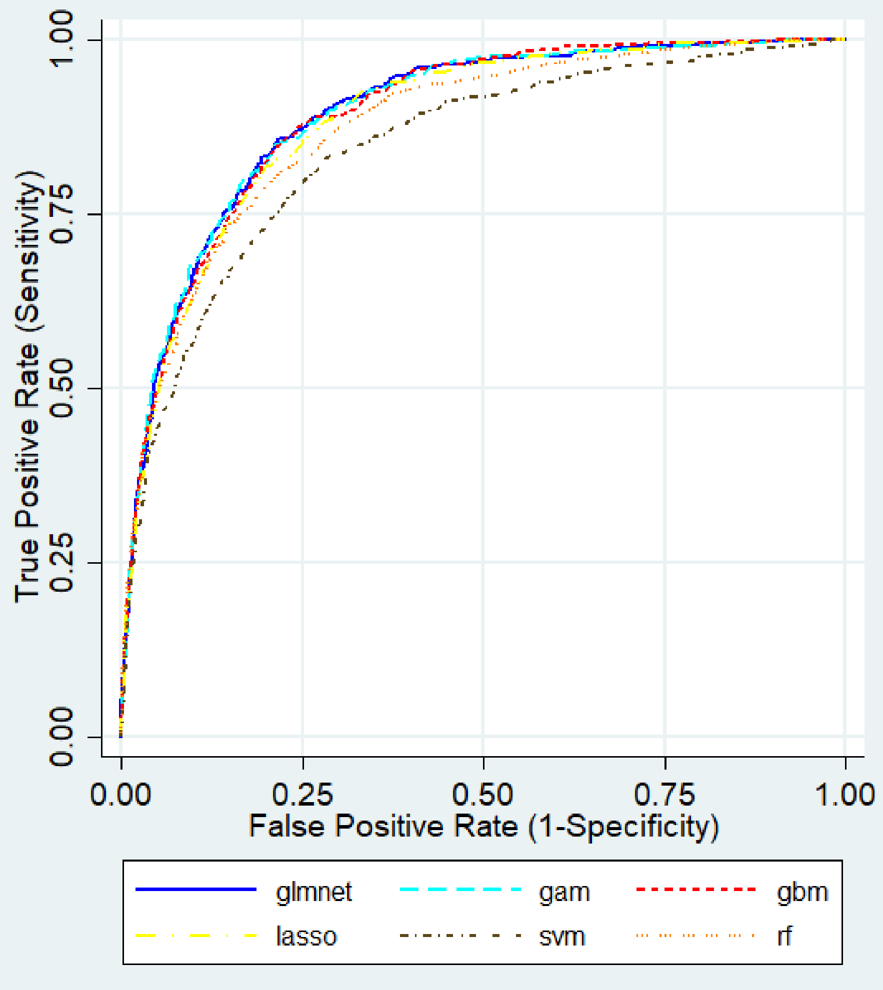
Panel C presents the sensitivity, false positive rate and accuracy for each algorithm predicting dropout in year *t*+2. In the first scenario, the random forest, elastic net and generalised additive models are the highest performers. In the second scenario, the boosted trees algorithm replaces the random forest algorithm among the group with the highest true positive rate.

Figure 4 and Figure 5 show the variable importance in the prediction of school dropout for glmnet and gbm. 13 variables can be found among the 20 most important in both algorithms.[10] Among the 5 most important in either of the models I find: age, average grade in years *t* and *t*−1, attendance in year *t*, relative average grade and attendance in year *t*, the grade (seventh to twelfth) at which the student is in year *t* and the previous average rate of dropout in the school. Income per capita plays a minor role in helping school dropout prediction in these two models.

Overall, differences in performance across the models exist but these are small in magnitude. In general, glmnet, gam and gbm are the top performers, while svm shows the worst results. The best MLA produce adequate predictions of school dropout. Regarding the true positive and false positive rates, my results are better or in the same region as 107 out of the 110 dropout flags analysed by Bowers et al. (2013). The results provided by glmnet are better than the ones obtained in Guatemala and Honduras (Adelman et al., 2017). The accuracy levels shown in the left part of Table 9, around 90%, are equivalent to the results obtained by the best performing MLA tested in North Carolina (Sorensen, 2018). My AUC findings are in line with the best-performing models of school dropout tested in Wisconsin (Knowles, 2015), where most of the algorithms have an AUC of between 0.860 and 0.870. However, these results are below the areas under the curve of 0.948 and 0.965 observed in Denmark (Sara et al., 2015). The policy implications of these results are discussed in the concluding section of the paper.

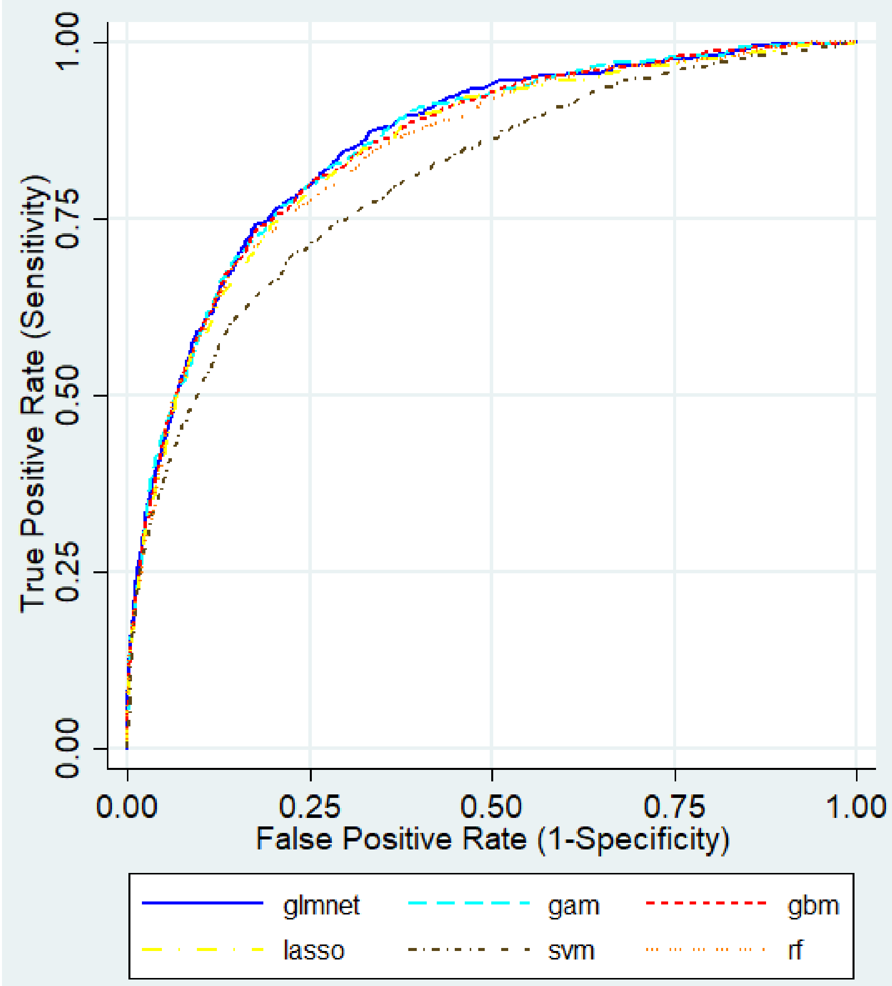
[10] In the Caret Package the maximum value of variable importance is 100. The procedure to estimate the variable importance varies by approach. For example, algorithms based on trees require permuting predictors to assess their accuracy while linear models utilise the absolute value of the *t*-statistic of each coefficient in the regression.

