The Accuracy Of Proxy Means Tests For Immigrant Populations: A Case Study In Colombia

Will Sims, MPP 2020

wlsims@gmail.com

The Gerald R. Ford School of Public Policy

University of Michigan

April 29, 2020

Abstract

This paper examines the accuracy of proxy means tests (PMTs) for identifying low-income households among immigrant populations. Specifically, it develops a PMT model based on Colombia's SISBEN system, and evaluates its ability to identify poverty among recent and established Venezuelan immigrants. It finds that these immigrant groups have significantly higher rates of exclusion errors relative to native Colombians, which could prevent them from accessing valuable social services.These findings are robust to a number of specifications, and the issue is not resolved by simply including immigration status within the model. Additionally, occupational downgrading is identified as the most likely mechanism for this effect, as Venezuelan immigrants in Colombia generally have lower returns to education when compared with native Colombians. These results should inspire caution when choosing to use PMTs for targeting, and it is recommended that all policymakers evaluate the accuracy of their PMTs for vulnerable subpopulations prior to implementation.

Introduction

A large share of workers in the Global South are employed in the informal sector, and as a result accurate and reliable data on income, and thus poverty, is frequently unobtainable. Economic theory suggests that there is a correlation between income and asset ownership, meaning linear regression models can be used to predict the income of informal workers based on assets such as vehicle or livestock ownership, as well as more readily observable characteristics such as education level (Grosh, 1994). These models, known as Proxy Means Tests (PMTs), have become a common tool over the past few decades for targeting anti-poverty interventions on both a local and national scale. Despite their widespread use, the efficacy of PMTs is still under debate, and much research remains to be done to understand both the theoretical and empirical limits of PMT methodologies.

This paper will make three main contributions to the literature on PMTs. First, it will provide the first academic assessment of PMTs in Colombia since Castañeda's study in 2005. Second, this paper will assess the accuracy of PMT methods for measuring income in specific sub-populations — in particular, large groups of newly-arrived immigrants. Prior to this work, analysis of sub-populations — including immigrant groups — has been largely absent from the literature on PMTs. Given the increasing size and vulnerability of immigrant populations across the globe, this is a particularly glaring omission. Finally, this paper seeks to contribute to broader debates about the efficacy of PMTs as a tool for targeting anti-poverty interventions. PMTs can be a useful tool in specific circumstances, but given their widespread use in the development and humanitarian communities, a comprehensive understanding of their limitations and drawbacks is essential for effective policy design.

Given the natural variation among individuals, erroneous predictions are an unavoidable side-effect of PMT methods. When discussing poverty targeting, the errors generated by this unexplained variation are typically classified into two types: *inclusion errors* refer to individuals who are classified as poor or receive a benefit even though their actual income exceeds the model's income threshold, while *exclusion errors* represent individuals who are below the threshold but for whom the PMT model predicts income above the threshold. Both types of errors can pose serious challenges for policymakers. While exclusion errors can prevent vulnerable households from receiving needed services, high rates of inclusion errors can result in higher costs for social welfare programs.

If we focus on immigrant populations in particular, two potential mechanisms could influence the accuracy of PMTs. First, immigrants typically engage in what economics literature has termed occupational downgrading — taking jobs for which they are overqualified or overeducated due to barriers to transferring qualifications or lack of formal em-

---

1

ployment status (Chiswick, 1978). This could result in higher rates of exclusion errors, as well-educated immigrants are predicted to have higher incomes than they are able to earn in their new country.

On the other hand, it may take new immigrants time to build up stocks of assets, so predicting income based on assets could underestimate their income, leading to higher inclusion errors. If this is the case, this effect would diminish over time, as new immigrants acquire assets that align with corresponding native-born income levels. Of course, it is also possible that these effects cancel each other out, or that neither of the effects is significant enough to impact the overall accuracy of the PMT model. This paper will primarily be concerned with the occupational downgrading mechanism, as exclusion errors can have a significant negative impact for already vulnerable immigrant populations, and the relatively small size of the immigrant population will likely limit the impact of inclusion errors on the overall solvency of social programs.

Contemporary Colombia is an ideal context to examine these questions because PMTs have become a significant element of the country's approach to both social welfare and poverty alleviation. The *Sistema de Selección de Beneficiarios para Programas Sociales* (SISBEN), a system of proxy means testing, has been used in Colombia since 1994. It is used to allocate services such as subsidized health insurance and conditional cash transfers (Vélez, Castaño & Deutsch, 1998; Attanasio & Mesnard, 2006). Furthermore, in the past few years, Colombia has experienced an influx of more than a million immigrants from Venezuela (UNHCR, 2019). The Colombian government is currently using SISBEN in order to provide these immigrants with subsidized healthcare and other social services (El Tiempo, 2018; CONPES, 2018). While this paper will focus on a theoretical PMT model (the current SISBEN formula is confidential), the concerns that it raises are applicable to any PMT model currently in use by policymakers.

Colombia's recent influx of Venezuelan immigrants also presents a particularly interesting case for statistical reasons. First, Venezuelan immigrants are a compelling comparison class because they have arrived recently and in large numbers, lending them legibility within the data. Second, Colombia and Venezuela are relatively similar in terms of shared history and culture, so it is less likely that divergent patterns in asset ownership are driven by cultural preferences. Third, Venezuelan immigration to Colombia is catalyzed by a political and economic crisis, so the immigrants likely represent a broader cross-section of education and pre-migration earnings than typical economic migrants (UNHCR, 2019). In terms of broader validity, these unique characteristics of the immigrant population may limit the cross-group and cross-country applicability of our conclusions. That said, we can also speculate that this analysis will likely underestimate the inaccuracy of PMTs relative to other contexts.

