Evaluating and Improving Proxy Selection Frameworks for Welfare Schemes

December 1, 2024

ECO-3500 Development Economics

Professor Bharat Ramaswami

We sincerely thank Professor Bharat for his invaluable guidance in conceptualizing the research framework and refining our methodology. We also extend our gratitude to our TFs, Rohit and Mritunjay, for their assistance with data files, technical support, and insightful feedback throughout the project.

1 Introduction

Welfare policies in developing countries often rely on outdated poverty line definitions for social programs, resulting in inaccurate targeting. Schemes targeting the bottom 40% of the population run into problems because income and consumption data used for these calculations are often misreported, hidden, or miscalculated. Mis-targeting in these schemes not only leads to resource wastage, but also results in significant inclusion and exclusion errors, undermining the schemes' effectiveness and reach. Since consumption is a better indicator of welfare, and households have the tendency to smooth consumption over time (unlike income), we would ideally want to observe a household's consumption expenditure to identify those eligible for social programs. However, given that consumption information is self-reported, it is not a very observable or verifiable metric for large-scale policies.

A way to overcome this is to use verifiable proxies that closely reflect consumption expenditure of households; this is exactly what a Proxy Means Test (PMT) does. Instead of income data, as used by the means test, PMTs use certain characteristics of households such as their possession of durable goods, and employment and residence categories as proxies to estimate their consumption and the corresponding decile they fall into.

The expenditure and behavior of people change over time, especially as income increases and trends change. Characteristics that previously indicated higher income may become easily accessible or even basic necessities as GDP and socio-economic conditions improve. For example, access to smartphones or internet—once markers of relative affluence—may no longer effectively distinguish between households in need and those that are not. These changes make it challenging to rely on static proxies for revising beneficiary lists, leading to potential errors in accurately identifying eligible households. Instead of a standard set of proxies, there is a need for a framework that gives the optimal set of proxies which best reflect household consumption expenditure.

2 Problem Statement

In this paper, we aim to find a framework that gives the optimal number and set of proxies that best predict household consumption expenditure. The framework's objective is to maximize the accuracy of the estimation (given by adjusted R square or $\bar{R}^2$), and minimize inclusion errors and exclusion errors. Then, we see whether these proxies can be used as time invariant estimates and if they would give accurate results even a decade later, given changes in environment, living standards, governance, technology and globalization. We also explore the changes in inclusion and exclusion errors with respect to changes in the poverty line.

Section 3 of the paper is a literature review that evaluates the effectiveness and shortcomings of existing PMT mechanisms. It delineates our motivation to overcome some of these drawbacks by introducing a novel framework that selects proxies to optimize the adjusted R-squared and better capture household consumption. *Section 4* defines inclusion and exclusion errors associated with PMTs and explains how our paper aims to minimize them. *Section 5* lays out the analytical framework and data used. *Section 6* presents the results of the optimal PMT model, its comparison to existing PMTs, and the changes in inclusion and exclusion errors with respect to different poverty lines. *Section 7* highlights the scope for future research and concludes the paper.

3 Motivation

PMTs have a rich literature, with a focus on targeting in developing countries. Grosh and Baker show that household characteristics can serve as reasonable proxies in assessing eligibility for social programs (Grosh and Baker). They further show that more information is generally better than less, though there are diminishing returns. Their model goes through a range of proxy sets: location, housing, ownership of durables, etc., which are chosen based on a high degree of correlation with consumption expenditure, and not with the goal of maximizing adjusted R square values. A similar approach is used by Mahamalik and Sahu where they create 6 binary proxies for possession of dwelling unit and land, caste, primary agriculture employment, and whether annual per capita expenditure on clothes is below Rs 216.29 (Mahamallik and Sahu). However, even their selection is based on a high correlation of these proxies with possession of Below Poverty Line (BPL) cards.

In a study using household survey data from Bangladesh, Indonesia, Rwanda and Sri Lanka, Kidd and Wylde findings show that the PMT is inherently inaccurate, especially at low levels of coverage, and it arbitrarily selects beneficiaries (Kidd and Wylde). They claim that it functions more like a simple rationing mechanism that selects some poor and non-poor but excludes large numbers of eligible poor from receiving benefits. Their argument relies on using R-squared as a measure of accuracy, which might mislead results. If we decide to evaluate the goodness of fit on R square instead, we will not be able to distinguish between variables that are good predictors and those that are not. Since $\bar{R}^2$ penalizes the addition of irrelevant variables, it's easy to observe when we should stop adding regressors. Hence, we improve on this research by focusing on adjusted R-square in our upcoming methodology.