Figure 2: ROC Curve for Models Predicting School Dropout in Year *t*+1



Source: own calculations using administrative datasets, Chilean Ministry of Education and Ministry of Social Development

Figure 3: ROC Curve for Models Predicting School Dropout in Year *t*+2



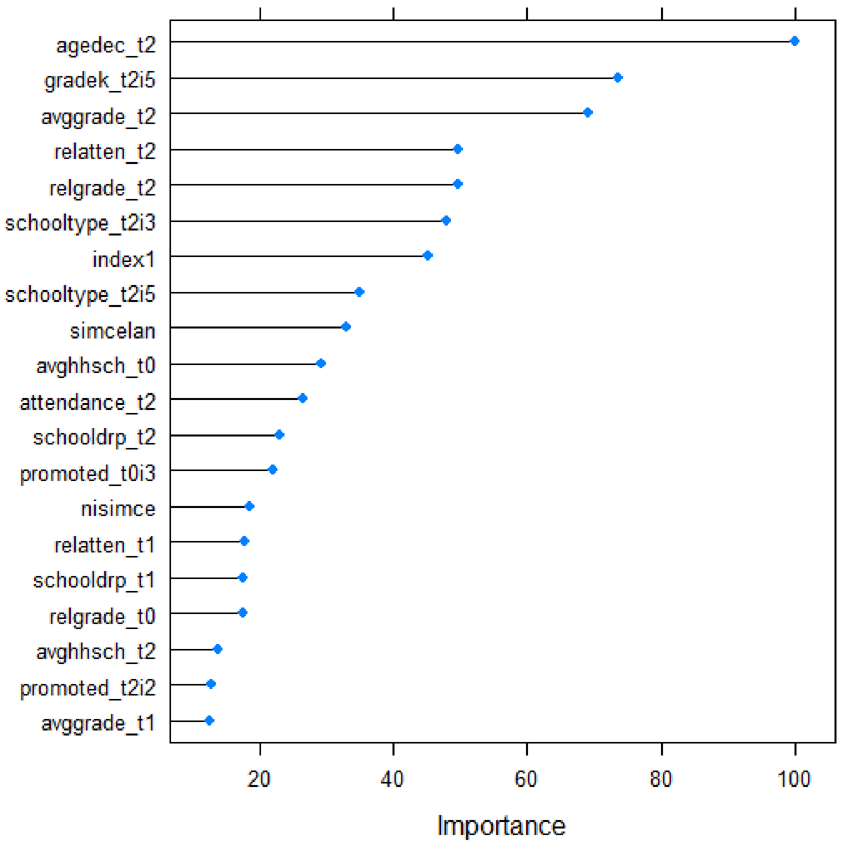
Source: own calculations using administrative datasets, Chilean Ministry of Education and Ministry of Social Development

Table 8: True Positive Rate (Sensitivity), False Positive Rate (1−Specificity) and Accuracy for Models Predicting School Dropout

Machine Learning Algorithms	Scenario 1: 10% of Adolescents Classified as Dropouts			Scenario 2: 30% of Adolescents Classified as Dropouts		
	True Positive Rate (Sensitivity)	False Positive Rate (1−Specificity)	Accuracy	True Positive Rate (Sensitivity)	False Positive Rate (1−Specificity)	Accuracy
Panel A: Dropout in Year t+1 or t+2						
glmnet	0.477	0.053	0.895	0.803	0.238	0.767
gam	0.474	0.054	0.894	0.810	0.237	0.768
gbm	0.471	0.054	0.894	0.794	0.239	0.765
lasso	0.461	0.055	0.891	0.789	0.239	0.764
svm	0.449	0.057	0.889	0.801	0.238	0.766
rf	0.438	0.057	0.887	0.770	0.236	0.765
Panel B: Dropout in Year t+1						
glmnet	0.567	0.067	0.909	0.879	0.259	0.750
gam	0.584	0.066	0.911	0.874	0.260	0.749
gbm	0.561	0.068	0.908	0.881	0.259	0.750
lasso	0.561	0.068	0.908	0.861	0.261	0.747
svm	0.494	0.072	0.899	0.807	0.264	0.740
rf	0.539	0.069	0.905	0.833	0.252	0.754
Panel C: Dropout in Year t+2						
glmnet	0.483	0.064	0.897	0.800	0.252	0.752
gam	0.486	0.063	0.898	0.800	0.252	0.752
gbm	0.481	0.064	0.897	0.799	0.253	0.752
lasso	0.476	0.064	0.896	0.794	0.253	0.751
svm	0.434	0.068	0.889	0.718	0.260	0.738
rf	0.491	0.063	0.899	0.773	0.250	0.752

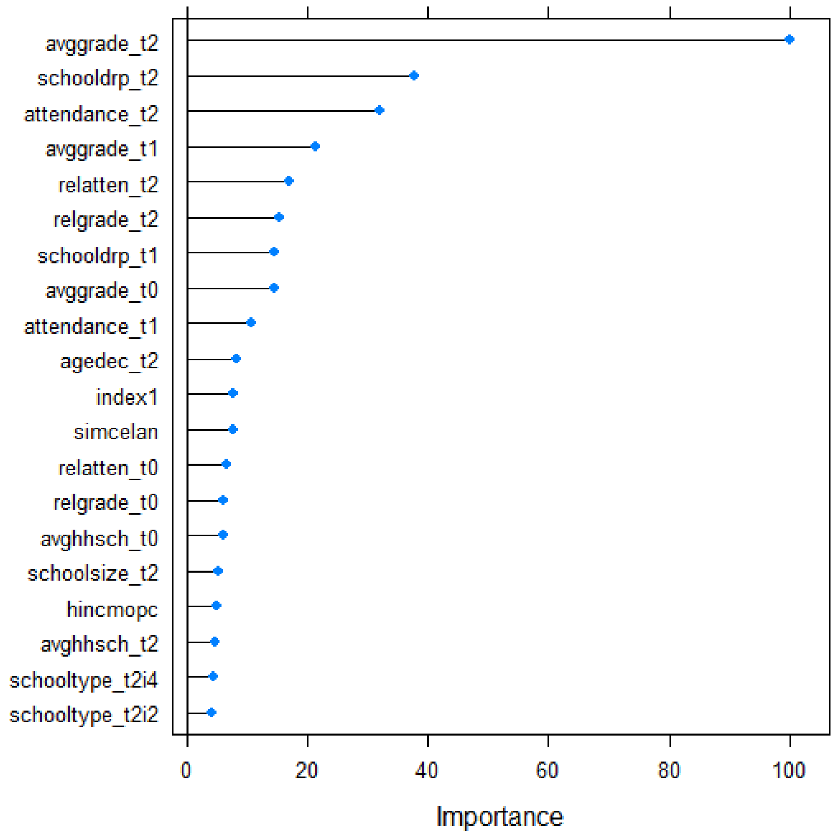
Source: own calculations using administrative datasets, Chilean Ministry of Education and Ministry of Social Development

Figure 4: Variable Importance for glmnet Algorithm



Source: own calculations using administrative datasets, Chilean Ministry of Education and Ministry of Social Development

Figure 5: Variable Importance for gbm Algorithm



Source: own calculations using administrative datasets, Chilean Ministry of Education and Ministry of Social Development

Results: Targeting Assessment

This section presents the results of the targeting assessment. The first part provides summary statistics. These statistics describe the relationship between poverty and school dropout and between each targeting mechanism and the last two outcomes. The second subsection presents the results for total leakage. The concluding part introduces the results for leaked welfare.

Summary Statistics

Table 10 provides bivariate summary statistics between targeting mechanisms (organised in quintile groups) and the outcomes of the targeting assessment. The outcomes, poverty status and school dropout, are presented in the columns. I offer both the mean value and the relative frequency. As poverty status and school dropout are dichotomous variables (they are either zero or one), the mean value can be interpreted as the proportion of poor adolescents and school dropouts within each quintile group, respectively. Contrarily, the relative frequency describes the distribution of poor adolescents and future school dropouts among the quintile groups.

Panel A describes the relationship between income quintile groups, poverty and school dropout in the test dataset. The rows of Panel A present the five income groups. The first group represents the bottom fifth of adolescents regarding household income per capita. As I define poverty by being in the first household income per capita quintile group of the sample, Panel A additionally partly reveals the relationship between the two outcomes of this assessment.

Panels B and C of Table 10 follow a similar logic to Panel A. I present the mean value and the relative frequency of poor adolescents and future school dropouts by quintile groups. Panel B focuses on the quintile groups of PMT scores in the Social Protection File. I assign adolescents in the bottom fifth of SPF scores to the first quintile group. In contrast, the quintile groups in Panel C are related to the predictions of the best performing algorithm in the previous section (glmnet). I assign adolescents with the 20% highest probability of dropping out of school to the first quintile group. Conversely, I allocate students with the lowest probability of dropping out (or a higher probability of remaining at school) to the fifth quintile group.

The measure of school dropout I present is dropout_12. Consistently, Panel C uses the quintile groups from the glmnet model predicting adolescents that leave school in years *t*+1 or *t*+2. Summary statistics for the other two measures of school dropout are available in Appendix D.

Panel A of Table 10 shows that a negative correlation exists between household income per capita and dropping out. The proportion of adolescents who leave school at any time within two years declines from the first income quintile group to the fifth. On average: i) 15 out of 100 adolescents left school from the 20% in the lowest income group in the sample, and ii) only 7 out of 100 adolescents dropped out of school when they belonged to the top 20% in terms of income in the sample. Regarding the relative distribution of future school dropouts among the income per capita quintile groups, 27.97% of adolescents who dropped out belonged to the first income quintile group in the sample. Only 12.13% of school dropouts were from the 20% with the highest income in the sample. Accordingly, it is possible to learn that there is not a big overlap between poor adolescents and future school dropouts in my sample.