Previous Analysis of PMT Accuracy

The theoretical basis for PMT methods is relatively straightforward: higher-income households tend to own more assets, so rough surveys of observable assets can be used to predict income in situations where reliable income data is unavailable. Some research has found that PMT targeting programs can effectively identify poor households, and PMTs have outperformed other targeting methods such as community-based targeting in some studies (Grosh et. al, 2008). Furthermore, research suggests that PMTs do not distort consumption, as asset accumulation habits do not change significantly in response to the implementation of PMTs (Banerjee et. al., 2018). Based on findings like these, PMTs have become a recommended approach to poverty targeting within international development organizations like the World Bank (World Bank, 2017).

Other research has questioned the overall accuracy of PMT methods, suggesting that high error rates in the model can undermine both their effectiveness and political viability (Kidd & Athias, 2019). Even the most accurate PMT models report R^2 values between 0.3 and 0.6, leaving 40% to 70% of total variation in income unexplained (Kidd & Wylde, 2011). They also point to the inherent trade-off between the size of the targeted population and PMT accuracy — narrower bands of income or consumption mathematically increase the probability of exclusion errors. Finally, these statistical issues with the model are compounded during the implementation of PMTs, including inaccuracy over time, problems with falsification of responses, and recipient dissatisfaction. Despite these criticisms, lack of consensus around alternative methods has allowed PMTs to remain a common approach to poverty targeting.

Latin America in particular has a long history of PMT projects and research. Much of the original development of PMT methods by Margaret Grosh and The World Bank took place in Chile in the 1990s (Grosh, 1994; Grosh & Baker, 1995). Since then, there have been numerous evaluations of PMT models across Latin America, including studies in Peru (Johannsen, 2005) and Bolivia (Klasen & Lange, 2015) that found exclusion error rates ranging from 8 to 26 percent. Going beyond models to evaluate implementation, Castañeda and Lindhert compared actual PMT performance in Mexico, Brazil, Costa Rica, Chile, and the United States (Castañeda & Lindhert, 2005). They found rates of exclusion as high as 84 percent for Chile's PASIS old-age pension. Specific to Colombia, a related study by the same authors estimated a best-case exclusion error rate of 19 percent for Colombia's SISBEN PMT program (Castañeda, 2005). The Castañeda and Lindhert research provides essential context for this paper, but the exclusion rates from these evaluations should not be compared directly. Their study also includes the exclusions that result from the uneven implementation of the PMT targeting schemes, while this piece solely addresses exclusions based on the statistical attributes of the model.

Despite this breadth of research, analysis of the accuracy of PMTs for immigrant

subpopulations has been largely absent from this discourse. For migrants in particular, the *Progresa/Oportunidades/Prospera* conditional cash transfer program in Mexico utilizes a PMT model that includes questions on the immigration status of household members, but the focus is on the role of immigration as a determinant of remittances for non-migrant household members (Skoufias, Davis, & Behrman, 1999; Orozco & Hubert, 2005). Similarly, the PMT survey for the Bolsa Familia in Brazil includes questions about internal migration, but does not identify foreign immigrants (Castañeda & Lindhert, 2005). We have found no evaluations that focus on the accuracy of PMTs applied to immigrant populations, either in Latin America or in other regions. This is a particularly glaring omission as the growing ubiquity of PMTs as a development tool has coincided with increased flows of immigrants and refugees around the globe.

In spite of its focus on the Colombian context, this paper will seek to contribute to both regional debates on PMT effectiveness, as well as the discourse on poverty targeting more generally. Globally, refugee and migrant populations are growing in size and increasingly vulnerable. Excluding these populations from social protection programs due to PMT selection methods risks undermining the effectiveness of these programs.

Methodology & Data

The data for this study is drawn from Colombia's *Gran Encuesta Integrada de Hogares* (GEIH) – a nation-wide household survey performed by the *Departamento Administrativo Nacional de Estadística* (DANE), the Colombian national statistics office. This data was compiled from three distinct datasets:

- GEIH Viviendas y Hogares: individual data on household assets

- Medición de Pobreza Monetaria y Desigualidad: household data on income and individual data on education

- GEIH Immigration Data: individual data on birthplace and place of residence 5 years and 1 year ago

The unique identifier variables *Directorio* and *Secuencia_P* were used to link households across the three datasets. In addition, the *Factor de Expansión* variable was used to assign survey weights in the model. The datasets were compiled for 2018 – the most recent year for which data was available.

Income in the GEIH is self-reported, and thus can run into the same issues with misrepresentation that PMTs are designed to avoid. This issue is unavoidable in any analysis of income in the informal sector, but is mitigated in this case by a number of factors. First, the GEIH data is collected by DANE explicitly for research purposes and not for use in any targeting program. This reduces the incentive for participants to lie about their income in order to receive benefits. Second, the GEIH survey is highly

detailed, and explicitly asks for incomes from work, rent, social programs, and other sources, reducing the risk that respondents misremember or misreport their incomes. Missing values are imputed by DANE based on similar households only where strictly necessary.

Two additional key variables, immigration status and education level, were only available at the individual level. For these variables, households were classified based on the status of their household head. Household-head status was self-reported by respondents. The creation of a more detailed household-level index for immigration status or education was considered, but decided against due to risks of researcher bias in the creation of the index.