With respect to the number of proxies, Klassen and Lange suggest that increasing the number of proxies (i.e. going from parsimonious models to models employing more proxies) exhibits quickly decreasing returns in terms of accuracy (Klasen and Lange). More importantly, parsimonious PMTs based on few easy-to-verify proxies such as geography and demographics perform worse in terms of targeting accuracy yet outcomes do not differ much from more sophisticated PMTs when it comes to poverty effects.

Given that there is high government intervention in India and a concerted effort to reduce poverty through various schemes like Pradhan Mantri Jan Dhan Yojana (PMJDY), Mahatma Gandhi National Rural Employment Guarantee Act (MGNREGA), and Public Distribution System (PDS), the need for effective targeting in India is crucial. Resources are scarce to begin with, but their distribution is further hampered by corruption, bureaucracy, and inadequate data. This makes it more essential to have accurate targeting mechanisms in India to ensure the best possible outcome for poverty alleviation. Our study aims to assess the effectiveness of the Proxy Means Test and provide a framework that would constitute an ideal targeting mechanism.

Our contribution is novel as it suggests an ideal, implementable framework for choosing the right proxies which will optimize targeting. We show that changing the goal of PMT models –from high degree of correlation to maximizing adjusted R-square (henceforth referred to as $\bar{R}^2$)–reduces errors and better predicts household consumption expenditure for the poor as compared to the above mentioned ones.

4 Inclusion and Exclusion Errors in PMTs

As we use a PMT instead of actual consumption data, it might be the case that some people who are eligible for the program (who truly fall in the bottom 40% bracket) may be misidentified as ineligible based on the proxies. This may happen if enumerators are not always objective during surveys and may lack the time to verify proxies. Or, in light of crises and shocks faced by households, households that fall into poverty but do not suffer a related change in the household characteristics and assets used as proxies may be excluded from receiving benefits. Such cases lead to exclusion errors or undercoverage, where eligible individuals fail to receive program benefits. This undercoverage makes the program ineffective in changing the welfare level of the intended beneficiaries, but it carries no budgetary cost. The other case of targeting error occurs when a person's "true" welfare level is above the cut-off but their predicted welfare is below it. Incorrectly identifying people as being eligible for program benefits is called exclusion error or leakage.

To address undercoverage, we increase the number of proxies to ensure that no single proxy dominates the classification process. As we include more variables, especially dynamic consumption variables like food habits or participation in subsidized programs, the $\bar{R}^2$) improves. This helps account for nuanced changes in household welfare over time, reducing the chances of excluding households experiencing recent income shocks. The model better captures their current welfare status by incorporating dynamic consumption indicators which highly correlate to a fall in incomes, such as the household's shift from consuming "costly dals" (like arhar and tur) to "poor dals," (like moong and urd), substituting ghee with vanaspati oil or reducing consumption of expensive, seasonal vegetables like spinach and lemons, and increasing consumption of basic vegetables like potatoes and onions. Adding more consumption variables balances proxies that reflect past wealth with those that indicate present conditions.

Other PMT models usually do not use consumption expenditure as proxies because they are self-reported and unverifiable values. In our model, instead of including expenditure on certain unobservable variables, we include the type of consumable goods which the household possesses. For example, instead of including the total household expenditure on pulses, we create a variable for costly pulses. This is as easy to observe as any durable. We argue that it is easier to verify because, with the knowledge of an evaluator coming for inspection, one might hide their durable goods (say, give their TV to their neighbors), but they cannot possibly hide or change their entire month's ration of pulses, grains and oil. Although this method slightly increases leakage, it prioritizes accurately identifying poor households, a critical goal in developing economies. Moreover, by minimizing undercoverage and optimizing the $\bar{R}^2$, our framework reduces inclusion and exclusion errors, providing a robust targeting solution.

5 Analytical Framework and Data

We first draw the PMT model from existing literature and use those proxies on 2011 Household Consumer Expenditure Survey (HCES) Data from the National Sample Survey (NSS 2010-11). Since the selection of these proxies is very subjective, we propose a more robust and systematic method to optimize the choice of proxies which will give the highest accuracy and predictive power. We start by choosing a single covariate amongst the 140 variables that give the highest $\bar{R}^2$. The first chosen covariate is noted and appended to the regression model. In the next iteration we choose the covariate which gives the next highest $\bar{R}^2$ from the remaining 139 covariates. At each run, the chosen covariate is appended to the model until the $\bar{R}^2$ starts to fall. Eventually, we get a final set of covariates which maximise the $\bar{R}^2$ for all possible combinations of consumption items. Further in the paper, we'll refer to this model as the $\bar{R}^2$ *maximising model*. We then predict the consumption value and distribution of household consumption expenditure for 2022, using HCES data from NSS 2022-23, controlling for inflation.

