**Title**
Assessing Targeting Peformance: The Case of Ghana’s LEAP Program



**Authors**
Ligon, Ethan
Trachtman, Carly

**Publication Date**
2024-04-10

# Assessing Targeting Peformance: The Case of Ghana's LEAP Program

Carly Trachtman[1] and Ethan Ligon[2]

[1]International Food Policy Research Institute; Markets, Trade, and Institutions Unit
[2]UC Berkeley; Department of Agricultural and Resource Economics

April 11, 2024

## 1 Introduction

### 1.1 Motivation

Identifying the poorest and neediest households to receive limited social program resources remains a challenge in many low-income countries. These countries often lack the types of detailed tax records that high-income countries use to assess social program benefit eligibility, especially given the prevalence of employment in the informal and agricultural sectors. In many cases, this means that low-income countries must collect and maintain social registry data. This can be both expensive and logistically challenging, as households are often scattered throughout remote areas. As a result, many countries do not have up-to-date lists of those individuals or households who may benefit most from such programs. For instance, in the case of Ghana, which is the focus of this chapter, the Ghana National Household Registry Initiative (which collects basic information on the socioeconomic status of households in Ghana) did not even exist until 2016, with (still ongoing) data collection beginning in 2020.

In the absence of up-to-date social registry data, governments and other policy implementers have sought out cheaper ways to estimate individuals' current welfare status. A commonly used method for such estimation is proxy means testing (PMT), which employs a small set of cheaply observable proxies (usually household demographic characteristics and ownership of visible assets) to predict a household's current welfare level [Grosh and Baker, 1995]. The prediction algorithm is generally developed using survey data from comparable populations, which contains information on both the proxies and the welfare metric the proxies will be used to predict. Once this algorithm is developed, only information on the proxies is needed to predict which households are poorest based on the chosen welfare metric, and hence eligible for social benefits. Given its simplicity and potential for cost savings, PMT is used for targeting benefits by many social protection programs across Africa [Schnitzer and Stoeffler, 2024].

Historically, the welfare metric that PMT scores use proxies to predict is a household's per capita consumption level [Brown et al., 2018]. This is in line with what have widely been considered "best practices" for measuring welfare in poor country contexts in the past few decades, given the challenges associated with accurately measuring income via household survey in low-data environments [Deaton, 1997]. However, using survey measures of per capita consumption as a welfare metric also has drawbacks. Per capita consumption aggregates are generally calculated by summing over households' reports of consumption expenditures on a variety of expenditure categories, and then dividing by the number of "adult equivalent" household members. Households might report expenditures in each category with some error, and hence adding the categories together simply sums their errors, potentially producing quite noisy estimates of total consumption. Indeed, reports of expenditures on different categories have been shown to be quite error prone, and can be highly sensitive even to small changes in survey module design [Gibson et al., 2015, Friedman et al., 2017]. Critically, such errors may be non-random, which can decrease the absolute targeting accuracy of PMT [Gazeaud, 2020].

Besides these practical measurement concerns, there are also theoretical issues with using per capita consumption expenditures as a welfare metric. Notably, policymakers often want to target the poorest or "worst off" households, who would gain the most from receiving program benefits. In terms of economic theory, these would be the households with the lowest current utility levels, who (assuming utility functions are concave) would get the highest marginal utility from making additional consumption expenditures. Notably, changes in utility do not directly correspond to changes in total consumption expenditures, except under a specific set of assumptions, that generally do not seem to hold empirically. One of these assumptions is that when households experience an increase (or decrease) in income, they will increase (or decrease) their consumption of each category of goods in fixed proportions. In other words, this assumption suggests that if a household is spending 20% of their budget on maize and then receives additional money to spend, they will spend exactly 20% of this money on maize as well. There is myriad evidence that households do not behave this way, for example spending a smaller share of their overall budget on food as their budget increases [Banks et al., 1997]. Moreover, because households buy different types of goods at different total expenditure levels, looking at *which* goods households choose to consume can provide additional information about households' welfare, that we discard if simply sum expenditures on all categories together. A related issue has to do with prices. If the relative price of different kinds of staple foods increases, for example, then very basic economic theory tells us that poorer households will be made differentially worse-off. This is not captured correctly by per capita consumption expenditure aggregates, even if one adjusts using a consumer price index, and will tend to understate the welfare cost of such price changes to poorer households.

## 1.2 Proposal of Alternative Metric

Given the drawbacks of using per capita consumption as a welfare metric, a natural question is whether there is an alternative, preferable welfare metric that we could estimate with PMT scores in order to target program beneficiaries. In this paper, we propose such an alternative welfare metric: the estimated index of marginal utility of expenditures from Ligon [2020], which we henceforth refer to as $w$ for brevity. $w$ can be interpreted as an indicator of the relative (negated) marginal utility that a household receives from an additional dollar (or unit) of expenditures. The negation is simply for ease of interpretation; that way, low values of $w$ identify the households that are worse off and high values of $w$ identify those who are better off. The households with low values of $w$ are the ones with the highest marginal utility of additional expenditures and should benefit most from receiving resources, which are precisely the households that social benefits programs often target.

Critically, $w$ is calculated by estimating a Constant Frisch Elasticity (CFE) demand system. By estimating a non-homothetic demand *system* for multiple goods, we are able to treat a household's demands for different goods as distinct (but correlated) pieces of information, that depend on the characteristics of the household (e.g. household size and composition), prices of goods, and good-specific total-expenditure elasticities. Notice that this is different from the standard PMT approach, which simply tries to predict *total* expenditures, and ignores differences in the composition of different households' consumption baskets. Thus, this framework allows households which spend more overall to potentially have a different consumption basket composition than households which spend less overall. Practically, the $w$ measure may also help alleviate some of the measurement error concerns that arise when computing consumption aggregates. By treating households' demands for goods as separate pieces of information (all of which are potentially measured with noise), we are able to extract the household-specific signal. Hence noise in a sense gets "netted out" instead of added together, as is the case when expenditures are simply summed up over categories. Importantly, the $w$ metric has similar or lesser data requirements to calculate than a consumption aggregate, using basic information on household composition (of the same kind needed to calculate adult equivalency scales) and expenditures on various categories of goods. $w$ can even be calculated with only a subset of goods, allowing for the omission of goods that households might report on with particularly high levels of noise, such as those they purchase infrequently.