Migration and citizenship status were also not reported directly and had to be defined by the researcher. The GEIH survey includes questions on the respondents' place of birth, as well as where they lived 5 years and 1 year ago. These questions were used to classify the immigration statuses in Table 1:

Table 1

Migration Status	Birthplace	5 years ago	1 year ago
Colombian	Colombia	Colombia	Colombia
Venezuelan Resident	Venezuela	Colombia	Colombia
Venezuelan Established Migrant	Venezuela	Venezuela	Colombia
Venezuelan Recent Migrant	Venezuela	Venezuela	Venezuela

Unlike Venezuela, Colombia does not grant birthright citizenship, so there is a chance that individuals born in the country do not have citizenship status. Nevertheless, we believe that the "Colombian" status is representative of individuals who have assets, income, and education characteristics that are indistinguishable in aggregate from Colombian citizens. "Venezuelan Residents" were included in order to characterize the impact of any characteristics specific to Venezuelan citizenship, but distinct from immigration status. Venezuelan "Recent" and "Established" immigrants were divided in order to evaluate the potential impact of asset accumulation over time. Other immigration statuses, including individuals who were born in neither Colombia nor Venezuela, were included when calculating the PMT models, but excluded from the overall analysis of model accuracy. These other immigrant categories are heterogeneous, including everything from temporary workers from neighboring South-American countries, to European and American retirees with high wealth and no income. In order to focus the analysis on the question of interest, this paper will focus only on comparisons of Colombians, long-term Venezuelan residents of Colombia, and the two types of Venezuelan immigrants.

Descriptive Analysis

There are significant differences between the different categories of Venezuelan immigrants and their Colombian counterparts in terms of age, education, and income. Table 2 breaks out the number of surveyed individuals in both the training and testing partitions, as well as population statistics by subgroup:

Table 2

Migration Status	Count	Age	Income	Education
Colombian	224,564	49.9	2,293	6.1
Venezuelan Resident	253	39.7***	2,326	6.7
Venezuelan Established Migrant	1,449	32.8***	1,519***	7.4***
Venezuelan Recent Migrant	1,392	33.1***	1,314***	7.3***

Note: *p<0.1; **p<0.05; ***p<0.01

Age is the mean age reported by respondents in years. Income is the median income, reported in thousands of Colombian pesos (1 million Colombian pesos is equivalent to approximately $300 USD at current exchange rates). Education is the mean highest grade level completed. For age and education, the p-values are generated by a Tukey's Test for pairwise differences in ANOVA, compared against Colombians as a baseline. For income, the p-values are generated by a pairwise Mood's median test, again compared against Colombians.

The results of these tests suggest that as a cohort both recent and established Venezuelan immigrants are generally younger, poorer, and better-educated than either Venezuelan-born or native-born Colombian residents. This lends credence to the possibility that Venezuelan immigrants are experiencing occupational downgrading, as they have lower earnings than native Colombians despite higher levels of education. Furthermore, the relative similarity in incomes between Colombians and Venezuelan Residents suggests that there is nothing inherent to being born in Venezuela that is causing this effect. Finally, the difference between incomes for Venezuelan Established Immigrants and Venezuelan Recent Immigrants suggests that there may be a process of adaptation occurring, as Venezuelans are gradually able to establish themselves in the Colombian labor market. The effect of these differences on the PMT model is the focus of the next section.

Model Fit

The analytic approach undertaken for this paper is fairly straightforward. First, I estimate an ordinary least squares PMT model for all Colombian residents and immigrants in the training partition of the dataset (a randomly-selected 25 percent of the

2018 data). The dependent variable for the model is the log of household income, and the independent variables are drawn from both previous research on PMTs (Hanna & Olken) and Colombia's own SISBEN III PMT model for household targeting:

$$ln(Y_{dhi}) = \alpha + \beta_d\theta_d + \beta_h\gamma_h + \beta_i\delta_i + \epsilon_{dhi}$$

In this regression equation, θ_d is a categorical variable for Departamento (the equivalent of states in Colombia). γ_h is a vector of household-level asset and housing quality variables, including household size and access to utilities like electricity and running water. Finally, δ_i is a set of variables representing characteristics of the self-identified household-head, including years of education and health insurance status. An abridged regression output is included in Table 3, and the full list of variables and regression results is available in Appendix Table 1.

Table 3

	Log of Household Income
Household characteristics	
Number of rooms	0.081*** (0.004)
Number of bedrooms	0.037*** (0.007)
Private bathroom	0.063*** (0.017)
Asset Ownership	
Phone line	0.070*** (0.010)
Refrigerator	0.061*** (0.013)
Washing Machine	0.097*** (0.010)
Color television	0.161*** (0.018)
Cable	0.092*** (0.009)
Water heater	0.081*** (0.011)
Microwave	0.123*** (0.011)
Air Conditioning	0.163*** (0.020)
Computer	0.067*** (0.010)
Stereo system	0.064*** (0.008)
Motorcycle	0.118*** (0.009)
Car	0.381*** (0.011)
24-hour water	0.040*** (0.012)
Electricity	0.117 (0.100)
Internet	0.148*** (0.011)
Houshold size	0.106*** (0.003)
Household Head Education Level (None baseline)	
Preschool	0.298 (0.371)

Primary school	0.097*** (0.018)
Secondary school	0.202*** (0.020)
High School	0.236*** (0.019)
University or higher	0.520*** (0.020)
No answer	0.048 (0.249)
Has health Insurance	0.229*** (0.015)
Constant	5.508*** (0.111)
Observations	51,191
R^2	0.423
Adjusted R^2	0.422
Residual Std. Error	6.237 (df = 51099)
F Statistic	411.811*** (df = 91; 51099)

Note: *p<0.1; **p<0.05; ***p<0.01 *Standard errors in parentheses. Other coefficients available in Appendix Table 1.*

The model overall has an adjusted R^2 value of .422. While this is not particularly high, it is well within the range of PMT models that are currently in use in development contexts. The model found statistically significant coefficients on a majority of the variables, including nearly every level of the Department factor. None of the variables included in the model had an unexpected sign or magnitude. Some of the largest and most significant effects came from flooring material in the home, car and motorcycle ownership, and household head education level. Overall, the model was considered to have sufficient predictive power to justify its use in the evaluation of PMT accuracy in this case study.

The model was then used to generate a predicted income for each household in the testing partition. The testing households were then classified as exclusion error, inclusion error, or correctly categorized, based on the relationship between the model-predicted income and the actual status of the household's income relative to the DANE-estimated poverty line. The error rates for Venezuelan groups were compared with Colombians using a simple z-test for difference-in-proportions. Several robustness checks were performed to establish the reliability of the results across different subsets of the data.