6 Methodology and Results

6.1 Finding the optimal list of covariates

As mentioned in *Section 5*, by using the $\bar{R}^2$ *maximizing model* on 2011 data we find that $\bar{R}^2$ maximizes at the 73^{rd} step, after which it starts reducing. Figure 1 shows how the $\bar{R}^2$ progresses as the next best predictor is appended to the model. As can be seen, the graph is increasing albeit at a decreasing rate, and maximizes at the 73^{rd} step after which adding another variable will only reduce the model's $\bar{R}^2$ (the slope's shift isn't clearly visible as the decrease is incredibly small).

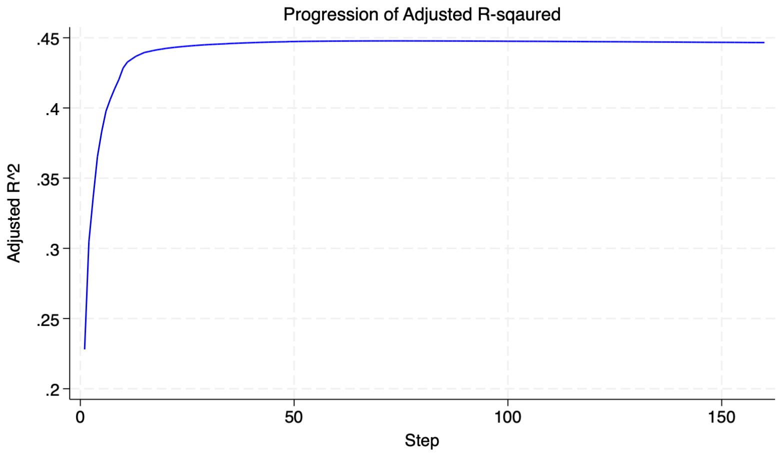


Figure 1: Progression of $\overline{R}^2$ as Variables as Sequentially Added to the Model

Figure 2 shows the trend of the change in $\bar{R}^2$, or the incremental increase (and eventual decrease) in the model's predictive capability of income. The highest $\bar{R}^2$ according to the model was reported to be 0.447. There comes a point after the 30^{th} iteration beyond which the marginal increase in $\bar{R}^2$ becomes extremely insignificant after 30 variables. Hence, we use only 30 covariates in the comparisons that ensue post this section.

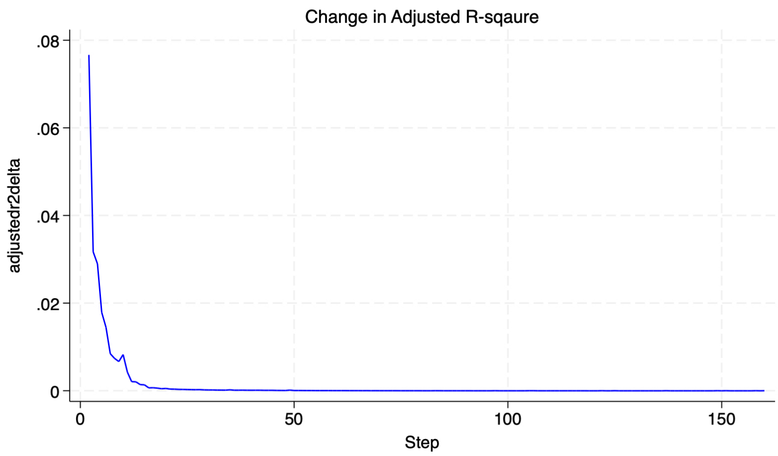


Figure 2: $\overline{R}^2$ Delta

6.2 Modifying the Model

After going through the list of these 30 covariates, we replaced those that are unobservable– like total expenditure on pulses, leather footwear, school uniforms, toiletries etc. This gives us the set of optimal proxies that maximise $\bar{R}^2$ and the predictive capability of consumption expenditure for the model. This regression provides the covariates and coefficents for 2011, which then become a base to predict 2022 consumption expenditure for households. The selection of largely observable household consumption characteristics lead to the regression in Table 1, which has an $\bar{R}^2$ value of 0.422.