Given these results, it is likely that a targeting instrument designed to find one specific group (such as the poor) will have a lower capacity to identify the other group (school dropouts).

Table 9: Mean and Relative Frequency of Poor and School Dropout by Quintile Groups

Quintile Groups	Poor		Dropout *t*+1 or *t*+2	
	Mean	Relative Frequency (%)	Mean	Relative Frequency (%)
Panel A: By Quintile Groups of Income per Capita				
1	1	100	0.15	27.97
2	0	0	0.13	23.05
3	0	0	0.11	19.81
4	0	0	0.09	17.04
5	0	0	0.07	12.13
Total	0.20	100	0.11	100
Panel B: By Quintile Groups of SPF Scores				
1	0.44	44.03	0.14	25.59
2	0.32	32.28	0.13	23.38
3	0.16	16.29	0.12	20.94
4	0.06	5.93	0.10	17.38
5	0.01	1.47	0.07	12.71
Total	0.20	100	0.11	100
Panel C: By Quintile Groups of the Predictive Model of School Dropout (glmnet)				
1	0.29	28.81	0.38	69.63
2	0.24	24.06	0.10	18.76
3	0.20	20.18	0.04	7.46
4	0.16	16.39	0.02	3.11
5	0.11	10.56	0.01	1.05
Total	0.20	100	0.11	100

Source: own calculations using administrative datasets, Chilean ME & MSD

As in Panel A, Panel B shows a negative correlation between SPF scores and leaving school. To illustrate, 14 out of 100 students in the bottom 20% of SPF scores dropped out but only 7 out of 100 did so among those in the fifth quintile group of SPF scores. Also, there is an inverse relationship between PMT scores and poverty. For example, 44 out of 100 adolescents in the first quintile group of the SPF are poor, while only 1 out of 100 students within the 20% highest PMT scores belongs to the 20% with the lowest income in the sample. Regarding relative frequencies, 25.59% of those adolescents who dropped out belong to the first SPF quintile group and only 12.71% of dropouts are from the fifth quintile group of SPF scores.

The SPF score is a better tool for finding poor adolescents than for finding future dropouts. Among the first quintile group of PMT scores 44.03% of poor adolescents can be found. Only 25.59% of future dropouts are located in the lowest quintile group of the PMT scores. These findings are not explained by problems in the SPF model, but rather by the low overlap between poverty and school dropouts in my sample. In fact, the relative frequencies of school dropout by quintile groups shown in Panel B are similar in magnitude to the ones presented in Panel A.

Panel C shows that the predictive model is a more effective tool to find future dropouts than poor adolescents. 69.63% of school dropouts can be found in the first quintile group of the predictive model. Conversely, only 28.81% of the poor are distributed among the 20% with the highest likelihood of dropping out in my sample. Regarding absolute values, there are more future dropouts than poor students in the first quintile group of the predictive model. The latter case occurs despite the population of future dropouts being smaller relative to poor adolescents.

I extract two key findings from Panels B and C. Firstly, to find poor adolescents the SPF score is better equipped (relative to machine learning outputs). In other words, using the PMT is more income progressive than using the predictive model of school dropout. Within the first quintile group of the SPF 44 out of 100 students are poor while in the first quintile group of the predictive model only 29 out of 100 students are poor. Secondly, to find future dropouts the PMT is less effective. Among the first quintile group of SPF scores 14 out of 100 adolescents dropped out. In contrast, in the first quintile group of the predictive model 38 out of 100 adolescents left school. Thus, prioritising the use of SPF scores to target a CCT increases the effectiveness in terms of finding the poor but decreases the capacity to find future dropouts.

Targeting Assessment: Total Leakage

This subsection presents the central results of the targeting assessment. For simplicity, the evaluation focuses on one measure of school dropout. This is the indicator that captures whether an adolescent dropped out in year *t*+1 or year *t*+2 (dropout_t12). Thus, I use the outputs of the best machine learning algorithm predicting dropout_t12 as a targeting mechanism (glmnet). The results for the other two measures of school dropout are available in an appendix.

Table 11 shows the results for the first two (out of four) targeting mechanisms. The left side of the table represents a mechanism based solely on the proxy means test score of the Social Protection File. Conversely, the right-hand side of the table reproduces the results for a mechanism based exclusively on the outputs of the machine learning algorithm.

Table 10: Targeting Indicators by Independent Approach and Budget Available

Targeting Based 100% in SPF Score		Targeting Based 100% in Predictive Model	
Panel A: The Budget Allows a CCT to Reach 5% of Adolescents			
Poor Undercoverage	0.867	Poor Undercoverage	0.922
Non-Poor Leakage	0.470	Non-Poor Leakage	0.689
Dropout Undercoverage	0.934	Dropout Undercoverage	0.696
Non-Dropout Leakage	0.855	Non-Dropout Leakage	0.329
Total Leakage	0.412	Total Leakage	0.232
Panel B: The Budget Allows a CCT to Reach 20% of Adolescents			
Poor Undercoverage	0.560	Poor Undercoverage	0.712
Non-Poor Leakage	0.560	Non-Poor Leakage	0.712
Dropout Undercoverage	0.744	Dropout Undercoverage	0.304
Non-Dropout Leakage	0.859	Non-Dropout Leakage	0.615
Total Leakage	0.493	Total Leakage	0.444
Panel C: The Budget Allows a CCT to Reach 40% of Adolescents			
Poor Undercoverage	0.237	Poor Undercoverage	0.471
Non-Poor Leakage	0.618	Non-Poor Leakage	0.736
Dropout Undercoverage	0.510	Dropout Undercoverage	0.116
Non-Dropout Leakage	0.865	Non-Dropout Leakage	0.756
Total Leakage	0.544	Total Leakage	0.563

Source: own calculations using administrative datasets, Chilean ME and MSD

Table 11 shows that a trade-off exists between finding the poor and future dropouts. Under any budget scenario, poor undercoverage and non-poor leakage increase when switching from SPF scores to the predictive model. For example, when the budget allows for providing the CCT to 5% of adolescents in the sample, poor undercoverage is 0.867 while non-poor leakage is 0.470 if I only use the SPF for targeting. If I use the predictive model, these indicators increase to 0.922 and 0.689, respectively. Conversely, undercoverage of dropouts and leakage of non-dropouts decrease when the output of glmnet replaces the PMT. To illustrate, when the budget allows for providing the CCT to 20% of students in the sample, dropout undercoverage is 0.744 and non-dropout leakage is 0.859 when targeting is based on the SPF. If I use the predictive model for targeting these indicators drop to 0.304 and 0.615, respectively.

An additional trade-off is related to expenditure. The more of the budget that is spent, the lower undercoverage becomes among both target groups. However, leakage rates also increase. Another interesting finding is that the optimal targeting mechanism depends on the budget available. In the first two budget scenarios (in Panel A and B) total leakage is higher if the SPF score is used (relative to the predictive model) for targeting. However, when the budget allows the CCT to reach 40% of adolescents in the sample, total leakage is lower if the PMT is used.

Table 12 switches from independent to combined mechanisms for targeting. As explained in subsection 2.3, I test the targeting performance of two mechanisms that use information from both sources. The left side of the table presents the results for the mechanism where I first allocate 25% of the budget based upon the PMT score (and the remaining 75% is given based on the machine learning algorithm). Conversely, the right-hand side of the table presents the setting where I apportion the first 75% of the budget through the SPF.

Table 11: Targeting Indicators by Combined Approach and Budget Available

Targeting Based 25% in SPF Score		Targeting Based 75% in SPF Score	
Panel A: The Budget Allows a CCT to Reach 5% of Adolescents			
Poor Undercoverage	0.882	Poor Undercoverage	0.852
Non-Poor Leakage	0.530	Non-Poor Leakage	0.408
Dropout Undercoverage	0.740	Dropout Undercoverage	0.854
Non-Dropout Leakage	0.426	Non-Dropout Leakage	0.679
Total Leakage	0.158	Total Leakage	0.232
Panel B: The Budget Allows a CCT to Reach 20% of Adolescents			
Poor Undercoverage	0.665	Poor Undercoverage	0.609
Non-Poor Leakage	0.665	Non-Poor Leakage	0.609
Dropout Undercoverage	0.349	Dropout Undercoverage	0.524
Non-Dropout Leakage	0.640	Non-Dropout Leakage	0.737
Total Leakage	0.419	Total Leakage	0.442
Panel C: The Budget Allows a CCT to Reach 40% of Adolescents			
Poor Undercoverage	0.441	Poor Undercoverage	0.298
Non-Poor Leakage	0.720	Non-Poor Leakage	0.649
Dropout Undercoverage	0.138	Dropout Undercoverage	0.243
Non-Dropout Leakage	0.762	Non-Dropout Leakage	0.791
Total Leakage	0.553	Total Leakage	0.508

Source: own calculations using administrative datasets, Chilean ME & MSD

Similar conclusions can be obtained from Table 12 as from Table 11. In the first instance, there is a trade-off associated with the selection of the mechanisms. To assign a higher fraction of the budget based on the SPF translates into lower undercoverage for the poor and non-poor leakage but a greater lack of coverage for future dropouts and non-dropout leakage. Secondly, when the budget increases, so do all the leakage rates, yet undercoverage for both target groups decreases. Thirdly, the mechanism with the lowest total leakage depends on the budget at disposal.