Results

PMT Error Rates

The first assessment of this model is a simple comparison of exclusion error rates between different immigrant groups, relative to the poverty line estimated by DANE. The poverty line varies for households based on their family size and estimated costs of basic

nutrition, ranging from 167,000 to 304,000 Colombian Pesos per month (approximately $50-$90 USD at current exchange rates), with an average of 271,000 Colombian Pesos ($80 USD). Households were considered 'Correct' if their PMT-predicted income correctly classified them as above or below their DANE-estimated poverty line. Exclusion errors represent individuals who have actual incomes below the DANE-estimated poverty line, but predicted incomes above the line. Inclusion errors are the converse.

Table 4

Migration Status	Correct	Exclusion Error	Inclusion Error
Colombian	76.1%	16.3%	7.7%
Venezuelan Resident	72.0%	21.1%*	6.7%
Venezuelan Established Migrant	59.1%***	34.2%***	6.7%
Venezuelan Recent Migrant	57.3%***	34.9%***	7.8%

Note: *p$<$0.1; **p$<$0.05; ***p$<$0.01

The rates of exclusion errors are nearly identical for both Established and Recent Venezuelan Immigrants, and both groups have a rate of exclusion error more than twice as high as the error rate for Colombians. Venezuelan residents also have slightly higher rates of exclusion error. This would suggest that the PMT is overestimating the wealth of both Recent and Established Venezuelan Immigrants, but after more than 5 years of residence in Colombia the effect begins to dissipate. The rates of inclusion error were similar across immigration-status groups, justifying our focus on exclusion errors.

To verify these observed differences in exclusion error rates, the exclusion error rates for immigrant groups were compared to Colombians in a pairwise two-sample difference of proportions z-test. The results of these tests are detailed in Table 5, including a 95 percent confidence interval for the difference in proportions:

Table 5 (Exclusion Errors)

Migration Status	Errors	Error Rate	95% CE	P-Value
Colombian	27,437	16.3%	–	–
Venezuelan Resident	41	21.1%	[-1%, 11%]	0.078*
Venezuelan Established Migrant	365	34.2%	[15%, 21%]	<0.000***
Venezuelan Recent Migrant	373	34.9%	[16%, 22%]	<0.000***

Note: 95% Confidence Interval represents difference vs Colombians.

Since these results are statistically significant at the .01 significance level, we can conclude that Recent and Established Venezuelan immigrants below the poverty line have higher rates of exclusion from the PMT model relative to Colombians. Venezuelan Residents who have been in Colombia for at least 5 years also have slightly higher rates of errors, but this difference is only significant at the p$<$0.1 level. In terms of practical implications, while a 16 percent exclusion error rate may be acceptable for the PMT

model in general, this model excludes approximately one in three Venezuelan immigrants below the poverty line threshold.

Robustness Check 1: Near-Poverty Error Rates

We might be concerned that these error rates are simply driven by the fact that there are a higher share of immigrants clustered around the poverty line, which would mechanically increase the probability of exclusion errors. This is further reinforced by our previous finding that Colombians have higher median incomes than Venezuelan immigrants. In order to address this issue, the analysis was rerun using a subset of the data restricted to only cases that have actual income less than 200 percent above the poverty line.[1] This resulted in the subset in Table 6, which captures approximately the bottom quartile to third of the income distribution — in line with other PMT targeting thresholds.

Table 6

Migration Status	N	Share of Test Data	Mean Income
Colombian	33,105	19.7%	508 K
Venezuelan Resident	45	23.3%	575 K**
Venezuelan Established Migrant	325	30.4%	556 K***
Venezuelan Recent Migrant	401	37.5%	534 K***

Note: *p<0.1; **p<0.05; ***p<0.01; *two-sample difference of mean income vs Colombians*

These results also indicate that concerns about the distribution of incomes across immigrant groups was well-founded. Colombians within the specified income range are a much smaller share of the overall test data compared to the immigrant populations. Once this subset is taken, however, the mean income across groups is comparable, suggesting that income distribution effects have been mitigated. The subset was then assessed using the same methods as the test-data PMT:

Table 7

Migration Status	Errors	Error Rate	95% CE	P-Value
Colombian	9,415	28.4%	–	–
Venezuelan Resident	23	51.1%	[7%, 38%]	0.001***
Venezuelan Established Migrant	184	56.6%	[23%, 34%]	<0.000***
Venezuelan Recent Migrant	222	55.4%	[22%, 32%]	<0.000***

Note: *p<0.1; **p<0.05; ***p<0.01

---

[1]Attempts were made to take an even smaller subset of incomes as a robustness check, but the resulting sample sizes were significantly smaller, resulting in concerns about the power of the resulting tests.

As expected, the error rates in Table 7 are higher across the board for the restricted dataset than for the whole income distribution. Nevertheless, the rates are still significantly higher for the three Venezuelan groups. This suggests that even within this restricted cohort, there are systemic factors that are reducing the accuracy of the PMT model for Venezuelan immigrants. Furthermore, these results imply that for immigrants with incomes near the poverty line, the PMT model is actually more inaccurate than accurate. With exclusion error rates higher than 50 percent, we would expect that a series of fair random coin-flips could categorize immigrants at a roughly similar rate to the PMT model. In practical terms, this inaccuracy would result in high levels of exclusion of immigrants from social programs that rely on PMT targeting.

Robustness Check 2: Work-Status Error Rates

As a further check, we estimated the amount of variation in the error rates that is attributable to different rates of employment among different immigration subgroups. The GEIH asks respondents how they spent a majority of their time in the previous week, with a majority of cases reporting that they were working or looking for work. As illustrated in Table 7, Colombians are less likely to be working, but also less likely to be looking for work compared with Venezuelans in Colombia. They also have a higher share in the 'Other' category, which includes students, homemakers, and the disabled.