Let $j = 1, \ldots, \bar{J}$ index some set of goods, and let prices for all goods be a vector $p$. Suppose that a particular household has a set of observable characteristics $z$ (e.g., household size and composition, rural vs. urban, etc.). We suppose that we observe $J \leq \bar{J}$ expenditures on some consumption items for this same household, denoting expenditures on good $j$ by $x^j$. Then the CFE demand system can be expressed as a set of equations for each of those $J$ different goods,

$$\log x^j = \alpha_j(p) + \gamma_j(z) + \beta_j w + \epsilon_j, \tag{1}$$

where (i) $\epsilon$ is an error term (perhaps related to unobservable household characteristics, or measurement error in expenditures); (ii) the function $\alpha_j(p)$ tells us how expenditures depend on the vector of prices; (iii) the function $\gamma_j(z)$ controls for household characteristics; and critically (iv) the coefficient $\beta_j$ tells us how changes in the household's welfare $w$ affects expenditures. That same $\beta_j$ can be interpreted as a *relative income elasticity*, so takes larger values for "luxuries" and smaller values for "staples".

So, if we have data on a particular household's expenditures and already *know* (or have estimates of) $(\alpha, \beta, \gamma)$, then we can simply find the value of $w$ which minimizes the sum of squared errors

$$\sum_j \epsilon_j^2 = \sum_j (\log x^j - \alpha_j(p) - \gamma_j(z) - \beta_j w).$$

This is just a least-squares problem, and can solved even for many households simply by regressing $y_j = \log x^j - \alpha_j(p) - \gamma_j(z)$ on the different values of $\beta_j$, so that we obtain estimated values of $w$ as household-specific coefficients on the (known or estimated) $\beta_j$.

Of course, this omits the important fact that we may not know the demand system in advance—if we assume that demands are in the CFE family, this means that we will need to estimate $(\alpha, \beta, \gamma)$. So ordinarily estimation of the welfare measures $w$ proceeds in two steps. There's a sort of "training" step, in which data on household item-level expenditures is used to estimate $(\alpha, \beta, \gamma)$. Depending on the assumptions one makes, one may be able to also formulate this training stage as a linear estimation problem. Then there's a "prediction" step, which just involves the Ordinary Least Squares (OLS) regression given above.

Identification assumptions for estimation of the demand system may vary with the application. Here we identify $\alpha(p)$ by assuming that all households in the same region (market) of the country face the same prices (though of course these can vary over time). Then the $\alpha(p)$ can be treated as latent variables, and estimated as good-time-market fixed effects. We identify $\gamma(z)$ by assuming that the $\gamma$ is linear, allowing estimation of the parameters of this function via least squares. Finally we assume that $z$ is orthogonal to $w$, allowing us to estimate $(\alpha, \gamma)$ in a first stage without knowing $\beta$. The second stage of estimation then exploits the fact that the covariance matrix of log expenditures conditional on $(p, z)$ is equal to the to the outer matrix product $\beta\beta'$, up to a normalization. This then gives us estimates of all of the objects that appear in the demand system $(\alpha, \beta, \gamma)$, and obtaining estimates of $w$ via the OLS prediction step described above. The code and pointers to the data used to estimate $\beta$ and the values of $w$ for households in the 2016–17 wave of the Ghana Living Standards Survey are described in the Appendix.

## 1.3 Research Question and Context

Even if the $w$ metric has benefits over per capita consumption aggregates as a welfare indicator, it remains to be seen whether it can be estimated with PMT scores in a similar manner to per capita consumption aggregates. PMT scoring tools such as the PPI have identified and validated small sets of relatively easily observable proxies that predict consumption aggregates [IPA, 2019], but no similar exercise has been done for $w$. Hence in this paper, we ask whether it is possible to generate PMT scoring functions that can predict $w$ with an accuracy similar or superior to the accuracy with which PMT functions predict consumption aggregates. Additionally, we ask whether the types of proxies used to predict per capita consumption expenditures can also predict $w$, or if other types of proxies may be more suitable.

We explore these questions in the context of targeting Ghana's flagship social program, the Livelihoods Empowerment Against Poverty (LEAP) program. LEAP provides bimonthly cash transfers and health insurance to nearly a quarter of a million extremely poor households across nearly all districts of Ghana. The program targets poor, vulnerable households, with elderly individuals, disabled individuals, orphans/vulnerable children, and mothers who are pregnant or have infants. LEAP's current beneficiary targeting regime has multiple steps. First, centralized geographical targeting identifies regions, districts, and communities in which poor households are likely located. In each area, local officials help to identify households who are likely eligible for LEAP. These nominated individuals then have their poverty status verified using a PMT. Little about the PMT formula LEAP uses to target households has been publicly disclosed, however it is possible to learn from surveys which households receive LEAP benefits. Hence, we also ask how targeting based on the PMT scores we estimate in this chapter (to predict both per capita consumption expenditures and $w$) compare to LEAP's actual targeting.

## 1.4 Summary and Contribution

In our analysis, we separately estimate PMT scores that are meant to predict per capita consumption expenditures and $w$, using data from the 2016–2017 round of the Ghana Living Standard Survey (GLSS). We use penalized ordinary least squares regression techniques, specifically least absolute shrinkage and selection operator (LASSO) regression, to estimate PMT formulas with four distinct sets of proxies that vary in the numbers and types of proxy variables used. These four sets include a sparse set of 10 proxies that are used to construct the PPI; two more extended sets of around 20 proxies each, which contain standard asset ownership and household demographics variables; and a final sparse proxy set of expenditures with only six categories of consumption goods. Additionally, as an alternative way to construct a PMT formula, we use CFE demand estimation techniques to predict the $w$ we estimated with the full set of consumption categories using only the reduced set of six consumption category "proxies." We then explore model fit and targeting accuracy both in-sample (in the 2016–2017 GLSS data) and out-of-sample, using supplementary data from an evaluation of the LEAP 1000 pilot. The LEAP 1000 pilot tested the expansion of LEAP benefits to mothers who are pregnant or have infants. (This eligibility category was not introduced until 2016.)