The last two tables facilitate the comparisons within each targeting approach, but not across them. Table 13 summarises the total leakage indicator for the two independent mechanisms and the two combined mechanisms (from Table 11 and Table 12). Within a fixed budget, Table 13 helps the reader to identify the targeting mechanism with the lowest total leakage.

Table 12: Total Leakage by Targeting Mechanism and Budget Available

Targeting Mechanisms	The Budget Allows a CCT to Reach x% of Adolescents		
	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.232	0.444	0.563
25% SPF; 75% Model	0.158	0.419	0.553
75% SPF; 25% Model	0.232	0.442	0.508
100% SPF; 0% Model	0.412	0.493	0.544

Source: own calculations using administrative datasets, Chilean ME & MSD

A combined approach is more effective in finding the poor or future dropouts relative to an independent approach. For example, in the context where the budget allows for reaching 5% of the sample, the mechanism that uses 25% of the SPF and 75% of the predictive model provides the lowest level of total leakage. In this example, only 15.8% of students who are assigned the hypothetical CCT are neither poor nor dropouts. In the other two budget scenarios, a mechanism that uses both sources of information also provides the optimal solution. In the second case, each combined mechanism performs better in the simulations than the independent mechanisms. When the budget increases to 40%, the optimal mechanism within the alternatives I analyse is to allocate the first 75% of the resources using the PMT score.

Appendix E discusses the robustness of the results in Table 13. I use multiple alternative specifications. First, I change the definitions of the poverty line and income. Second, I modify the measure of school dropout. Third, I use an alternative combined approach. This consists of a single composite score derived from weighting both instruments (the SPF and the predictive model) and assigning the hypothetical CCT using this new index. Finally, I replace the best machine learning algorithm (glmnet) with two others: boosted trees (gbm) and lasso.

Overall, Appendix E shows that my findings are robust to alternative specifications. A targeting mechanism that uses the PMT score in conjunction with the predictive model minimises total leakage (relative to independent mechanisms) in every scenario. This finding does not change depending on the budget, poverty line, income definition, dropout measure or algorithm I select.

In practice, changing the targeting mechanism of a CCT from a PMT to a mechanism that additionally requires using a predictive model of school dropout implies new targeting costs. Appendix F explores in detail whether the conclusions of this part of the assessment hold after adding administrative costs. A targeting approach that relies on both sources of information remains more effective than an independent approach. This holds for all combinations of fixed and variables costs added to targeting mechanisms that incorporate the predictive model.

Targeting Assessment: Leaked Welfare

Table 14 presents the results of the targeting assessment when the social valuation of the target groups differs. Panel A describes two cases where society places greater value on targeting a CCT at a poor adolescent rather than at a future dropout. Panel B does the opposite, in this case finding a future school dropout is twice or four times more important than finding a poor student.

The measure in Table 14 is not comparable to the indicator I provide in Table 13. Unlike total leakage, leaked welfare is affected by differences in the social valuation of target groups. For example, where the poor are twice more important than future dropouts, leaked welfare is: i) zero if all hypothetical recipients of a CCT are poor, ii) one if all beneficiaries are non-poor and non-future dropouts, and iii) 0.5 if all potential recipients are future dropouts but non-poor. Where finding future school dropouts is four times more important than finding the poor, leaked welfare is: i) zero if all recipients are future school dropouts, ii) one if they are neither poor nor future dropouts, and iii) 0.75 if they are all poor but not future school dropouts.

Table 13: Leaked Welfare by Social Valuation of Target Groups

Panel A: Higher Social Valuation of the Poor Relative to Dropouts						
Targeting Mechanisms	The Poor are Twice More Important than Dropouts			The Poor are Four Times More Important than Dropouts		
	The Budget Allows a CCT to Reach x% of Adolescents					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.460	0.578	0.649	0.575	0.645	0.692
25% SPF; 75% Model	0.344	0.542	0.637	0.437	0.603	0.678
75% SPF; 25% Model	0.320	0.525	0.578	0.364	0.567	0.614
100% SPF; 0% Model	0.441	0.526	0.581	0.456	0.543	0.600
Panel B: Higher Social Valuation of Dropouts Relative to the Poor						
Targeting Mechanisms	Dropouts are Twice More Important than the Poor			Dropouts are Four Times More Important than the Poor		
	The Budget Allows a CCT to Reach x% of Adolescents					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.280	0.529	0.659	0.304	0.572	0.707
25% SPF; 75% Model	0.292	0.529	0.657	0.359	0.585	0.709
75% SPF; 25% Model	0.456	0.589	0.649	0.567	0.663	0.720
100% SPF; 0% Model	0.633	0.676	0.704	0.744	0.767	0.784

Source: own calculations using administrative datasets, Chilean ME & MSD

Panel A of Table 14 shows that when the poor are valued more highly than future dropouts it is beneficial to make extended use of the Social Protection File to select beneficiaries. The left side of the Panel shows that the combined mechanism that assigns the first 75% of the budget using the SPF provides the lowest leaked welfare. On the right-hand side of Panel A, where the poor are four times more important, the optimal mechanism in two out of three scenarios is to use the PMT score exclusively.

Panel B of Table 14 demonstrates that relying exclusively on the predictive model is mostly the optimal mechanism when future dropouts are valued more highly than the poor. On the right-hand side of Panel B, where dropouts are four times more important than the poor, not using the SPF minimises leaked welfare in all three budget scenarios. According to the left side of Panel B, a combined mechanism is only superior in the context of a large budget available for a CCT.

Overall, the leaked welfare measure I provide in this subsection improves our understanding of the targeting performance of different mechanisms. When the social valuation of the target groups differs to a large extent, the preferred mechanism is the one designed to find the target group that is most socially valued. When the welfare weight γ assigned to a future dropout is much higher than that of a poor student, using solely the predictive model is the optimal mechanism to maximise welfare. Conversely, when finding a poor adolescent has a much higher social valuation γ than finding a future dropout, prioritising the PMT mostly provides higher levels of welfare.

Conclusion

This paper has analysed whether a PMT and alternative targeting mechanisms based on a predictive model of school dropout built with MLA are effective tools to reach the poor and future school dropouts. Its primary motivation has been the improvement of the targeting design and evaluations of CCTs. Overall, the paper provides novel contributions to the social policy targeting field. Its findings are not only relevant for Chile but for all developing countries that have CCTs, wish to develop predictive models of school dropout using administrative records, or wish to strengthen the targeting effectiveness of their social policies.

A first distinctive contribution of my paper is the predictive model of school dropout. The literature is extensive on the topic of determinants but less so on predictions. The core of this research comes from developed countries, especially the United States. My paper is one of the first, along with Adelman et al. (2017), to use large administrative datasets from a developing nation to study this topic. Furthermore, there are not many applications of MLA for school dropout. The most effective MLA produce results that are in line with the related literature (Adelman et al., 2017; Knowles, 2015; Sorensen, 2018) and that are better than most of the dropout flags analysed in Bowers et al.'s (2013) summary. The best model in predicting school dropout at any point within two years reaches an area under the ROC curve of 0.866 in my test dataset.

These results show that appropriate predictive models of school dropout using administrative datasets are at hand for public officials. Naturally, the selection of variables is restricted by the availability of administrative records, as the models I implement rely solely on information that is currently available from the Chilean government. No variables are provided by costly surveys. This finding has policy implications beyond CCTs, more generally, for every policy that defines students at risk of dropping out of school as their target group. For example, Early Warning Systems can improve their impact by strengthening their ability to find those more likely to drop out of school. In contexts where countries improve their administrative records these lessons deserve attention.

Future research could complement what has been advanced in this paper. For example, longitudinal and multilevel models could be tested for the predictive part. In fact, my approach to find f does not precisely match that of Lamote et al. (2013), which is in the category of longitudinal multilevel modelling. Longitudinal growth models have provided the most accurate predictions on school dropout (Bowers et al., 2013). Additionally, future research projects could consider the case of bringing back to school those who are outside of it when a CCT is implemented.