<div align="center">Table 8</div>

Migration Status	Working	Looking for Work	Other
Colombian	62.8%	2.3%	34.9%
Venezuelan Resident	76.7%***	4.7%**	18.7%***
Venezuelan Established Migrant	80.8%***	3.3%**	15.9%***
Venezuelan Recent Migrant	80.9%***	5.4%***	13.7%***

*Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$; Two-sample difference of proportions vs Colombians*

In order to mitigate the effect of these differential work statuses on our evaluation of the PMT's accuracy, we restricted the exclusion error analysis to only include self-identified workers. Furthermore, we can go beyond work status to restrict the analysis to individuals who have income. In this data "zero income" is different from "not working," because most non-workers (including students, homemakers, and the disabled) have some source of income, such as rents or transfer programs. The PMT does a very poor job of prediction for individuals with zero income, due to the relatively high intercept term in the model and the fact that every case in the data had at least some housing attributes or asset ownership. Table 8 documents exclusion error rates for various combinations of these data subsets, including the worker-only, positive income, and near-poverty (NP)

subsets. As before, p-values are based on pairwise difference in proportions tests against Colombians:

Table 9

Migration Status	Workers	NP Workers	NP w/ Income
Colombian	14%	47%	50%
Venezuelan Resident	19%*	52%	51%
Venezuelan Established Migrant	31%***	54%**	56%**
Venezuelan Recent Migrant	31%***	51%	55%**

Note: *p<0.1; **p<0.05; ***p<0.01

These results suggest that near the poverty line the model is less accurate for workers, and at least some of the low error rates for near-poverty Colombians are attributable to the correct classification of Colombian non-workers. This implies that at least some proportion of the differences in model accuracy is attributable to the higher share of workers in the Venezuelan immigrant populations. This is encouraging, as it seems like rather than being biased against immigrant groups in particular, the model is biased against workers near poverty, a group in which immigrants are over-represented. Unfortunately, while this seems to resolve some of the inaccuracy, it is likely not feasible to include work status in an actual PMT because it would be very easy for respondents to misrepresent their occupational status. Ideally, there would be another observable factor that could be used to decrease exclusion error rates for immigrant populations without relying on easily misrepresented information.

Identifying the Occupational Downgrading Mechanism

In order to isolate the educational downgrading mechanism, we looked at the model's accuracy across different educational strata. Overall, Venezuelan immigrants have higher levels of education relative to Colombians, with a lower share having a primary school or lower education and a higher share with *media* (10$^{\text{th}}$ and 11$^{\text{th}}$ grade) education:

Table 10

Migration Status	≤Primary	Secondary	Media	≥University
Colombian	34%	13%	26%	27%
Venezuelan Resident	16%***	20%	32%*	38%***
Venezuelan Established Migrant	9%***	20%***	41%***	30%***
Venezuelan Recent Migrant	9%***	20%***	40%***	31%***

Note: *p<0.1; **p<0.05; ***p<0.01; *Two-sample difference of proportions vs Colombians*

To roughly assess the returns to education for different immigrant and non-immigrant groups, we can look at a regression of the interaction of immigration and education on

income using Colombians as a baseline, controlling for age in order to mitigate the effect of age-differences between the cohorts:

$$ln(Y_{abc}) = \alpha + [\beta_a(EDUCATION) * \beta_b(IMMIGRATION)] + \beta_c AGE + \epsilon_{abc}$$

We can see in Table 11 that the overall positive coefficient for education level has a weakly negative interaction with Venezuelan Established Immigrants and significantly negative interaction with Venezuelan Recent Immigrants.

<div align="center">

Table 11[2]
</div>

	Log Household Income
Education Level	0.328*** (0.001)
Age (Years)	0.014*** (0.000)
Education Level*Venezuelan Resident	0.051 (0.066)
Education Level*Venezuelan Established Migrant	−0.059* (0.029)
Education Level*Venezuelan Recent Migrant	−0.087** (0.028)
Constant	5.196*** (0.012)
R^2	0.178
Adjusted R^2	0.178

Note: *p<0.1; **p<0.05; ***p<0.01

This suggests that occupational downgrading is indeed occurring within the data: returns to education are lower for Venezuelan immigrants than they are for Colombians. In terms of the PMT model, if we compare rates of exclusion errors across both immigration and education-level subcategories, Table 12 finds lower rates of exclusion error among university graduates across all immigration categories:

<div align="center">

Table 12
</div>

Migration Status	≤Primary	Secondary	Media	≥University
Colombian	22%	20%	17%	6%
Venezuelan Resident	37%*	21%**	28%**	10%
Venezuelan Established Migrant	34%**	42%***	40%***	22%***
Venezuelan Recent Migrant	40%***	42%***	38%***	24%***

Note: *p<0.1; **p<0.05; ***p<0.01; *Two-sample difference of proportions vs Colombians*

This effect is particularly strong for Colombians. For both Recent and Established Venezuelan immigrants, exclusion errors at the university level are still equal or higher than the exclusion rates for Colombians at any education level. Furthermore, these

---

[2]Other immigration statuses were also included for completeness, but were excluded from the regression output.

differences between the rates are higher as we move up education levels. This implies that the observed occupational downgrading is indeed increasing the inaccuracy of the PMT model for Venezuelan immigrants.