Using LASSO regression, we estimate models that explain between 14–50% of the variation in per capita consumption expenditures depending on the set of proxies used, and between 29–41% of variation in $w$. We find that proxies related to household demographics and asset ownership are better predictors of per capita consumption expenditures, whereas consumption expenditures on a reduced subset of goods are better predictors of $w$. Using the CFE demand estimation-based PMT score that uses expenditures on only 6 goods, we find that the model explains about 7% of the variation in per capita consumption expenditures, and 34% of the variation in $w$ (when measured with a more robust set of consumption goods). When looking at targeting errors, linear models that use mostly asset ownership and household demographic proxies outperform current LEAP targeting in correctly including the poorest households based on the per capita consumption benchmark in-sample, but perform worse out-of-sample. In contrast, our linear PMT models which aim at predicting $w$ do not outperform LEAP in terms of minimizing $w$-based targeting errors in-sample. However, out of sample, we instead see significant improvements in $w$-based targeting accuracy when the limited set of consumption expenditure categories are used to predict $w$, and a similar quality of per capita consumption-based targeting between our models and LEAP's current regime. Additionally, when we use CFE demand estimation with a limited set of goods to calculate PMT scores, we see improvements in targeting accuracy based on both per capita consumption expenditures and $w$, both in-sample and out-of-sample. Hence policymakers may want to consider using a limited set of consumption categories as proxies (rather than assets and demographic information) and CFE estimation-based PMT score formulas when calculating PMT scores, especially when we want to use $w$ rather than per capita consumpiton expenditures as a welfare metric.

This work contributes to the literature on using alternative welfare metrics to assess targeting accuracy. Previous literature has considered welfare metrics that move beyond per capita consumption, such as an asset index [Karlan and Thuysbaert, 2019, Aiken et al., 2023], self-assessed welfare ladder steps [Alatas et al., 2012], and a food security score [Premand and Schnitzer, 2021]. In most of these cases, alternative targeting benchmarks are introduced to better understand the type of welfare benchmarks that community members might target in a different method of targeting called community-based targeting.

In this paper, we instead propose the use of an alternative targeting benchmark metric because of theoretical and practical reasons that make the alternative metric better at capturing household welfare more generally. Notably Trachtman et al. [2022] also uses this $w$ metric to assess targeting accuracy of community-based targeting and PMT methods. However, there is little (if any) existing work that seeks to generate PMT score formulas that seek to predict alternative welfare metrics for the purposes of program targeting. Hence our paper helps to answer key questions about whether the types of proxies and techniques that work best for predicting consumption scores will also work well to predict alternative welfare metrics.

# 2 Data/Methods

## 2.1 Data

We use two data sets to conduct our empirical analysis. The first is the 2016-2017 round of the GLSS, which is the most recent survey collection round for which the data has been made publicly available. The

GLSS contains data on around 14,000 households from all ten regions of Ghana that existed in 2017 (before the 2018 referendum further splitting Ghana into 16 regions). The data set is rich, containing information about household composition and economic activities, asset ownership/housing quality, and consumption expenditures on around 200 different food items. Households also were asked about any transfers that they receive, with the option to report receiving LEAP benefits. Only around 1.5% of households in this data set report receiving LEAP benefits. Official estimates suggest that LEAP covers over 1.4 million individuals, which is over 4% of Ghana's population. Hence it is possible that individuals under-report receipt of LEAP benefits on GLSS, which is one potential drawback of assessing target accuracy in this data set.

To circumvent this issue and to test the out-of-sample performance of our estimated PMT scores, we use a second data set, which is baseline data from the UNC Transfer Project's evaluation of the LEAP 1000 Pilot, collected in 2015. This is a much smaller data set containing responses from around 2,500 households in only the Northern and Upper East regions of Ghana. The survey modules used were abbreviated versions of those in the GLSS survey, and hence data on the ownership of some assets and on consumption expenditures in some categories is not available. However, because this data was used for an impact evaluation, it contains only households that had been nominated to potentially receive LEAP benefits locally, and that were either just above or just below the PMT threshold score that determines LEAP benefit receipt. Hence around 50% of these relatively similar households receive LEAP, and LEAP receipt status data comes from administrative records rather than self-reports. Hence despite the more limited geographical coverage and fewer welfare variables collected, the LEAP 1000 data remains a favorable setting in which to assess the targeting accuracy of our estimated PMT formulas.

## 2.2 Methods

### 2.2.1 Estimating Welfare Metrics

In order to estimate PMT scores and assess targeting performance, we need estimates of per capita expenditures and $w$ in both the GLSS and the LEAP 1000 data. In both data sets, we use the estimates of adult-equivalent food expenditures that were already calculated by the authors of each data set. These estimates were generated by summing over household food expenditures in all food categories and dividing by the number of adult-equivalent household members. Adult-equivalency is based on the relative caloric needs of different household members. For instance, adult men ages 25-50 are counted as 1, whereas females ages 11-50 are counted as 0.76, children ages 7-10 are counted as 0.69, etc. We calculated $w$ in the GLSS data using 77 consumption categories (where some categories are combinations of similar goods that were asked about separately), controlling for any good-specific demand driven by simple household attributes (log household size, number of adult men, number of adult women, number of girls and number of boys), and by region-specific market conditions. The estimated total expenditure elasticities for various goods can be seen in Figure 1. As one might expect, "luxury" goods like chewing gum and cocoa have relatively high estimated elasticities, while staples like maize and maize flour have relatively low estimated elasticities. Given the more limited set of goods in the LEAP 1000 data, we estimated $w$ using only a small set of six goods, which we will discuss in more detail later in this section.