Another potential direction for further research would be to improve the capabilities of any predictive model by adding new variables. For example, in Chile it is well documented that pregnancy and motherhood are relevant drivers of school dropout (Opazo et al., 2015). Young mothers can be identified through the Civil Register administrative datasets and added to the predictive model. Additionally, the frequency of some predictors I use in my model could be enhanced. For example, the Chilean Ministry of Education has monthly attendance records at the individual level. This information could be useful if attendance levels in the last months of an academic year are a stronger predictor of future dropout than attendance when an academic year starts. The variable absences in the last month is one of the features used by the top-performing algorithm in predicting high school dropout in the literature (Sara et al., 2015).

Another distinctive contribution of my paper is the emphasis on a double vision. Despite having multiple target groups, CCTs have been primarily targeted towards low-income households or individuals. In my targeting assessment, both the poor and future dropouts count. In the paper I offer targeting indicators that combine information on these two key target groups of CCTs. Few papers, beyond Azevedo and Robles (2013), have analysed the quality of CCT targeting by considering more dimensions than just income. My paper and theirs are alike in the sense that both offer a multidimensional targeting approach that fosters the notion that more than one target group and more than one targeting criterion should exist for CCT design and assessment.

The results of the assessment show that a trade-off exists between using the PMT relative to the predictive model. Using the PMT for targeting, instead of the predictive model, is more income progressive, as poor undercoverage and non-poor leakage are reduced. However, future dropout undercoverage and non-dropout leakage increase. This trade-off is explained by the low level of overlap between poverty and future school dropout in Chile. Generally, it is more effective to use these two mechanisms in conjunction rather to use them independently. For different fixed budgets, the proportion receiving the CCT who are neither poor nor a future dropout is minimised with a combined approach. These results hold after considering administrative costs.

These results are partly in line with the findings of Azevedo and Robles (2013). These authors find that their multidimensional targeting approach is better suited to identifying beneficiaries with higher rates of school non-attendance and child labour. This is comparable to my results. Using a combined approach reduces future dropout undercoverage relative to using only a PMT. However, the authors also find that their model identifies the income monetary poor as well as the mechanism used by the CCT. Therefore, in their case no trade-off exists between the two targeting mechanisms assessed. Using multidimensional targeting is always superior.

Another key finding of my paper is that the use of a combined approach is not necessarily more effective when the social valuation of the two target groups differs to a large extent. In the cases where allocating a CCT to a poor adolescent is four times more valuable relative to a future dropout, and vice-versa, it is common to observe that using only one instrument is the optimum. Regarding policy implications, my paper advances the idea that the targeting of CCTs can be improved when other dimensions beyond income are considered. This finding invites policymakers to broaden the targeting design of CCTs by adding the human capital accumulation dimension. Achieving a better balance among target groups in CCT allocation could also help to enrich and diversify the targeting assessment of these schemes, where a unidimensional outlook has prevailed (Maluccio, 2009; Robles et al., 2015; Skoufias et al., 2001; Stampini & Tornarolli, 2012). An essential and implicit takeaway from the paper is that effective targeting depends on consistency. Targeting design must follow the goals of the policy and its consequential definition of the target groups. If a cash transfer has multiple purposes and target groups, then unidimensional targeting may not be the most effective design for this programme.

The latter conclusion does not necessarily hold if public officials strongly prioritise finding the poor over future dropouts. In this case maintaining the status quo, which is targeting CCTs on the basis of income, is appropriate. Alternatively, policy designers should evaluate the cost-effectiveness of adopting a new targeting mechanism for CCTs. A first step in this sense would be to: i) estimate the costs of developing and implementing a new targeting mechanism, ii) estimate the gains in targeting effectiveness, and iii) compare these with the default scenario.

The results of the paper also contribute to enriching the theoretical literature that seeks to minimise poverty or maximise social welfare (Coady & Skoufias, 2004; De Wachter & Galiani, 2006; Glewwe, 1992; Ravallion & Chao, 1989) when these models are applied to CCTs. As in the case of CCT allocation design and evaluation, moving from considering only one dimension in these theoretical models towards multiple dimensions seems desirable. For example, in welfare maximisation models it might be necessary to consider not only the utility provided by the transfer through the income dimension but also by preventing future dropout. In other cases, it might be necessary to include the elasticity of school dropout to extra income. Finally, regarding poverty minimisation problems it may be useful to incorporate future poverty alleviation explained by increased schooling in addition to current poverty alleviation due to the transfer.

Using the framework of social welfare models can enrich the discussion of what effective targeting is. In theory, CCTs should be prioritised towards the groups that would be most impacted. These are the poorest among the poor and adolescents that would drop out of school but would not because of the CCT. For example, a CCT might have an increased impact for an adolescent whose household needs money relative to a peer with little motivation to continue studying because of low school quality. In practice though, targeting CCTs using these criteria requires not only a flawless measurement of the degrees of poverty but also a perfect understanding of the causes of potential school dropout for each adolescent.

Building upon this paper, future research could strengthen my social welfare analysis. One limitation of my targeting assessment is that I use only undercoverage and leakage rates. For example, I make no distinction between those at the bottom of the distribution and those who are marginally poor. I have assumed that the social valuation of finding any poor is the same. A similar shortcoming exists in the case of dropouts. This analysis could also be enriched if the size of the transfers differs, as higher transfers increase the likelihood of obtaining the desired effects.

Additional angles for future research related to this paper are to: i) include more dimensions than education (such as health: include children who are not attending preventive check-ups as a target group), ii) consider other stages in the educational cycle (such as entry to pre-school), iii) model take-up, not everyone that is eligible for the CCT would end up accessing it, and iv) use new predictive models or means tests instead of the PMT used in this paper.

CCTs continue to be a relevant social policy across the globe. Their goals of poverty alleviation and human capital accumulation remain valid in multiple countries. This paper has intended to collaborate in their improved design and assessment concerning targeting. In Chile, a country where administrative datasets are large and rich, using a PMT in conjunction with a predictive model of school dropout allows for finding more adolescents who are either poor or future school dropouts. Public officials that value these two target groups equally, may find opportunities for increased targeting effectiveness by modifying the allocation rules of CCTs.

Appendix A. Predictors Included in Machine Learning Algorithms

Table 14: Predictors of School Dropout Included in Machine Learning Algorithms

Average Grade Year *t*	Urban/Rural
Attendance (%) Year *t*	Same Gender School
Relative Grade Year *t*	Age (Years) at the End of Academic Year *t*
Relative Attendance Year *t*	Male
Promoted Year *t*	Indigenous Background
Mobility Year *t*	Household Number of Rooms
Average Grade Year *t*−1	Head of Household (HH) is Female
Attendance (%) Year *t*−1	Head of Household Lives with a Partner
Relative Grade Year *t*−1	HH Years of Schooling
Relative Attendance Year *t*−1	HH Employed and with Social Security
Promoted Year *t*−1	Relationship with HH
Mobility Year *t*−1	Household Ownership
Average Grade Year *t*−2	Number of Less Than Six Years Olds
Attendance (%) Year *t*−2	Household Income per Capita ($CLP)
Relative Grade Year *t*−2	Average HH Years of Schooling in the Academic Cohort Year *t*
Relative Attendance Year *t*−2	Average HH Years of Schooling in the Academic Cohort Year *t*−1
Promoted Year *t*−2	Average HH Years of Schooling in the Academic Cohort Year *t*−2
Number of Students in the Cohort Year *t*	Index 1 Management of Schools
School Size Year *t*	Index 2 Management of Schools
School Dropout Rate Year *t*	Index 3 Management of Schools
School Dropout Rate Year *t*−1	Index 4 Management of Schools
School Type	Index 5 Management of Schools
Grade	Index 6 Management of Schools
School Administrative Dependency	School SIMCE in Mathematics
School Region	School SIMCE in Language

Appendix B. Indicators Derived from a Classifier of School Dropout

Table 15: Examples of Indicators Derived from a Classifier of School Dropout

Formula	Names
FN / D	False Negative Rate, Miss Rate, Type II Error Rate
TP / D	True Positive Rate, Sensitivity, Recall, Power, 1–Type II Error Rate
TN / ND	True Negative Rate, Specificity
FP / ND	False Positive Rate, 1–Specificity, Type I Error Rate, False Alarm Rate
FP / P	False Discovery Rate
TP / P	Positive Predicted Value, Precision, 1–False Discovery Rate
FN / N	False Omission Rate
TN / N	Negative Predicted Value
(TN+TP) / T	Accuracy

Table 16 shows examples of indicators for a classifier of school dropout. These indicators, which are derived from the nine shaded cells in Table 1, have various names depending on the discipline. Some can be compared to the ones commonly found in the social policy targeting literature. In the context of poor targeting, undercoverage is the proportion of poor households or individuals who do not receive an intervention targeted at the poor (Coady et al., 2004). This indicator is comparable to the false negative rate. In Table 16 this rate is equivalent to the proportion of future dropouts who are not predicted as such. In the context of poor targeting, the false negative rate can be calculated using the number of the poor who are incorrectly predicted as non-poor in the numerator and the total number of the poor in the denominator.