This statistic in isolation does not tell us much about whether we have actually been able to improve predictions of expenditure. It's therefore crucial that we compare the $\bar{R}^2$ with models that are currently being used, and also see how well it predicts consumption over time.

Source	SS	df	MS			
Model	6.9980e+12	47	1.4889e+11	Number of obs	=	100,351
Residual	9.5644e+12	100,303	95355130.6	F(47, 100303)	=	1561.47
				Prob > F	=	0.0000
Total	1.6562e+13	100,350	165046510	R-squared	=	0.4225
				Adj R-squared	=	0.4223
				Root MSE	=	9765

Cons_exp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
cons_clothes_tot	2.867983	.03585	80.00	0.000	2.797718	2.938249
cons_entertainment_tot	6.917222	.1386698	49.88	0.000	6.645431	7.189013
cons_lemon	111.8916	2.054903	54.45	0.000	107.864	115.9192
cons_med_insti_tot	1.024448	.0144747	70.78	0.000	.9960779	1.052818
cons_misc_HH_consumables_tot	7.2481	.235505	30.78	0.000	6.786513	7.709687
cons_educ_exp_tot	1.011077	.0139518	72.47	0.000	.9837313	1.038422
cons_non_insti_med_tot	1.265053	.040189	31.48	0.000	1.186283	1.343823
cons_egg_meat	1.608287	.0759773	21.17	0.000	1.459373	1.757202
WH_car	2378.127	134.9584	17.62	0.000	2113.611	2642.644
cons_milk	1.086139	.0517	21.01	0.000	.9848077	1.18747
cons_servant	1.781702	.1110405	16.05	0.000	1.564064	1.999341
cons_refined_liquor	1.99706	.1772259	11.27	0.000	1.649699	2.34442
Religion_num						
christian	518.493	322.9916	1.61	0.108	-114.5666	1151.553
hindu	-64.70047	302.5857	-0.21	0.831	-657.7647	528.3638
jain	-149.0918	627.7411	-0.24	0.812	-1379.456	1081.273
muslim	11.32978	313.3957	0.04	0.971	-602.9218	625.5814
other	-329.625	442.2875	-0.75	0.456	-1196.503	537.2531
parsi	54122.48	5694.533	9.50	0.000	42961.27	65283.7
sikh	-509.2218	382.6152	-1.33	0.183	-1259.143	240.6991
cons_costly_dals	1.463294	.2416989	6.05	0.000	.9895669	1.937021
cons_internet_expense	5.006211	.4720118	10.61	0.000	4.081074	5.931348
cons_cable_TV	-4.614038	.4107009	-11.23	0.000	-5.419007	-3.809069
cons_electricity	1.428515	.1327405	10.76	0.000	1.168345	1.688685
Type_of_land_owned_num						
homestead & other	-779.6425	108.0405	-7.22	0.000	-991.4006	-567.8843
homestead only	-1074.53	102.8218	-10.45	0.000	-1276.06	-873.0006
other only	57.69755	311.5853	0.19	0.853	-553.0059	668.401
cons_jowar	1.443957	.3904656	3.70	0.000	.6786488	2.209264
cons_dryfruits_tot	2.158256	.2790305	7.73	0.000	1.61136	2.705152
cons_cinema	-5.792737	.5447567	-10.63	0.000	-6.860453	-4.725021
Reg_salary_earner	435.5334	72.12312	6.04	0.000	294.173	576.8938
cons_vanaspati_oil	-1.787773	.8264548	-2.16	0.031	-3.407615	-.1679323
cons_country_liquor	4.969433	.2711754	18.33	0.000	4.437932	5.500933
cons_quilt_matress	2.625459	.3010116	8.72	0.000	2.03548	3.215438
cons_firewood	1.135562	.1816908	6.25	0.000	.7794499	1.491673
Cooking_Code_num						
charcoal	654.7243	909.1803	0.72	0.471	-1127.258	2436.706
coke, coal	-397.9478	255.4947	-1.56	0.119	-898.7141	102.8186
dung cake	100.4993	161.3635	0.62	0.533	-215.7711	416.7697
electricity	332.0359	555.8425	0.60	0.550	-757.4084	1421.48
firewood and chips	-287.3735	96.26706	-2.99	0.003	-476.0558	-98.69127
gobar gas	-262.0427	788.6295	-0.33	0.740	-1807.747	1283.661
kerosene	-254.1627	211.5939	-1.20	0.230	-668.8841	160.5586
no cooking arrangement	1211.758	253.0071	4.79	0.000	715.8675	1707.649
others	12.57132	229.3669	0.05	0.956	-436.985	462.1277
cons_costly_cereal	.9825165	.07682	12.79	0.000	.8319503	1.133083
cons_leather_sandals	-.6404885	.3035795	-2.11	0.035	-1.235501	-.0454763
cons_uniform_boys	-2.477231	.2982687	-8.31	0.000	-3.061834	-1.892628
WH_PC_laptop	1207.811	139.9675	8.63	0.000	933.4766	1482.146
_cons	953.5948	320.2535	2.98	0.003	325.902	1581.288