### 2.2.2 Estimating PMT Scores: LASSO Regression

Our first set of PMT formulas simply use ordinary least squares to estimate the relationship between a set of proxies and the welfare metric we are looking to estimate (either per capita consumption or $w$). We

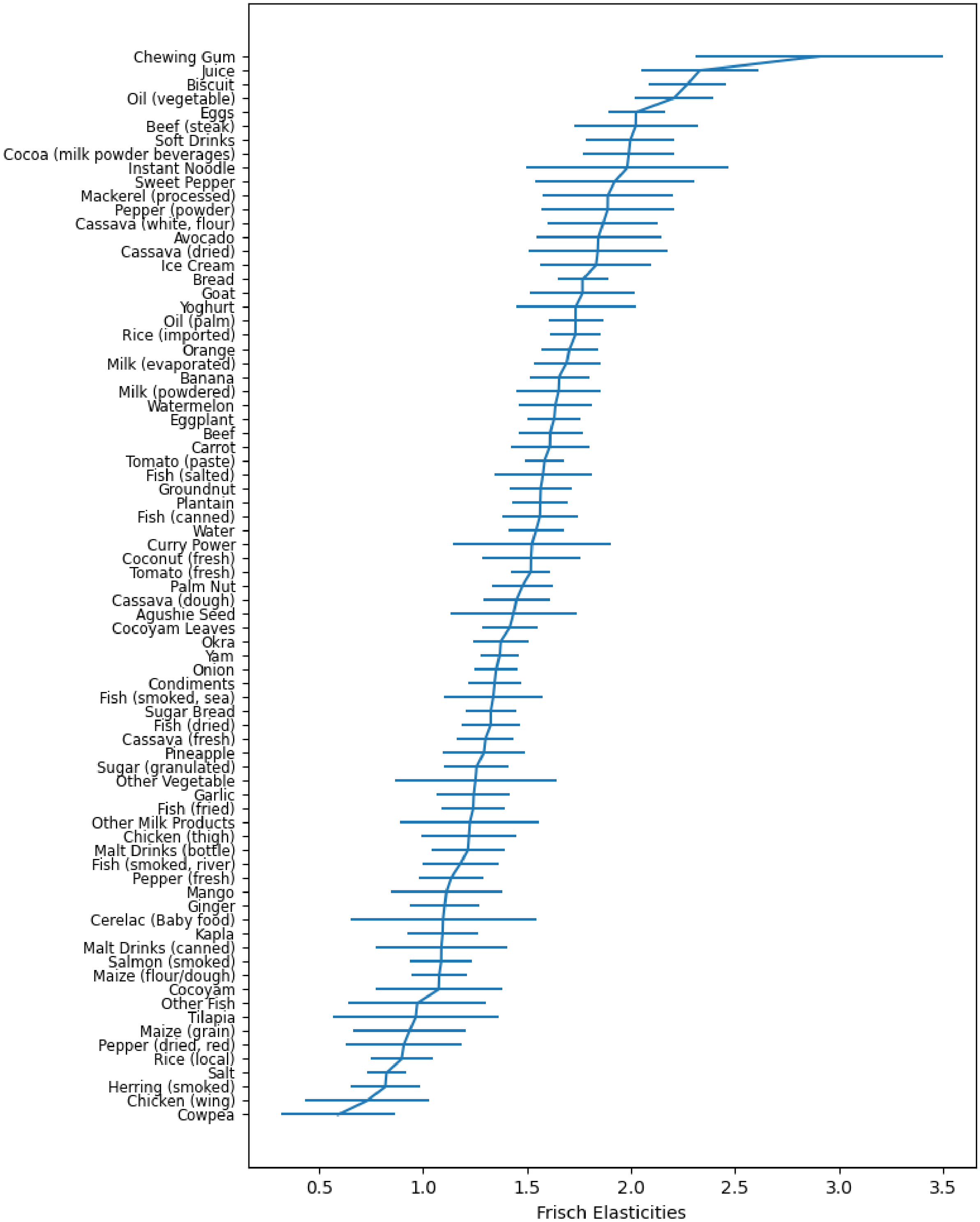


Figure 1: Estimated elasticities of consumption categories

Table 1: Proxies Used to Estimate PMT Scores

| Proxy Set 1:<br>Ghana PPI Variables | Proxy Set 2:<br>Variables likely used by LEAP (Pop 2015) | Proxy Set 3:<br>Variables from previous categories that are in LEAP 1000 Evaluation Data | Proxy Set 4:<br>Consumption expenditures (subset of goods) |
|---|---|---|---|
| Region<br>HH size categories<br>Purchased chicken eggs<br>Purchased beef<br>Outer wall material<br>Cooking fuel source<br>Own gas stove<br>Own refrigerator<br>Own TV<br>Own fan fan | Region<br>HH size<br>Elderly HH member<br>Share HH adults<br>HH head wage employ.<br>HH head works on own farm<br>HH member categories<br>Rooms in house categories<br>Has electricity<br>Water source categories<br>Toilet type categories<br>Outer wall material<br>Roof material<br>Floor material<br>Own land<br>Own livestock categories<br>Own stove<br>Own refrigerator/freezer<br>Own TV<br>Own motorcycle<br>Own car<br>Rural | Region<br>HH size count/ categories<br>Purchased chicken eggs<br>Purchased beef<br>Elderly HH member<br>Share HH adults<br>HH head wage employ.<br>HH head works on own farm<br>HH members categories<br>Rooms in house categories<br>Has electricity<br>Water source categories<br>Toilet type categories<br>Outer wall material<br>Roof material<br>Floor material<br>Cooking fuel source<br>Own livestock categories | Region<br>Expenditures on corn dough<br>Expenditures on bread<br>Expenditures on eggplant<br>Expenditures on okra<br>Expenditures on onion<br>Expenditures on salt |