Modified PMT Model with Immigration Status Term

A possible solution to this issue is to simply include immigration status in the PMT model itself. To test this approach, the previous regression model was modified to include the β_kIMMIGRATION term – a factor variable for immigration status:

$$ln(Y_{dhik}) = \alpha + \beta_d\theta_d + \beta_h\gamma_h + \beta_i\delta_i + \beta_k\text{IMMIGRATION} + \epsilon_{dhik}$$

The full model result is available in Appendix Table 2, and relevant coefficients for the immigration-status factor variable are outlined in Table 13, relative to the Colombian reference category:

Table 13

	Log Household Income
Venezuelan Resident	0.130 (0.107)
Venezuelan Established Migrant	0.273*** (0.049)
Venezuelan Recent Migrant	0.334*** (0.050)
Controls from PMT Model	Yes
R²	0.425
Adjusted R²	0.423
Residual Std. Error	6.230 (df = 51091)

Note: *p<0.1; **p<0.05; ***p<0.01

The overall predictive strength of the model has barely increased: Adjusted R^2 has gone up by .001, and RSE has gone down by just .007 relative to the model without immigration. The specific coefficients for Venezuelan immigrants, however, are statistically significant at the .01 level. In terms of magnitudes, the coefficient on Recent Immigrants is comparable to owning a car, and the coefficient on Established Immigrants is comparable to having health insurance. It is unclear from this result exactly why the coefficients are positive — the most plausible scenario is that immigrants occupy a uniquely underprivileged position in regards to their combination of housing and asset variables, so their average income conditional on the other PMT variables requires a positive correction. According to Table 14, due to this low increase in predictive power, the revised model does almost nothing to decrease the level of bias in the exclusion error rates.

14

Table 14

Migration Status	Errors	Error Rate	95% CE	P-Value
Colombian	17,433	16.3%	–	–
Venezuelan Resident	41	21.2%	[-1%, 11%]	0.077*
Venezuelan Established Migrant	366	34.3%	[15%, 21%]	<0.000***
Venezuelan Recent Migrant	375	35.1%	[16%, 35%]	<0.000***

Note: *p<0.1; **p<0.05; ***p<0.01

This result implies that the bias in the model cannot be corrected by simply including immigration status in the model, but requires a reevaluation of the model itself. Optimization to remove this bias will likely require reworking the set of variables included in the model, or providing a post-hoc correction to the PMT predictions for immigrant populations. While this likely involves significant data gathering and pre-testing on the part of the PMT designer, it may be the only way to ensure that vulnerable populations are not excluded from programs in an unfair manner due to statistical errors.

Policy Prescriptions

The assessment in this paper is a case study of a particular choice of PMT model, applied to a specific threshold in Colombia, compared against a unique subpopulation. That said, it is possible to use this case to draw two general conclusions regarding the distributional equity of PMTs:

1. Policymakers should be cautious in the implementation of PMTs, avoiding them if at all possible.

If the results of this paper were to hold in the implementation of a nationwide PMT program in Colombia, not only would the PMT erroneously exclude significant numbers of Colombians, significant numbers of Venezuelan immigrants would also be excluded due to statistical bias. As a rough calculation, UNHCR estimates that by the end of 2018 there were approximately 1.2 million displaced Venezuelans in Colombia; the 34 percent exclusion error rate found in this model could leave more than 400,000 individuals without benefits and services to which they are entitled. There is a robust evidence base supporting both universal programs and other non-statistical targeting methods, which should be considered alongside PMTs wherever possible (Kidd & Athias, 2019).

2. If PMTs must be used, they should be checked for accuracy within sub-populations prior to implementation.

Given the demonstrated possibility for unequal coverage of vulnerable sub-populations within a PMT model, it is important that the evaluation of PMTs include estimates of coverage rates and exclusion errors prior to implementation. This paper proposes a methodology for the evaluation of PMTs for sub-populations that can be drawn upon in

other country contexts and for other gender, disability, or ethnic sub-populations. The comparison of exclusion error rates against a baseline category is only a single dimension of PMT bias, but it could be easily identified using this type of pre-analysis. Once identified, it is incumbent upon policymakers to design targeting programs that mitigate or avoid these types of biases.

Conclusion

In addition to these recommendations, this case study also identifies several areas for future research. There is ample room to test for this type of PMT exclusion bias across different models, country contexts, and sub-populations to develop a sense of its extent. In terms of immigrant sub-populations, it will be important to see how PMT models hold up over time. PMT models are typically recalculated once every several years, so it is an open question as to whether a model that may be unbiased based on a single year's results is able to maintain accuracy in the face of a large wave of immigration. Providing an alternative approach to PMT modeling was outside the scope of this paper, so it is incumbent on future research to determine how to best excise exclusion bias from targeting approaches. Finally, as a general project it is important to continue to probe the limitations and biases of PMT models in order to advance a more just and equitable paradigm for economic development and social protection.

Appendix

Appendix Table 1: Baseline PMT Model

	Log of Household Income
Department (Antioquia baseline)	
Atlántico	0.056*** (0.021)
Bogota	0.075*** (0.013)
Bolívar	−0.008 (0.023)
Boyacá	−0.095*** (0.023)
Caldas	−0.100*** (0.026)
Caquetá	−0.102*** (0.038)
Cauca	−0.248*** (0.024)
Cesar	−0.237*** (0.028)
Córdoba	−0.100*** (0.028)
Cudinamarca	0.128*** (0.020)
Chocó	−0.384*** (0.048)
Huila	−0.203*** (0.027)
La Guajira	−0.081** (0.040)
Magdalena	−0.095*** (0.031)
Meta	−0.022 (0.029)
Nariño	−0.166*** (0.023)
Norte de Santander	−0.116*** (0.026)
Quindío	−0.083*** (0.026)
Risaralda	−0.046* (0.024)
Santander	−0.025 (0.020)
Sucre	−0.011 (0.034)
Tolima	−0.064*** (0.024)
Valle del Cauca	−0.072*** (0.015)
Home Ownership Status (Self-owned baseline)	
Mortgaged	−0.018 (0.020)
Rented	−0.255*** (0.009)
Usufruct	−0.116*** (0.011)
Squatting or collectively owned	−0.299*** (0.025)
Other	−0.559*** (0.108)