Table 1: Regression results- With 2011 coefficients
Using $\overline{R}^2$ maximising model

6.3 Comparing the predictive capability

We have used two datasets in our comparison: household expenditure surveys for 2011 and 2022. We find the covariates of the $\bar{R}^2$ maximizing model from 2011 data and note down the coefficients. A similar approach is taken for the coefficients of the Grosh-Baker model for comparison purposes. To establish our model's predictive capability: (i) we compare the $\bar{R}^2$ for these two models (ii) we look at how well it predicts consumption income for 2022.

The Grosh-Baker model, published by the World Bank, is representative of the existing PMT methodology (ie. selecting proxies on the basis of high correlation of certain standard characteristics). For analytical purposes, we compare our results with this model to see how better does our $\bar{R}^2$ *maximizing model* fare than the existing norm.

6.3.1 Comparing Adjusted-R Square

The $\bar{R}^2$ of our model and the one of Grosh-Baker is 0.42 and 0.16 respectively(The latter is seen in Table 2). Our model is better at predicting consumption expenditure by using the optimal combination of covariates.

Source	SS	df	MS			
				Number of obs	=	99,522
				F(29, 99492)	=	664.37
Model	2.6247e+12	29	9.0507e+10	Prob > F	=	0.0000
Residual	1.3554e+13	99,492	136230118	R-squared	=	0.1622
				Adj R-squared	=	0.1620
Total	1.6179e+13	99,521	162563793	Root MSE	=	11672

Cons_exp	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
rural	464.34	91.45751	5.08	0.000	285.0844	643.5956
Dwelling_unit_Code_num	-650.3634	40.79347	-15.94	0.000	-730.3181	-570.4087
cons_electricity	5.018252	.1562174	32.12	0.000	4.712067	5.324436
cons_water_bill	4.294743	.7962306	5.39	0.000	2.73414	5.855345
HH_Size_num	792.8436	18.1106	43.78	0.000	757.347	828.3402
Education_num						
02	-220.1121	730.9088	-0.30	0.763	-1652.684	1212.46
03	-998.1222	1521.679	-0.66	0.512	-3980.594	1984.349
04	553.4163	748.2905	0.74	0.460	-913.2239	2020.056
05	44.47575	140.2491	0.32	0.751	-230.4107	319.3622
06	92.46202	133.6529	0.69	0.489	-169.4959	354.42
07	529.1588	127.1047	4.16	0.000	280.0351	778.2825
08	410.4356	135.4971	3.03	0.002	144.8629	676.0083
10	778.3259	157.8562	4.93	0.000	468.9297	1087.722
11	2102.577	337.7548	6.23	0.000	1440.582	2764.573
12	876.2526	165.7386	5.29	0.000	551.407	1201.098
13	1719.375	237.3167	7.25	0.000	1254.237	2184.513
Age	8.632137	3.057036	2.82	0.005	2.640383	14.62389
Sex	6.094483	121.8287	0.05	0.960	-232.6882	244.8772
Reg_salary_earner	854.0273	89.19677	9.57	0.000	679.2027	1028.852
Land_owned	.2634127	.0224612	11.73	0.000	.219389	.3074364
WH_telephone	2944.291	153.1941	19.22	0.000	2644.033	3244.55
WH_stove	779.8697	95.07499	8.20	0.000	593.5238	966.2155
WH_fridge	2153.303	108.624	19.82	0.000	1940.402	2366.205
WH_AC_cooler	-115.877	123.5455	-0.94	0.348	-358.0248	126.2708
WH_electric_fan	93.58415	99.56841	0.94	0.347	-101.5687	288.737
WH_Radio	483.5018	89.08505	5.43	0.000	308.8961	658.1074
WH_motorcycle	1452.256	95.65581	15.18	0.000	1264.772	1639.74
WH_car	4355.089	160.3963	27.15	0.000	4040.714	4669.464
cons_servant	4.053995	.1314014	30.85	0.000	3.79645	4.31154
_cons	1707.836	234.2084	7.29	0.000	1248.79	2166.881