start with four distinct sets of proxies to use in prediction, which are summarized in Table 1. The first set of proxies is a set of 10 variables identified (also using the 2016-2017 GLSS data) by the Poverty Probability Index to be highly predictive of per capita consumption expenditures. While the LEAP program does not share which variables it uses to compute PMT scores, Pop [2015] identifies a set of variables that is likely used, which we use as our second set of proxies. However, the LEAP 1000 data does not contain information of the full set of proxies in either of the previous two proxy sets. So to facilitate out-of-sample prediction, for the third set of proxies, we choose any of the proxies in the first two sets that are also present in the LEAP 1000 evaluation data. Finally, the fourth set of proxies is distinct, as instead of using "standard" proxies like asset ownership and household demographics, it includes expenditures in six food categories. Given that $w$'s calculation depends on a demand system and is meant to capture current utility levels rather than long-term wealth, we hypothesized that a limited set of consumption goods might be better at predicting $w$ than the sets of other proxies. Additionally, use of these proxies will allow us to compare different PMT estimation procedures using roughly the same sets of variables, as we also use these same proxy goods to estimate CFE demands-based PMT scores. We choose these six expenditure categories in particular because they are commonly purchased and have clearly analogous categories in the consumption modules in both the GLSS and LEAP 1000 data, allowing for comparable in-sample and out-of-sample prediction.

With each of these sets of proxies, we choose the optimal model using a LASSO regression procedure. Specifically, for each set of proxies, we take $X$ to be set of all of the proxies, plus all possible pairwise interactions between these proxies, and take $Y$ to be the welfare metric of interest. We then solve:

$$\min_{\beta}(Y - X\beta)'(Y - X\beta) + \lambda \sum_{j=1}^{p} |\beta_j|$$

where $\lambda$ is a penalty for additional the the inclusion of predictors with non-zero coefficients. The inclusion of this penalty term helps mitigate the risk of over-fitting, which could lead to poor out-of-sample prediction. In our case, we choose a value of $\lambda$ that minimizes mean squared error, using the `cv.glmnet` command from the `glmnet` package in R.

#### 2.2.3 Estimating PMT Scores: CFE Demand Estimation With a Subset of Goods

We use a second technique to estimate PMT scores that predict $w$ as well. In the GLSS data, we estimate $\beta$ using a large set of goods, and use that to predict $w$ for households within the GLSS dataset. However, a useful feature of the CFE demand estimation procedure is that we can still estimate $w$ for households with missing information on consumption in some categories. Hence another way to estimate a PMT score is to only use a subset of goods as proxies, and treat expenditure on all other categories of goods as "missing." In this sense, we are using the fitted elasticities from the model with many goods to estimate $w$ in the case where we have data on a much smaller set of goods, which is the general idea of a PMT formula. For PMT score estimation, we use the same set of six consumption expenditure categories listed under Proxy Set 4 in Table 1.

#### 2.2.4 Calculating Targeting Errors

Once all of the PMT formulas are estimated, we then calculate the predicted (fitted) values from each set of proxies predicting each welfare indicator (per capita expenditures or $w$). Next, we calculate targeting errors as follows. First, we rank households in each region from poorest to richest based on our estimates of per capita expenditures, our estimates of $w$, and the predicted values of each metric under each of our estimated models. Then we assume that the number of transfers available is fixed at the regional level, and is equal to the number of households that are currently receiving LEAP in that region, $n_r$. We do this given that LEAP's actual program targeting features a geographical targeting component that first assigns eligible districts before identifying recipients. Then we consider every household with a rank of $n_r$ or less as potential transfer recipients under that metric/model. In the case of ties at ranks near $n_r$ which make it such that the number of targeting recipients using the described procedure does not actually yield the number of transfers assigned to that region, we randomly assign the remaining number of transfers to households near the threshold, such that the overall percentage of households in the data set receiving LEAP is fixed. Then for each welfare metric $j \in \{\text{PCE}, w\}$, we calculate an indicator variable for a household being "correctly included" under a given model $k$ which equals 1 if metric $j$ and model $k$ both suggest that the household is in the $n_r$ poorest households of their region, an indicator for "correctly excluded" that equals 1 if both metric $j$ and model $k$ suggest that the household is not in the $n_r$ poorest households of their region, an indicator for "incorrectly included" that equals 1 if metric $j$ suggests a household is not in the $n_r$ poorest, but model $k$ suggests it is, and an indicator for "incorrectly excluded" that equals 1 if metric $j$ suggests a household is in the $n_r$ poorest but model $k$ suggests it is not. To compare targeting accuracy among models, we compare whether the percentage of households correctly included, correctly excluded, incorrectly included and incorrectly excluded are statistically different using a t-test of means.

# 3 Results: PMT Formulas/in-sample prediction

Model fit of the linear models estimated using the GLSS data is shown in Table 2. For each set of proxies and predicted welfare metric, we list the number of variables (including proxies and pairwise interactions between proxies) that are included with non-zero coefficients in our selected model, penalty parameter value chosen, the percentage of total deviation in the outcome variable explained by the model, and the number of observations used to estimate the model. The number of observations used varies between models due to differences in the number of observations with missing values for the proxies belonging to each set. Only observations with non-missing values for both welfare metrics (per capita expenditures and $w$) are used to estimate models that predict either of the benchmarks, for straightforward comparison. For the models that predict per capita consumption expenditures, the amount of deviation explained by the four sets of proxies ranges from 14.5% to 49.1%. Notably, the sets of proxies that are not "sparse" and are mostly comprised of household demographics and asset ownership variables (sets 2 and 3) have the most predictive power. This is unsurprising given that these are the standard types of variables used to predict per capita expenditures. However, it is worth noting that the first set of proxies, which contains less than 50% of the number of proxies used in sets 2 and 3, has more than half of the predictive power of the models using sets 2 and 3. This affirms that even these ten proxies in set 1 are quite predictive of per capita consumption expenditures. On the other hand, proxy set 4, which contains the 6 expenditure category proxies (and region), has only around half of the predictive power of proxy set 1 despite having 70% of the number of variables.