Cooking Fuel (Electricity baseline)	
Gasoline or kerosene	0.133 (0.193)
Natural gas lines	−0.110*** (0.028)
Natural gas or propane tanks	−0.122*** (0.029)
Wood or wood charcoal	−0.346*** (0.034)
Mineral charcoal	0.154 (0.146)
Waste materials	−0.528 (0.388)
Room where food is prepared (Kitchen baseline)	
Bedroom	0.011 (0.033)
Dining room with sink	0.010 (0.018)
Dining room without sink	−0.021 (0.041)
Outside	−0.002 (0.036)
Type of Sewage System (Sewer connection baseline)	
Septic tank	−0.072*** (0.014)
No connection	−0.208*** (0.028)
Latrine	−0.214*** (0.076)
Tidal flows	0.104 (0.067)
Water Source (Indoor central plumbing baseline)	
Indoor plumbing from other source	−0.198*** (0.070)
Well with pump	−0.110 (0.112)
Well without pump	−0.282** (0.119)
Rainwater	−0.119 (0.075)
River or spring	0.245** (0.110)
Public well	0.124 (0.147)
Delivery truck	−0.034 (0.126)
Delivery person	0.152** (0.076)
Bottled or bagged water	0.118*** (0.028)
Exterior Wall Material (Brick baseline)	
Polished wood	−0.454*** (0.075)
Adobe	−0.082*** (0.030)
Wood and mud-brick	−0.074*** (0.027)
Rough wood	−0.104*** (0.030)
Bamboo	0.597*** (0.162)
Cane or other plant matter	−0.056 (0.082)
Metal or scrap	0.156 (0.099)
No walls	0.720** (0.348)

Floor Material (Dirt baseline)	
Cement	0.025 (0.030)
Rough wood	0.107*** (0.038)
Brick or tile	0.195*** (0.031)
Marble	0.576*** (0.068)
Polished wood	0.627*** (0.044)
Carpet	0.202** (0.095)
Household characteristics	
Number of rooms	0.081*** (0.004)
Number of bedrooms	0.037*** (0.007)
Private bathroom	0.063*** (0.017)
Asset Ownership	
Phone line	0.070*** (0.010)
Refrigerator	0.061*** (0.013)
Washing Machine	0.097*** (0.010)
Color television	0.161*** (0.018)
Cable	0.092*** (0.009)
Water heater	0.081*** (0.011)
Microwave	0.123*** (0.011)
Air Conditioning	0.163*** (0.020)
Computer	0.067*** (0.010)
Stereo system	0.064*** (0.008)
Motorcycle	0.118*** (0.009)
Car	0.381*** (0.011)
24-hour water	0.040*** (0.012)
Electricity	0.117 (0.100)
Internet	0.148*** (0.011)
Houshold size	0.106*** (0.003)
Household Head Education Level (None baseline)	
Preschool	0.298 (0.371)
Primary school	0.097*** (0.018)
Secondary school	0.202*** (0.020)
High School	0.236*** (0.019)
University or higher	0.520*** (0.020)
No answer	0.048 (0.249)
Has health Insurance	0.229*** (0.015)
Constant	5.508*** (0.111)

Observations	51,191
R^2	0.423
Adjusted R^2	0.422
Residual Std. Error	6.237 (df = 51099)
F Statistic	411.811*** (df = 91; 51099)

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix Table 2: Immigration-Adjusted PMT Model

	Log of Household Income
Migration Status (Colombian Baseline)	
Colombian Other Returnee	0.144*** (0.046)
Colombian Venezuelan Returnee	0.099** (0.042)
Other Migrant	−0.656*** (0.125)
Other Resident	0.235*** (0.084)
Undefined	0.901** (0.399)
Venezuelan Established Migrant	0.273*** (0.049)
Venezuelan Recent Migrant	0.334*** (0.050)
Venezuelan Resident	0.130 (0.107)
Department (Antioquia baseline)	
Atlántico	0.054*** (0.021)
Bogota	0.079*** (0.013)
Bolívar	−0.008 (0.023)
Boyacá	−0.090*** (0.023)
Caldas	−0.098*** (0.026)
Caquetá	−0.096** (0.037)
Cauca	−0.243*** (0.024)
Cesar	−0.237*** (0.028)
Córdoba	−0.099*** (0.028)
Cudinamarca	0.129*** (0.020)
Chocó	−0.382*** (0.048)
Huila	−0.202*** (0.027)
La Guajira	−0.086** (0.040)
Magdalena	−0.096*** (0.031)
Meta	−0.018 (0.029)

Nariño	−0.158*** (0.023)
Norte de Santander	−0.123*** (0.026)
Quindío	−0.083*** (0.026)
Risaralda	−0.048** (0.024)
Santander	−0.025 (0.020)
Sucre	−0.016 (0.034)
Tolima	−0.059** (0.024)
Valle del Cauca	−0.072*** (0.015)
Home Ownership Status (Self-owned baseline)	
Mortgaged	−0.019 (0.020)
Rented	−0.257*** (0.009)
Usufruct	−0.113*** (0.011)
Squatting or collectively owned	−0.294*** (0.025)
Other	−0.552*** (0.108)
Cooking Fuel (Electricity baseline)	
Gasoline or kerosene	0.139 (0.193)
Natural gas lines	−0.109*** (0.028)
Natural gas or propane tanks	−0.122*** (0.029)
Wood or wood charcoal	−0.344*** (0.034)
Mineral charcoal	0.148 (0.146)
Waste materials	−0.526 (0.387)
Room where food is prepared (Kitchen baseline)	
Bedroom	0.006 (0.033)
Dining room with sink	0.008 (0.018)
Dining room without sink	−0.026 (0.041)
Outside	0.001 (0.036)
Type of Sewage System (Sewer connection baseline)	
Septic tank	−0.072*** (0.014)
No connection	−0.207*** (0.028)
Latrine	−0.216*** (0.076)
Tidal flows	0.098 (0.067)
Water Source (Indoor central plumbing baseline)	
Indoor plumbing from other source	−0.201*** (0.069)
Well with pump	−0.111 (0.112)
Well without pump	−0.284** (0.119)
Rainwater	−0.119 (0.075)
River or spring	0.244** (0.110)
Public well	0.125 (0.146)
Delivery truck	−0.027 (0.126)