Leakage is defined as the proportion of non-poor households or individuals that receive a programme among all its beneficiaries (Coady et al., 2004). This indicator can be compared to the false discovery rate. In Table 16 this indicator is equivalent to the proportion of adolescents incorrectly classified as future dropouts among all those predicted as future dropouts. In the context of poor targeting, the false discovery rate would be estimated using the number of non-poor who access the programme (those who are incorrectly predicted as poor) in the numerator and all the recipients (all those predicted as poor) in the denominator.

Undercoverage and leakage are usually referred to in the social policy targeting literature as the exclusion and inclusion errors, respectively. However, these last two terms have also been used to explain conceptually false negatives and false positives. Given the potential misunderstanding that could arise from using such a range of different names, I mostly use the terms undercoverage and leakage in the targeting assessment.

Appendix C. Out of Time Validation for the Predictive Model

This appendix contains the results of the predictive model of school dropout in the second test dataset (year-cohort 2014). I use this dataset to assess the quality of the predictions over time. Table 17 shows the AUC for six MLA (similarly than Table 8). Table 18 presents the true positive rate, false positive rate and accuracy for two different scenarios (as in Table 9).

Table 16: Area Under the ROC Curve for Models Predicting School Dropout

Machine Learning Algorithms	School Dropout Measures		
	Dropout in Year *t*+1 or *t*+2	Dropout in Year *t*+1	Dropout in Year *t*+2
glmnet	0.878	0.906	0.876
gam	0.875	0.904	0.873
gbm	0.870	0.901	0.870
lasso	0.872	0.899	0.868
svm	0.866	0.862	0.832
rf	0.860	0.886	0.861

Source: own calculations using administrative datasets, Chilean ME & MSD

The elastic net algorithm (glmnet) also prevails in this dataset. The AUC in the second test dataset reaches 0.878 for dropout_t12 and 0.906 for dropout_t1. The generalised additive models (gam) reach the second highest area under the curve in all three measures. The third highest AUC is provided by the boosted trees algorithm (gbm) or lasso. Conversely, the random forest (rf) and support vector machines (svm) algorithms have the worst performances in all measures.

From Table 18 we can distinguish a similar pattern to the results in Table 17. For example, in the case of dropout_t12, the glmnet algorithm repeats the best performance, finding future dropouts at a rate of 531 out of 1000 in the setting where 10% of adolescents are classified as future school dropouts. In this context, the elastic net algorithm incorrectly classifies non-dropouts at a rate of 56 cases out of 1000 and correctly classifies 90.6% of adolescents. Like in Table 17, gam is among the best performers (in many cases the highest), while svm and rf are the worst performing algorithms in respect to the true positive rate and accuracy.

Overall, the results in this appendix show similar tendencies as in the body of the paper. The relative performance of the MLA does not vary between the two test datasets. Hence, the results in the body of the paper are not sensitive to using information only from the most recent cohort.

Table 17: True Positive Rate (Sensitivity), False Positive Rate (1–Specificity) and Accuracy for Models Predicting School Dropout

Machine Learning Algorithms	Scenario 1: 10% of Adolescents Classified as Dropouts			Scenario 2: 30% of Adolescents Classified as Dropouts		
	True Positive Rate (Sensitivity)	False Positive Rate (1–Specificity)	Accuracy	True Positive Rate (Sensitivity)	False Positive Rate (1–Specificity)	Accuracy
Panel A: Dropout in Year t+1 or t+2						
glmnet	0.531	0.056	0.906	0.829	0.246	0.761
gam	0.530	0.056	0.906	0.834	0.246	0.762
gbm	0.521	0.057	0.904	0.824	0.247	0.760
lasso	0.512	0.058	0.902	0.832	0.246	0.761
svm	0.512	0.058	0.902	0.818	0.247	0.759
rf	0.506	0.058	0.902	0.806	0.245	0.760
Panel B: Dropout in Year t+1						
glmnet	0.648	0.067	0.917	0.904	0.264	0.745
gam	0.662	0.066	0.918	0.899	0.264	0.745
gbm	0.635	0.068	0.915	0.882	0.265	0.743
lasso	0.635	0.068	0.915	0.887	0.265	0.744
svm	0.568	0.072	0.908	0.808	0.270	0.735
rf	0.622	0.069	0.914	0.871	0.262	0.746
Panel C: Dropout in Year t+2						
glmnet	0.556	0.063	0.908	0.845	0.256	0.751
gam	0.546	0.064	0.907	0.849	0.256	0.752
gbm	0.541	0.065	0.906	0.842	0.256	0.751
lasso	0.532	0.065	0.905	0.837	0.257	0.750
svm	0.510	0.067	0.902	0.757	0.263	0.738
rf	0.531	0.065	0.905	0.811	0.250	0.754

Source: own calculations using administrative datasets, Chilean Ministry of Education and Ministry of Social Development

Appendix D. Summary Statistics for Other Measures of School Dropout

Table 19 provides bivariate summary statistics. Its structure is like Table 10. The rows contain the different quintile groups of the targeting mechanisms. While Panel A focuses on the quintile groups of household income per capita, Panel B and Panel C concentrate on SPF scores and the predictive model of school dropout, respectively. The columns show the mean value and relative frequency of school dropout for two measures (dropout_t1 and dropout_t2).

Table 18: Mean Values and Relative Frequency for School Dropout in Years *t*+1 and *t*+2

Quintile Groups	Dropout *t*+1		Dropout *t*+2	
	Mean	Rel. Freq. (%)	Mean	Rel. Freq. (%)
Panel A: By Quintile Groups of Income per Capita				
1	0.09	27.74	0.13	28.84
2	0.07	22.89	0.10	23.46
3	0.07	19.98	0.09	19.63
4	0.06	17.16	0.07	16.61
5	0.04	12.23	0.05	11.45
Total	0.07	100	0.09	100
Panel B: By Quintile Groups of SPF Scores				
1	0.08	25.39	0.11	25.94
2	0.08	23.30	0.10	23.79
3	0.07	20.91	0.09	20.97
4	0.06	17.57	0.08	17.21
5	0.04	12.83	0.05	12.09
Total	0.07	100	0.09	100
Panel C: By Quintile Groups of the Predictive Model of School Dropout (glmnet)				
1	0.25	77.92	0.31	71.78
2	0.05	14.59	0.08	17.49
3	0.02	4.96	0.03	6.84
4	0.01	1.87	0.01	2.86
5	0.00	0.66	0.00	1.03
Total	0.06	100	0.09	100

Source: own calculations using administrative datasets, Chilean ME & MSD

Similar patterns as in Table 10 are observed in Table 19. Among them, there is a negative correlation between income per capita or SPF scores and school dropout. Table 19 also confirms that the predictive model outperforms the SPF in terms of identifying future school dropouts. There are minor differences in the relative frequency of school dropouts in Panels A and B. However, there are substantive differences in the relative frequency of school dropouts in Panel C. Other new findings emerge from Table 19. A higher relative frequency of school dropouts can be found in the first quintile group of the predictive model when the measure of school dropout considered in the analysis is dropout_t1 instead of dropout_t2.

Appendix E. Sensitivity Analysis for Total Leakage Targeting Assessment

I present four distinct types of sensitivity analyses in this appendix. In the first type, I change the methodological approach to measure poverty. On the left side of Table 20 I use a higher poverty line. Applying a higher poverty line has the effect of reducing leakage (as more adolescents are classified as poor, it is more likely that poor students will be found among the recipients). In all cases at least one combined mechanism is more effective than the best independent mechanism.

The right-hand part of Table 20 presents the results for an alternative definition of income. In this case I replace income per capita with household income over an index of needs. This index considers economies of scale of households with multiple members. The outcomes of the targeting assessment are almost unresponsive to these modifications. I observe minor changes in total leakage for the mechanisms that rely primarily on the Social Protection File.

Table 19: Total Leakage: Sensitivity Analysis by Poverty Line and Income Definition

Targeting Mechanisms	Using a Higher Poverty Line			Using a Needs Index		
	The Budget Allows a CCT to Reach x% of Adolescents					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.166	0.321	0.418	0.232	0.444	0.563
25% SPF; 75% Model	0.113	0.272	0.406	0.158	0.417	0.552
75% SPF; 25% Model	0.059	0.295	0.338	0.224	0.440	0.507
100% SPF; 0% Model	0.159	0.312	0.347	0.403	0.490	0.543

Source: own calculations using administrative datasets, Chilean ME & MSD

The targeting assessment in the body of the paper considers dropout_t12. The next sensitivity analysis assesses whether the results vary when I use dropout_t1 and dropout_t2 in the prediction of the MLA and as an outcome in the assessment. Table 21 presents these results.