Table 2: Regression results- With 2011 coefficients
Using Grosh-Baker model

6.3.2 How well does it actually predict consumption expenditure

In order to determine this we test it on the 2022 dataset for consumption expenditure. We start by creating predicted values for consumption expenditure (2022) using covariates from our $\bar{R}^2$ *maximizing model* for 2011, as well as the Grosh-Baker model. After this, we plot the predicted expenditure overlapped with actual expenditure. Figure 3 overlaps the $\bar{R}^2$ *maximizing model* and Figure 4 overlaps the Grosh-Baker model with actual expenditure from NSS HCES 2022. The graph has household consumption expenditure (in Rs) on the x axis and the distribution of households on the y axis.

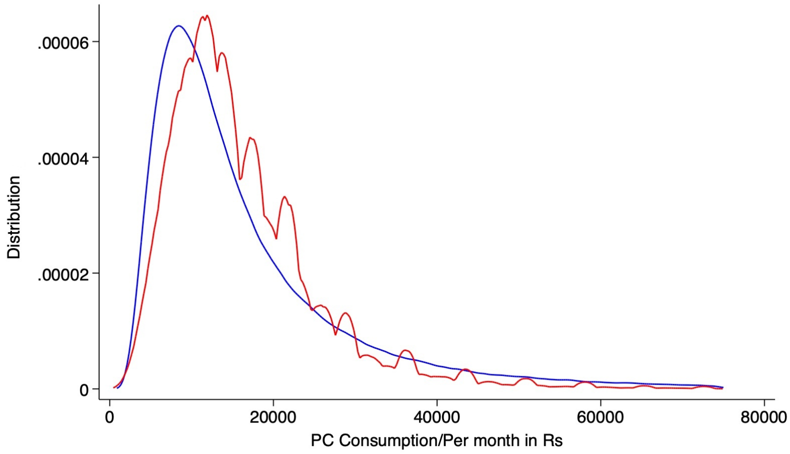


Figure 3: $\overline{R}$ Maximizing Model Overlapped with Actual Consumption

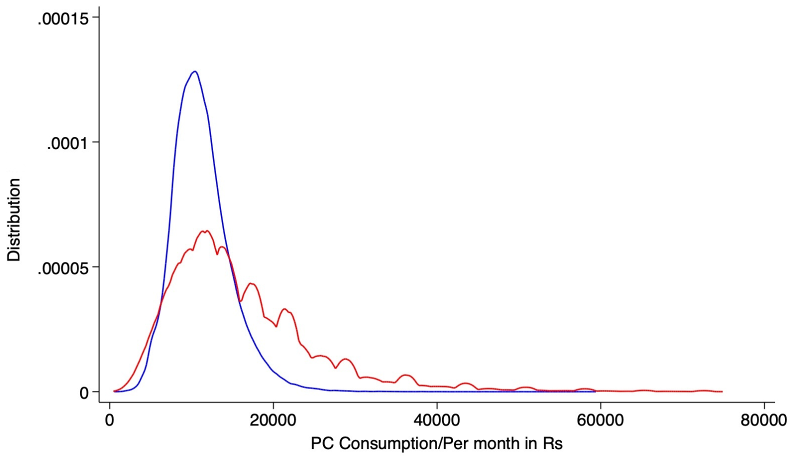


Figure 4: Grosh-Baker Model Overlapped with Actual Consumption

Legend: Red line represents real per household consumption expenditure in 2022, and Blue line represents predictions.

It is evident that the Grosh-Baker model inaccurately predicts household consumption expenditure, and the $\bar{R}^2$ *maximizing model* is in sync with true consumption expenditure. This is not only a testament to the relatively superior predictive ability, but also a proof that the covariates chosen while maximizing $\bar{R}^2$ are better at yielding time-invariant characteristics as they have a high degree of predictability in 2022 as well.

We now look at applications of the $\bar{R}^2$ *maximizing model* to study how better predictors can help us in delivering targetted social programs more effectively. We understand this by comparing the errors across various poverty lines, using them synonymously with eligibility thresholds.