Delivery person	0.144* (0.076)
Bottled or bagged water	0.115*** (0.028)
Exterior Wall Material (Brick baseline)	
Polished wood	−0.466*** (0.075)
Adobe	−0.078*** (0.030)
Wood and mud-brick	−0.072*** (0.027)
Rough wood	−0.100*** (0.030)
Bamboo	0.597*** (0.162)
Cane or other plant matter	−0.043 (0.081)
Metal or scrap	0.160 (0.099)
No walls	0.711** (0.348)
Floor Material (Dirt baseline)	
Cement	0.021 (0.030)
Rough wood	0.106*** (0.038)
Brick or tile	0.187*** (0.031)
Marble	0.561*** (0.068)
Polished wood	0.615*** (0.044)
Carpet	0.196** (0.094)
Household characteristics	
Number of rooms	0.081*** (0.004)
Number of bedrooms	0.038*** (0.007)
Private bathroom	0.067*** (0.017)
Asset Ownership	
Phone line	0.070*** (0.010)
Refrigerator	0.068*** (0.013)
Washing Machine	0.101*** (0.010)
Color television	0.172*** (0.018)
Cable	0.094*** (0.009)
Water heater	0.079*** (0.011)
Microwave	0.123*** (0.011)
Air Conditioning	0.165*** (0.020)
Computer	0.069*** (0.010)
Stereo system	0.067*** (0.008)
Motorcycle	0.121*** (0.009)
Car	0.382*** (0.011)
24-hour water	0.040*** (0.012)
Electricity	0.105 (0.100)
Internet	0.146*** (0.010)
Household size	0.103*** (0.003)

Household Head Education Level (None baseline)	
Preschool	0.322 (0.371)
Primary school	0.095*** (0.018)
Secondary school	0.195*** (0.020)
High School	0.230*** (0.019)
University or higher	0.510*** (0.020)
No answer	0.042 (0.249)
Has health Insurance	0.270*** (0.016)
Constant	5.463*** (0.111)
Observations	51,191
R^2	0.423
Adjusted R^2	0.422
Residual Std. Error	6.237 (df = 51099)
F Statistic	411.811*** (df = 91; 51099)

Note: *p<0.1; **p<0.05; ***p<0.01

References

Attanasio, O, & Mesnard, A. (2006). The Impact of a Conditional Cash Transfer Programme on Consumption in Colombia. *Journal of Applied Public Economics, Institute for Fiscal Studies*

Banerjee, A. Hanna, R. Olken, B. & Sumarto, S. (2018). The (lack of) Distortionary Effects of Proxy-Means Tests: Results from a Nationwide Experiment in Indonesia. *NBER Working Paper No. w25362.*

Camacho, A. & Conover, E. (2011). Manipulation of Social Program Eligibility *American Economic Journal: Economic Policy.*

Castañeda, T. (2005). Targeting Social Spending To The Poor With Proxy–Means Testing: Colombia's SISBEN System. *Social Protection Discussion Paper Series, The World Bank.*

Castañeda, T. & Lindert, K. (2005). Designing and Implementing Household Targeting Systems: Lessons from Latin American and The United States. *Social Protection Unit, Human Development Network, The World Bank.*

Grosh, M. (1994) Administering Targeted Social Programs in Latin America: From Platitudes to Practice *The International Bank for Reconstruction and Development, The World Bank*

Grosh, M. & Baker, J. (1995) Proxy Means Tests for Targeting Social Programs: Simulations and Speculation. *Living Standards Measurement Survey Working Paper, The World Bank*

Grosh, M., Del Ninno, Carlo., Tesliuc, E. & Ouerghi, A. (2008). For Protection and Promotion: The Design and Implementation of Effective Safety Nets. *The International Bank for Reconstruction and Development, The World Bank*

Hanna, R. & Olken, B. (2019). Universal Basic Incomes vs. Targeted Transfers: Anti-Poverty Programs in Developing Countries. *NBER Working Paper Series.*

Johannsen, J. (2006). Operational Poverty Targeting in Peru – Proxy Means Testing with Non-Income Indicators. *United Nations Development Programme, International Poverty Centre Working Paper.*

Kidd, S. & Athias, D. (2019). Hit and Miss: An assessment of targeting effectiveness in social protection *Working Paper, Development Pathways.*

Kidd, S. & Wylde, E. (2011). Targeting the Poorest: An assessment of the proxy means test methodology *AusAID.*

Klasen, S. & Lange, S. (2015). Accuracy and Poverty Impacts of Proxy Means-Tested Transfers: An Empirical Assessment for Bolivia. *Courant Research Centre, Georg-August-Universität Göttingen.*

Orozco, M. & Hubert, C. (2005). La Focalizacion En El Programa De Desarrollo Humano Oportunidades De Mexico *Social Protection Unit, The World Bank.*

Skoufias, E., Davis, B. & Behrman, J. (1999). An Evaluation of the Selection of Beneficiary Households in the Education, Health, and Nutrition Program (PROGRESA) of Mexico. *International Food Policy Research Institute*

UNHCR. (2019). The 2018 Global Trends Report. *United Nations High Commissioner for Refugees.*

Vélez, C., Castaño, E. & Deutsch, R. (1998). An Economic Interpretation of Colombia's SISBEN: A Composite Welfare Index Derived from the Optimal Scaling Algorithm. *Poverty and Inequality Advisory Unit, The Interamerican Development Bank.*

World Bank. (2017). Closing the Gap: The State of Social Safety Nets, 2017. *World Bank Group.*