Table 20: Total Leakage: Sensitivity Analysis by Measure of School Dropout

Targeting Mechanisms	Dropout in Year *t*+1			Dropout in Year *t*+2		
	The Budget Allows a CCT to Reach x% of Adolescents					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.344	0.542	0.635	0.295	0.486	0.592
25% SPF; 75% Model	0.240	0.504	0.621	0.209	0.457	0.583
75% SPF; 25% Model	0.272	0.494	0.560	0.257	0.471	0.534
100% SPF; 0% Model	0.439	0.523	0.575	0.427	0.510	0.561

Source: own calculations using administrative datasets, Chilean ME & MSD

Compared to Table 13 total leakage increases in every case. This is an expected result because dropout_t12 corresponds to a higher population relative to dropout_t1 and dropout_t2. Accordingly, each targeting mechanism has higher difficulties in finding future dropouts. Despite these changes, a combined approach remains more effective relative to an independent approach. For example, as shown in the first column, assigning 75% of the resources through the predictive model reduces total leakage to 0.240. Conversely, if only the predictive model had been used total leakage would have reached 0.344 while using only the SPF would have made this indicator 0.439.

In the third type of sensitivity analysis I test a new version of the combined approach. Table 22 focuses on these modifications. The new mechanism uses a composite score that I create through the combination of the PMT and the predictions of the glmnet algorithm. I assign different weights to each instrument. The left side of the table presents results that are comparable to Table 13. For example, the fourth row in Table 22 is the same as in Table 13. This can be explained by the fact that using a composite score relying 100% upon the SPF is equivalent to distributing the budget using the SPF exclusively. In the second and third rows total leakage increases relative to Table 13. Moreover, composite scores are not always more effective than an independent mechanism. In the scenario with the lowest budget using only the predictive model produces a total leakage of 0.232. Total leakage for the two composite indexes I test reaches 0.234 and 0.329, respectively.

The right-hand side of Table 22 shows the results with an additional change. In this case, total leakage corresponds to the proportion of students who are not simultaneously poor and future school dropouts. Under this definition the composite scores I present are more effective than the independent approach. This result illustrates the relevance of using a targeting mechanism that is consistent with the definition of the target group(s) of a CCT. When finding the poor or school dropouts matters, using a mechanism that takes the best information available from both sources produces better results than allocation through a composite index. When the target group is adolescents who are both poor and future dropouts, a composite index is better suited.

Table 21: Total Leakage: Sensitivity Analysis Using a Composite Score

Targeting Mechanisms (% of Weight in Composite Index)	Target: Poor or Dropout			Target: Poor and Dropouts		
	The Budget Allows a CCT to Reach x% of Adolescents					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.232	0.444	0.563	0.786	0.883	0.929
25% SPF; 75% Model	0.234	0.400	0.533	0.733	0.874	0.927
75% SPF; 25% Model	0.329	0.443	0.523	0.825	0.900	0.932
100% SPF; 0% Model	0.412	0.493	0.544	0.914	0.926	0.939

Source: own calculations using administrative datasets, Chilean ME & MSD

In the last type of sensitivity analysis, I exchange glmnet for two other MLA. The left side of Table 23 focuses on one of the best performing models: boosted trees (gbm). The right-hand side gives the results for lasso, a model that is never the better or the worst predictor in the previous section. The results from gbm are like the ones in Table 13. This is consistent with the equivalent performances of glmnet and gbm as predictors of school dropout. In the case of lasso, total leakage is slightly higher in all contexts relative to glmnet. Despite the latter, the two combined mechanisms are generally more effective than every independent mechanism.

Table 22: Total Leakage: Sensitivity Analysis by Machine Learning Algorithm

Targeting Mechanisms	Boosted Trees Model			Lasso		
	The Budget Allows a CCT to Reach x% of Adolescents					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.233	0.450	0.566	0.259	0.467	0.574
25% SPF; 75% Model	0.156	0.424	0.555	0.176	0.437	0.563
75% SPF; 25% Model	0.230	0.443	0.509	0.236	0.449	0.513
100% SPF; 0% Model	0.412	0.493	0.544	0.412	0.493	0.544

Source: own calculations using administrative datasets, Chilean ME & MSD

Appendix F. Targeting Assessment Including Administrative Costs

The paper has relied on an unrealistic assumption, the inexistence of targeting costs. The social policy targeting literature has identified different families of costs for targeted programmes, among them administrative, incentive, private, social and political (Besley & Kanbur, 1990). Accounting for all these costs is beyond the scope of this paper; however I do consider administrative costs. Within the context of implementing targeted transfers, Coady et al. (2004) associate administrative costs with expenses related to collecting information or building a poverty map.

Given three fixed budgets, this appendix assesses what proportion of the budget goes to students that are non-poor and non-dropouts or is spent on administrative costs. I consider two types of administrative costs. These are a fixed cost of using the predictive model and a variable cost per student selected through the predictive model. The logic behind this design is that using the model is associated with fixed costs such as organising the administrative information and running the model (which is independent of the number of students selected with the instrument) and variable costs (such as outreach through channels other than the ones used by the PMT). No costs are associated with using the PMT score. This assumption is justified on the basis that generally these PMTs are country-level instruments that would not see their cost-structure affected when one programme, out of many, changes its targeting design.

Most formally, for each targeting mechanism I estimate the “leaked budget” as follows:

$$Leaked\ Budget = \frac{Budget\ Received\ by\ Non\text{-}Poor\ \&\ Non\text{-}Dropouts + Admin. Costs}{Total\ CCT\ Budget}$$

The leaked budget indicator can range between zero and one. The targeting mechanism with the lowest value is preferred in this exercise. Table 24 presents the results for the leaked budget indicator. Panel A and Panel B differ by the fixed cost related to implementing a targeting mechanism that includes the predictive model. The left and right-hand sides of each Panel are differentiated from each other by the variable cost of reaching each student with the information from the predictive model. I analyse all these scenarios for the usual three hypothetical budgets.

When no administrative costs exist and when the transfers are equal for every recipient the leaked budget is equivalent to the total leakage. In fact, the results for the rows in Table 24 where I only use the SPF are the same as in Table 13 (because no fixed and variable costs are associated with this option). However, for all the rest of the rows the results are higher relative to Table 13. The addition of administrative costs explains this. These costs reduce the amount of resources that can be directed towards the beneficiaries. Logically, the higher these administrative costs are, the higher the proportion of leaked budget is.

Table 23: Leaked Budget by Administrative (Fixed and Variable) Costs

Panel A: Fixed Cost of Implementing the Model is 0.5% of the Cost of Universal Coverage						
Targeting Mechanisms	Variable Cost is 2%			Variable Cost is 4%		
	Budget Allows to Reach x% of Adolescents (if no adm. costs)					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.310	0.461	0.571	0.322	0.469	0.577
25% SPF; 75% Model	0.237	0.435	0.560	0.246	0.441	0.564
75% SPF; 25% Model	0.319	0.459	0.514	0.321	0.462	0.515
100% SPF; 0% Model	0.412	0.493	0.544	0.412	0.493	0.544
Panel B: Fixed Cost of Implementing the Model is 1.0% of the Cost of Universal Coverage						
Targeting Mechanisms	Variable Cost is 2%			Variable Cost is 4%		
	Budget Allows to Reach x% of Adolescents (if no adm. costs)					
	x=5%	x=20%	x=40%	x=5%	x=20%	x=40%
0% SPF; 100% Model	0.377	0.470	0.575	0.388	0.478	0.580
25% SPF; 75% Model	0.307	0.445	0.564	0.315	0.451	0.568
75% SPF; 25% Model	0.406	0.474	0.518	0.406	0.476	0.519
100% SPF; 0% Model	0.412	0.493	0.544	0.412	0.493	0.544

Source: own calculations using administrative datasets, Chilean ME & MSD

Table 24 provides one key finding. A combined approach remains predominant relative to an independent approach. This is true despite the addition of administrative costs when using the predictive model. When the budget of the programme allows for reaching 5% or 20% of the sample, assigning 25% of the budget with the SPF and 75% with the predictive model is the optimal mechanism. This holds for all the combinations of fixed and variable costs that I consider. In the case where the budget allows for reaching 40% of students in the sample, selecting 75% of the recipients first with the SPF and the rest with the predictive model minimises the leaked budget.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2014). Finite Population Standard Errors. http://www.nber.org/papers/w20325.