Looking at the models that predict $w$, we see a different picture. The models with proxy sets 2 and 3 explain only 34.2% and 31.1% of the variation in $w$ respectively, which is less predictive power than those proxies had in predicting per capita consumption expenditures. Proxy set 1 predicts around 30% of the variation in $w$, which is similar to the performance of these proxies in predicting per capita expenditures. However, the big difference in the models predicting $w$ is that consumption category proxies in proxy set 4 have the most explanatory power in predicting $w$, explaining 40.3% of overall variation. It is perhaps striking that these 7 proxies in set 4 alone have more predictive power than the much larger sets of proxies in sets 2 and 3. However, this highlights the importance of identifying the proper proxies to best predict a given welfare metric, as they may not be the same across metrics. This is despite the fact that per capita expenditures and $w$ are calculated using the same set of underlying data. It is also worth noting that the relative variance in the distribution of the $w$ measure is greater than that of per capita consumption. The coefficient of variation for $w$ is about 4.5, while the coefficient of variation in per capita consumption is about 0.91. This reflects that superior informational content of $w$ relative to the PCE, and the value of exploiting variation in the not just the scale but the composition of households' consumption baskets.

For the CFE demands-based PMT score formula, we use the variables from proxy set 4 (and five variables related to basic household demographic makeup: these are the number of boys, number of girls, number of men, number of women, and the logarithm of total household size) to estimate $w$. Notably, this model is not designed to estimate total expenditures, and as a result, the percentage of deviation of per-capita expenditures explained by this CFE demands-based score is only 7%. However, for $w$ this estimation procedure performs similarly to the linear models, explaining 34.3% of the variation in $w$.

Table 2: Linear Model Fit Summary

| Outcome Var: | PCE | $w$ | PCE | $w$ | PCE | $w$ | PCE | $w$ |
|---|---|---|---|---|---|---|---|---|
| Proxy Set # | 1 | 1 | 2 | 2 | 3 | 3 | 4 | 4 |
| # variables (w/ interact.) chosen | 86 | 91 | 302 | 334 | 377 | 204 | 62 | 80 |
| Penalty Parameter Value | 6.92 | 0.001146 | 5.81 | 0.001074 | 2.22 | 0.002142 | 7.71 | 0.000427 |
| % of Deviation Explained | 30.5% | 29.8% | 47.3% | 34.2% | 49.1% | 31.1% | 14.5% | 40.3% |
| N | 10387 | 10387 | 10225 | 10225 | 10242 | 10242 | 10408 | 10408 |

Note: "PCE" stands for adult equivalent per capita consumption expenditures.

Table 3: Correlations between welfare metrics and PMT scores

| | | Proxy Set 1 | Proxy Set 2 | Proxy Set 3 | Proxy Set 4 | CFE Estimation |
|---|---|---|---|---|---|---|
| GLSS | PCE | 0.55 | 0.69 | 0.70 | 0.38 | 0.26 |
| GLSS | w | 0.55 | 0.59 | 0.56 | 0.63 | 0.59 |
| LEAP 1000 | PCE | | | 0.25 | 0.19 | 0.40 |
| LEAP 1000 | w | | | 0.09 | 0.31 | 0.99 |

Note: "PCE" stands for adult equivalent per capita consumption expenditures.

## 4 Results: Targeting Accuracy

Now that we have explored model fit and have seen that our models do predict a significant portion of the variation in per capita expenditures and $w$, we now explore the targeting performances of each of our models and how they compare to LEAP's targeting performance. As a first pass, we explore the correlations between our estimated PMT scores and the welfare metrics they aim to predict both in and out of sample in Table 3. In sample, we see that correlations between the PMT scores and the welfare metrics they aim to predict is both large and positive, at around 0.55–0.70. The one exception is that the consumption proxy-based PMT scores do a worse job at predicting per capita consumption, regardless of whether linear or CFE estimation is used, at 0.38 and 0.26 for these methods respectively. Out of sample, the picture is different. None of the linear models seem to perform in an outstanding manner; correlations between the PMT scores and their predicted metrics range from 0.09 to 0.31. (Recall that not all proxies from proxy sets 1 and 2 are included in the LEAP 1000 data, and we hence cannot estimate PMT scores for these sets.) However, the CFE demands-based PMT scores perform much better. For $w$, the correlation with the CFE demands-based PMT scores is 0.99! Even per capita consumption (which is not directly what the CFE demands estimation aims to predict) has a correlation of 0.40 with these PMT scores, which is higher than the correlation for either of the linear model-based PMT scores out of sample. Hence using CFE demands to estimate PMT scores with a limited set of consumption goods may be a reasonable method to predict either of these welfare metrics.