6.4 Examining changes in errors with shifts in poverty line

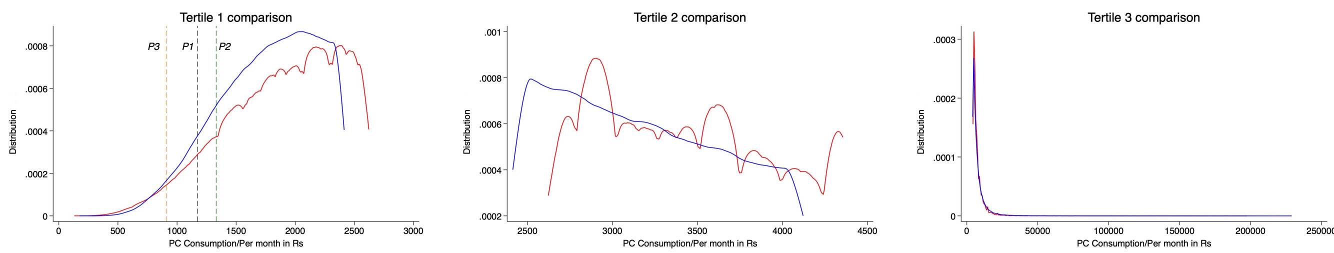
The choice of an ideal poverty threshold for targeting programs is widely debated, especially regarding its impact on welfare. Literature suggests that raising the poverty line reduces undercoverage but increases leakages, and vice versa for lowering the poverty line (Diether et al.; Balaji).

Poverty Lines	Exclusion Error	Accuracy	Inclusion Error
P1 @ Rs. 1172.5 per person/mo (*2011 poverty line*)	2.2%	95.23%	2.54%
P2 @ Rs. 1331.925 per person/mo (*World Bank poverty line*)	3.35%	92.46%	4.18%
P3 @ Rs. 908 per person/mo (*Tendulkar Committee poverty line*)	0.89%	98.35%	0.75%

Table 3: Impact of Change in Poverty Lines

To test this hypothesis, we adjusted the national poverty line by ±INR 200 (@2011 rates).[1] Our results, in Table 3, show that both undercoverage and leakage increase when the poverty line is raised and decrease when it is lowered. From our results we cannot conclude whether we should shift the poverty line up or down to change undercoverage and leakage. Rather, our results are possibly a reflection of the fact that our model better predicts consumption expenditure of the poorest (bottom 30%). To verify this, we analyzed the real versus predicted distributions of consumption expenditure, dividing households into three groups based on the income distribution, or tertiles (see Figure 5 and 6). Our findings confirm that our model performs best for households in the lowest income group, outperforming the Grosh-Baker model. This reflects our model's prioritization of proxies that strongly correlate with extreme poverty, ensuring effective targeting of the most vulnerable populations.

Figure 5: Side-by-Side Comparison of Tertiles for Per capita Consumption Expenditure vs Predicted Consumption Expenditure by $\overline{R}^2$ maximising model

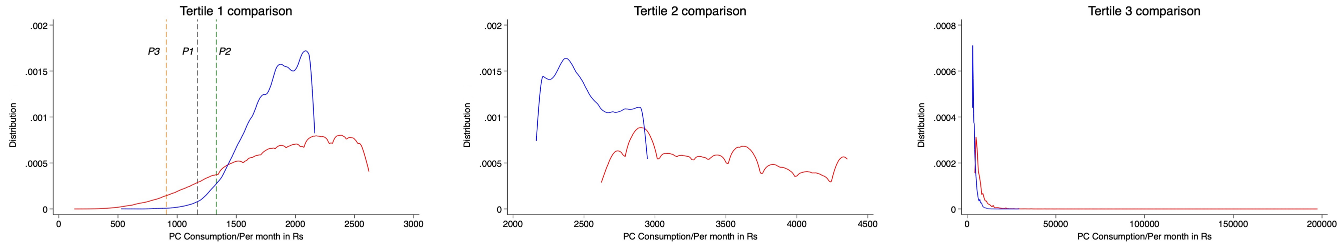


(a) Tertile 1: Total Exp vs Prediction (b) Tertile 2: Total Exp vs Prediction (c) Tertile 3: Total Exp vs Prediction

Legend: Red line represents real per capita consumption expenditure in 2022, and Blue line represents predictions.

[1]We take the national average (across rural and urban) poverty line to be Rs. 1172.5 per person per day. The higher poverty line (denoted by P2) is calculated according to the World Bank poverty line of $2.15 per person per day (PPP 2017), which amount to Rs. 1331.925 per person per month. The lower poverty line (denoted by P3) is based on the 2011 Tendulkar committee recommendations, which amounts to Rs. 908 per person per month.