Adelman, M., Haimovich, F., Ham, A., & Vazquez, E. (2017). Predicting School Dropout With Administrative Data. http://documents.worldbank.org/curated/en/273541499700395624/pdf/WPS8142.pdf.

Athey, S., & Imbens, G. W. (2015a). Machine Learning Methods for Estimating Heterogeneous Causal Effects. https://www.gsb.stanford.edu/gsb-cmis/gsb-cmis-download-auth/406621.

Athey, S., & Imbens, G. W. (2015b). A Measure of Robustness to Misspecification. American Economic Review, 105(5), 476-480.

Attanasio, O., Fitzsimons, E., Gomez, A., Gutiérrez, M. I., Meghir, C., & Mesnard, A. (2010). Children's Schooling and Work in the Presence of a Conditional Cash Transfer Program in Rural Colombia. Economic Development and Cultural Change, 58(2), 181-210.

Azevedo, V., & Robles, M. (2013). Multidimensional Targeting: Identifying Beneficiaries of Conditional Cash Transfer Programs. Social Indicators Research, 112(2), 447-475.

Barrera-Osorio, F., Bertrand, M., Linden, L. L., & Perez-Calle, F. (2011). Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia. American Economic Journal: Applied Economics, 3(2), 167-195.

Bassi, M., Busso, M., & Muñoz, J. S. (2015). Enrollment, Graduation and Dropout Rates in Latin America: Is the Glass Half Empty or Half Full? Economía, 16(1), 113-156.

Baulch, B. (2002). Poverty Monitoring and Targeting Using ROC Curves: Examples from Vietnam. http://www.ids.ac.uk/publication/poverty-monitoring-and-targeting-using-roc-curves-examples-from-vietnam.

Besley, T., & Kanbur, R. (1990). The Principles of Targeting. http://documents.worldbank.org/curated/en/212811468739258336/pdf/multi0page.pdf

Bowers, A., Sprott, R., & Taff, S. (2013). Do We Know Who Will Drop Out? A Review of the Predictors of Dropping Out of High School: Precision, Sensitivity and Specificity. The High School Journal, 96(2), 77-100.

Coady, D., Grosh, M., & Hoddinott, J. (2004). Targeting of Transfers in Developing Countries: Review of Lessons and Experience. http://siteresources.worldbank.org/SAFETYNETSANDTRANSFERS/Resources/281945-1138140795625/Targeting_En.pdf.

Coady, D., & Skoufias, E. (2004). On the Targeting and Redistributive Efficiencies of Alternative Targeting Instruments. Review of Income and Wealth, 50(1), 11-27.

Comité de Expertos Ficha de Protección Social. (2010). Informe Final Comité de Expertos Ficha de Protección Social. http://www.ministeriodesarrollosocial.gob.cl/btca/txtcompleto/mideplan/c.e-fps-infinal.pdf.

De Wachter, S., & Galiani, S. (2006). Optimal Income Support Targeting. International Tax and Public Finance, 13(6), 661-684.

Fiszbein, A., & Schady, N. (2009). Conditional Cash Transfers: Reducing Present and Future Poverty. https://siteresources.worldbank.org/INTCCT/Resources/5757608-1234228266004/PRR-CCT_web_noembargo.pdf.

Galiani, S., & McEwan, P. J. (2013). The Heterogeneous Impact of Conditional Cash Transfers. Journal of Public Economics, 103, 85-96.

Glewwe, P. (1992). Targeting Assistance to the Poor: Efficient Allocation of Transfers When Household Income Is not Observed. Journal of Development Economics, 38(2), 297-321.

Glewwe, P., & Olinto, P. (2004). Evaluating the Impact of Conditional Cash Transfers on Schooling: An Experimental Analysis of Honduras' PRAF Program. http://web.worldbank.org/archive/website01404/WEB/IMAGES/GLEWWEOL.PDF.

Grosh, M., & Baker, J. (1995). Proxy Means Tests for Targeting of Social Programs: Simulations and Speculation. http://documents.worldbank.org/curated/en/750401468776352539/pdf/multi-page.pdf.

Handa, S., & Davis, B. (2006). The Experience of Conditional Cash Transfers in Latin America and the Caribbean. Development Policy Review, 24(5), 513-536.

Honorati, M., Gentilini, U., & Yemtsov, R. G. (2015). The State of Social Safety Nets. http://documents.worldbank.org/curated/en/415491467994645020/pdf/97882-PUB-REVISED-Box393232B-PUBLIC-DOCDATE-6-29-2015-DOI-10-1596978-1-4648-0543-1-EPI-1464805431.pdf.

Hunt, F. (2008). Dropping Out from School: A Cross Country Review of Literature. http://www.create-rpc.org/pdf_documents/PTA16.pdf.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: With Applications in R. New York: Springer.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction Policy Problems. The American Economic Review, 105(5), 491-495.

Knowles, J. (2015). Of Needles and Haystacks: Building an Accurate Statewide Early Warning System in Wisconsin. Journal of Educational Data Mining, 7(3), 18-67.

Kuhn, M. (2008). Building Predictive Models in R Using the Caret Package. Journal of Statistical Software, 28(5), 1-26.

Lamote, C., Van Damme, J., Van den Noortgate, W., Speyboreck, S., Boonen, T., & Bilde, J. (2013). Dropout in Secondary Education. An Application of a Multilevel Discrete-Time Hazard Model Accounting for School Changes. Quality & Quantity, 47(5), 2425-2446.

Maluccio, J. (2009). Household Targeting in Practice: The Nicaraguan Red de Protección Social. Journal of International Development, 21(1), 1-23.

Maluccio, J., & Flores, R. (2005). Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social. http://www.ifpri.org/publication/impact-evaluation-conditional-cash-transfer-program-2.

McBride, L., & Nichols, A. (2016). Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning. The World Bank Economic Review, 32(3), 531-550.

Mullainathan, S., & Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives, 31(2), 87-106.
Nichols, A. (2018). Implementing Machine Learning Methods in Stata. https://www.stata.com/meeting/uk18/slides/uk18_Nichols.pdf.

Opazo, V., Ormazabal, C., & Crespo, C. (2015). Informe Final de Evaluación. Beca de Apoyo a la Retención Escolar. http://www.dipres.gob.cl/597/articles-141242_informe_final.pdf.

Peyre Dutrey, A. (2007). Successful Targeting? Reporting Efficiency and Costs in Targeted Poverty Alleviation Programmes. http://www.unrisd.org/80256B3C005BCCF9/(httpAuxPages)/0B87C67449C938EDC12573D10049830B/$file/Peyrepap.pdf.

Ravallion, M., & Chao, K. (1989). Targeted Policies for Poverty Alleviation Under Imperfect Information. Journal of Policy Modelling, 11(2), 213-224.

Robles, M., Rubio, M., & Stampini, M. (2015). Have Cash Transfers Succeeded in Reaching the Poor in Latin America and the Caribbean. https://publications.iadb.org/handle/11319/7223.

Rumberger, R. W., & Lim, S. A. (2008). Why Students Drop Out of School. A Review of 25 Years of Research. https://www.issuelab.org/resources/11658/11658.pdf.

Sara, N. B., Halland, R., Igel, C., & Alstrup, S. (2015). High-School Dropout Prediction Using Machine Learning: A Danish Large-Scale Study. In M. Verleysen (Ed.), Proceedings. ESANN 2015: 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (pp. 319-324). Louvain-la-Neuve, Belgium: Ciaco.

Schady, N. R., & Araujo, M. C. (2008). Cash Transfers, Conditions, and School Enrollment in Ecuador. Economía, 8(2), 43-70.

Schultz, T. P. (2004). School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program. Journal of Development Economics, 74(1), 199-250.

Skoufias, E., Davis, B., & De la Vega, S. (2001). Targeting the Poor in Mexico: An Evaluation of the Selection of Households into PROGRESA. World Development, 29(10), 1769-1784.

Sorenson, L. (2018). “Big Data” in Educational Administration: An Application for Predicting School Dropout Risk. Education Administration Quarterly, 1-43.

Stampini, M., & Tornarolli, L. (2012). The Growth of Conditional Cash Transfers in Latin America and the Caribbean: Did They Go Too Far? http://ftp.iza.org/pp49.pdf.

Theodoridis, S., & Koutroumbas, K. (2009). Pattern Recognition. California: Elsevier.

Van der Vaart, A. W., Dudoit, S., & Van der Laan, M. J. (2006). Oracle Inequalities for Multi-Fold Cross Validation. Statistics and Decisions, 24(3), 351-371.

Varian, H. (2014). Big Data. New Tricks for Econometrics. Journal of Economic Perspectives, 28(2), 3-28.

Wodon, Q. (1997). Targeting the Poor Using ROC Curves. World Development, 25(12), 2083-2092.