However, for the purposes of targeting performance, we do not necessarily care as much about the overall correlations between PMT scores and the benchmarks they seek to predict. Instead, we are concerned with whether these PMT scores are right "where it counts," in positioning households as either poor/eligible for benefits or non-poor/ineligible for benefits. We also want to benchmark this predictive accuracy against the current targeting accuracy of the LEAP program. We first look at targeting accuracy in-sample in Table 4. Only 1.5% of the GLSS sample are LEAP recipients. So if targeting were perfect then the percentage correctly included would be 1.5%, the percentage correctly excluded would be 98.5%, and the percentages incorrectly included and excluded would be 0%. Note here that the percentages of inclusion and exclusion errors are equal mechanically because we are allocating transfers based on relative poverty status rather than absolute poverty status; that is, we are assessing whether the households assigned transfers are indeed the poorest in their geographical area, not whether each individual household falls below the poverty line. Hence if a household is incorrectly excluded, there must be an incorrectly included household taking its place. Table 4 shows that targeting by LEAP as well as by all of our PMT formulas is far from perfect; the percentage of correctly included households ranges from 0.10% to 0.60%, which is nowhere close to 1.5%. However, some of this discrepancy is likely due to the fact that our analysis does not condition on households meeting all of the eligibility criteria for the LEAP program (such as households having an elderly individual, vulnerable child, etc.). Regardless, some of our estimated PMT scores still show significant targeting improvements over LEAP's current targeting. For targeting per capita consumption expenditures, models using proxy sets 2 and 3 more than double the percentage of households correctly included compared to LEAP's current targeting. Hence, even using similar proxies to the ones that are likely used in LEAP's current targeting procedures could potentially be improved by using a different model, such as the ones we calculate. Notably, the CFE-demands based PMT score is significantly better than LEAP at targeting both per capita expenditures and $w$, with the percentage of households included almost tripling in each case compared to LEAP's current targeting. Hence despite the relatively low correlations between the CFE-based PMT scores and per capita consumption, these scores seem to be accurate "where it counts," in terms of minimizing targeting errors. Hence CFE demands-based PMT scores may be favorable to use even if a policymaker seeks to use per capita consumption expenditures as their targeting benchmark. Moreover, in targeting $w$, the CFE model only targets 1.8% of households incorrectly, which reduces the error rate compared to LEAP and the other models by nearly a third.

Critically, looking at out-of-sample targeting errors tells quite a consistent story. Targeting error rates by both LEAP and our estimated models are in general quite high. If targeting were perfect, then the percentages of correctly included and correctly excluded households would each be around 50%. However, for most of our models these percentages are between 25-35%. In this case, the inaccuracy cannot be explained by households not meeting LEAP eligibility conditions, as all of these households are eligible, should they be sufficiently poor. However, we do see some improvement over LEAP targeting when targeting $w$ with the consumption variables in proxy set 4, and when targeting both per capita consumption and $w$ with the the CFE-based PMT scores. Indeed, targeting of $w$ using the CFE methods is nearly perfect, with only about 2.8% of households incorrectly classified. One may argue that this targeting accuracy is perhaps artificially high given that the $w$ estimates from the LEAP 1000 data and the PMT scores are calculated using the exact same proxy variables as inputs. However, a high correlation was not necessarily guaranteed, given that the PMT scores are calculated using the elasticity estimates from the GLSS data-based model. But even for targeting total consumption aggregates (which would not be subject to this concern as all consumption categories were used), using the CFE demands-based PMT scores decreases targeting error rates by about one-third.

Table 4: In sample targeting errors– GLSS

| | Benchmark | LEAP | Proxy Set 1 | Proxy Set 2 | Proxy Set 3 | Proxy Set 4 | CFE |
|---|---|---|---|---|---|---|---|
| Correctly Included | PCE | 0.10% | 0.17% | **0.22%** | **0.21%** | 0.11% | **0.27%** |
| | | (0.03) | (0.04) | (0.05) | (0.04) | (0.03) | 0.05 |
| Correctly Excluded | PCE | 97.1% | 97.2% | 97.2% | 97.2% | 97.1% | 97.3% |
| | | (0.17) | (0.16) | (0.16) | (0.16) | (0.17) | (0.16) |
| Incorrectly Included | PCE | 1.40% | 1.33% | 1.28% | 1.29% | 1.38% | 1.22% |
| | | (0.12) | (0.11) | (0.11) | (0.11) | (0.12) | (0.11) |
| Incorrectly Excluded | PCE | 1.40% | 1.33% | 1.28% | 1.29% | 1.38% | 1.22% |
| | | (0.12) | (0.11) | (0.11) | (0.11) | (0.12) | (0.11) |
| Correctly Included | $w$ | 0.12% | 0.11% | 0.18% | 0.18% | 0.13% | **0.60%** |
| | | (0.03) | (0.03) | (0.04) | (0.04) | (0.04) | (0.08) |
| Correctly Excluded | $w$ | 97.1% | 97.1% | 97.2% | 97.2% | 97.1% | **97.6%** |
| | | (0.17) | (0.17) | (0.16) | (0.16) | (0.17) | (0.15) |
| Incorrectly Included | $w$ | 1.38% | 1.39% | 1.32% | 1.32% | 1.37% | **0.90%** |
| | | (0.12) | (0.12) | (0.11) | (0.11) | (0.12) | (0.09) |
| Incorrectly Excluded | $w$ | 1.38% | 1.39% | 1.32% | 1.32% | 1.37% | **0.90%** |
| | | (0.12) | (0.12) | (0.11) | (0.11) | (0.12) | (0.09) |

Note: "PCE" stands for adult equivalent per capita consumption expenditures, and "CFE" refers to the CFE demand system-based PMT scores. Bolded numbers suggest a significantly better targeting performance than LEAP's actual targeting ("LEAP" column) using a t-test of means at the 10% level.

Table 5: Out of sample targeting errors–LEAP 1000

| | | LEAP | Proxy Set 3 | Proxy Set 4 | CFE |
|---|---|---|---|---|---|
| Correctly Included | PCE | 26.6% | 27.9% | 26.2% | **34.5%** |
| | | (0.89) | (0.91) | (0.89) | (0.96) |
| Correctly Excluded | PCE | 25.8% | 27.1% | 25.4% | **33.7%** |
| | | (0.89) | (0.90) | (0.88) | (0.96) |
| Incorrectly Included | PCE | 23.8% | 22.5% | 24.2% | **15.9%** |
| | | (0.86) | (0.85) | (0.87) | (0.74) |
| Incorrectly Excluded | PCE | 23.8% | 22.5% | 24.2% | **15.9%** |
| | | (0.86) | (0.85) | (0.87) | (0.74) |
| Correctly Included | $w$ | 25.6% | 25.3% | **31.8%** | **49.0%** |
| | | (0.88) | (0.88) | (0.94) | (1.01) |
| Correctly Excluded | $w$ | 24.9% | 24.6% | **31.0%** | **48.2%** |
| | | (0.87) | (0.87) | (0.94) | (1.01) |
| Incorrectly Included | $w$ | 24.8% | 25.1% | **18.6%** | **1.39%** |
| | | (0.87) | (0.88) | (0.79) | (0.24) |
| Incorrectly Excluded | $w$ | 24.8% | 25.4% | **18.6%** | **1.39%** |
| | | (0.87) | (0.88) | (0.79) | (0.24) |

Note: "PCE" stands for adult equivalent per capita consumption expenditures, and "CFE" refers to the CFE demand system-based PMT scores. Bolded numbers suggest a significantly better targeting performance than LEAP's actual targeting ("LEAP" column) using a t-test of means at the 10% level.