Figure 6: Side-by-Side Comparison of Tertiles for Per Capita Consumption Expenditure vs Predicted Consumption Expenditure using Grosh-Baker Model



(a) Tertile 1: Total Exp vs Prediction (b) Tertile 2: Total Exp vs Prediction (c) Tertile 3: Total Exp vs Prediction

Legend: Red line represents real per capita consumption expenditure in 2022, and Blue line represents predictions.

7 Scope and Conclusion

This paper underscores the importance of refining Proxy Means Tests to enhance targeting in welfare schemes. By optimizing $\bar{R}^2$ and introducing dynamic, observable proxies for consumables, the proposed model minimizes inclusion and exclusion errors. It highlights the need for adaptive targeting mechanisms that account for changing socio-economic conditions, ensuring effective and equitable resource allocation over time. But, it is important to note that the selected covariates referred to in the paper are not to be misunderstood as the most optimal ones. We simply recommend the framework used to arrive at the said covariates. These obviously change with time, and can be updated accordingly. Future research should explore the integration of machine learning to dynamically adapt proxy sets in real time, potentially reducing model errors further.

We propose a novel framework to optimally select PMT covariates that best reflect consumption expenditure using national level data. However, the proxies that delineate the bottom 40% in say, Rajasthan (a North Indian High Focus State) may be very different from that in Kerala (a South Indian Low Focus State) given state specific household characteristics, institutions, political climate, and different state poverty lines. Further research can be undertaken using the $\bar{R}^2$ maximizing model in exploring state-specific adaptations to account for regional socio-economic disparities and cultural heterogeneity. Investigations into how PMT frameworks can incorporate time-series data to dynamically adjust for evolving household characteristics and poverty thresholds across states would be valuable. Moreover, examining the role of localized administrative systems and governance structures in implementing and refining PMT-based targeting mechanisms could yield insights for practical policy applications and help further reduce the errors.

8 References

Beuermann, Diether W., et al. "Shooting a Moving Target: Evaluating Targeting Tools for Social Programs When Income Fluctuates." *Journal of Development Economics*, vol. 172, 2024, p. 103395, https://doi.org/10.1016/j.jdeveco.2024.103395. Accessed 1 Dec. 2024.

Grosh, Margaret, and Judy Baker. "Proxy Means Tests for Targeting Social Programs : Simulations and Speculation." *World Bank*, 2019, documents.worldbank.org/en/publication/documents-reports/documentdetail/750401468776352539/proxy-means-tests-for-targeting-social-programs-simulations-and-speculation. Accessed 1 Dec. 2024.

Kidd , Stephen , and Emily Wylde. "Targeting the Poorest: An Assessment of the Proxy Means Test Methodology." *Australian Government Department of Foreign Affairs and Trade*, Sept. 2011, www.dfat.gov.au/about-us/publications/Pages/targeting-the-poorest-an-assessment-of-the-proxy-means-test-methodology.

Klasen, Stephan, and Simon Lange. "Targeting Performance and Poverty Effects of Proxy Means-Tested Transfers: Trade-Offs and Challenges." *Ibero America Institute for Econ. Research (IAI) Discussion Papers*, 18 Mar. 2015, ideas.repec.org/p/got/iaidps/231.html. Accessed 1 Dec. 2024.

Mahamallik , Motilal, and Gagan Bihari Sahu. "Identification of the Poor: Errors of Exclusion and Inclusion." Economic and Political Weekly, vol. 46, no. 9, 2011, pp. 71–77. *JSTOR*, www.jstor.org/stable/41151842, https://doi.org/10.2307/41151842.

Murugesan, Balaji. "Changing poverty line-Inclusion and Exclusion of Poor in Indian States." *International Journal of Business Excellence*, vol. 1, no. 1, 2019, p. 1, https://doi.org/10.1504/ijbex.2019.10023037. Accessed 2 Feb. 2020.

National Sample Survey Office. Household Consumption Expenditure Survey: 2022-23. Ministry of Statistics and Programme Implementation, Government of India, 7 June 2024, https://microdata.gov.in/nada43/index.php/catalog/194.

National Sample Survey Office. Household Consumer Expenditure: Type 1, July 2011 - June 2012, NSS 68th Round. Ministry of Statistics and Programme Implementation, Government of India, 20 Nov. 2017, https://microdata.gov.in/nada43/index.php/catalog/1.

Moharir, Advait. *hces*_2022. GitHub, 2023, https://github.com/advaitmoharir/hces2022.