# 5 Conclusion

In this paper, we propose a novel choice of welfare metric with which to assess targeting performance, the $w$ metric presented in Ligon [2020]. We ask whether it is possible to predict $w$ using PMT scores in the same ways in which we would to predict per capita consumption, and whether the most suitable proxies to use for prediction would be the same. We find that we can indeed estimate PMT scores to predict $w$. However the standard types of easily observable asset ownership and household demographic proxies that are often used to predict per capita consumption are less effective at predicting $w$. Yet there exists an alternative set of proxies that can be used to predict $w$ using similar estimation methods, which is a small set of consumption expenditure categories. Moreover, this estimation is relatively efficient; using 10 standard proxies (proxy set 1) in a linear model to predict per capita expenditures can explain 30.5% of variation in this metric, while using 7 proxies (in proxy set 4) can explain 40.3% of the variation in $w$.

Moreover, if instead of using a linear model we estimate PMT scores using CFE demands estimation methods with a restricted set of goods, targeting accuracy rates both in-sample and out-of-sample show significant improvement over LEAP's current targeting regime in predicting both per capita expenditures and $w$. Hence the LEAP program may want to consider the use of CFE demand estimation methods to calculate PMT scores, even if they want to continue using per capita consumption expenditures as their welfare benchmark to target.

At the same time, in the Ghanaian context where most households are poor, the LEAP transfer amount is relatively low, and a fairly small percentage of the population receives any LEAP benefits at all, policymakers may be less inclined to dedicate resources to improve targeting. Moreover, there are other well-documented implementation challenges with LEAP that may be worth addressing first, such as payment delays [Handa et al., 2013], and corruption in the beneficiary selection process [Tweneboah-Koduah, 2020]. Yet, perhaps this exercise can motivate further applications of CFE-based PMT scoring for targeting of other similar anti-poverty programs.

A caveat worth mentioning is that there may be some inherent challenges with the use of consumption expenditure categories as proxies to calculate PMT scores. The types of proxies that are usually used to predict PMT scores (asset ownership and household demographic makeup) were not solely chosen because they are predictive of per capita consumption, but also because they tend to be cheap to observe and verify. For instance, finding out the material that a household's roof is made out of likely takes less time than finding out how much sugar they purchased last week. Moreover, such proxies may be less subject to recall bias and manipulation. It may be challenging for households to remember how much they have spent on particular food items, especially when recall periods are longer than a week or when the individual surveyed does not conduct household food purchases. Additionally, a household that knows program eligibility is based on consumption expenditures could fairly easily misreport food expenditures or change food expenditure behavior in the short term, in a way they likely could not for ownership of very visible assets. Hence for methods using consumption expenditures as proxies to be viable, policy implementers will have to be careful with the ways in which they advertise program eligibility criteria, or employ hybrid targeting methods which require verification of poverty status using observable welfare proxies. However, given the potential of being able to better target the households who are neediest and would benefit most from the transfers using $w$ as a welfare benchmark, policymakers may want to pursue creative solutions that allow for the use of consumption expenditures as proxies.

# References


E. L. Aiken, G. Bedoya, J. E. Blumenstock, and A. Coville. Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. *Journal of Development Economics*, 161:103016, 2023.

V. Alatas, A. Banerjee, R. Hanna, B. A. Olken, and J. Tobias. Targeting the poor: evidence from a field experiment in Indonesia. *American Economic Review*, 102(4):1206–1240, 2012.

J. Banks, R. Blundell, and A. Lewbel. Quadratic Engel curves and consumer demand. *Review of Economics and statistics*, 79(4):527–539, 1997.

C. Brown, M. Ravallion, and D. Van de Walle. A poor means test? Econometric targeting in Africa. *Journal of Development Economics*, 134:109–124, 2018.

A. Deaton. *The analysis of household surveys: a microeconometric approach to development policy*. World Bank Publications, 1997.

J. Friedman, K. Beegle, J. De Weerdt, and J. Gibson. Decomposing response error in food consumption measurement: Implications for survey design from a randomized survey experiment in Tanzania. *Food Policy*, 72:94–111, 2017.

J. Gazeaud. Proxy means testing vulnerability to measurement errors? *The Journal of Development Studies*, 56(11):2113–2133, 2020.

J. Gibson, K. Beegle, J. De Weerdt, and J. Friedman. What does variation in survey design reveal about the nature of measurement errors in household consumption? *Oxford Bulletin of Economics and Statistics*, 77(3):466–474, 2015.

M. Grosh and J. L. Baker. Proxy means tests for targeting social programs. *Living standards measurement study working paper*, 118:1–49, 1995.

S. Handa, M. Park, R. O. Darko, I. Osei-Akoto, B. Davis, and S. Daidone. Livelihood empowerment against poverty program impact evaluation. *Chapel Hill: Carolina Population Center, University of North Carolina*, 2013.

*Ghana 2016 PPI User Guide*. IPA, 2019.

D. Karlan and B. Thuysbaert. Targeting ultra-poor households in Honduras and Peru. *The World Bank Economic Review*, 33(1):63–94, 2019.

E. Ligon. Estimating household welfare from disaggregate expenditures. 2020. URL https://escholarship.org/uc/item/2cg4q0b9.

L. Pop. Options for improving the targeting of safety nets in ghana. *Safety Nets in Africa: Effective Mechanisms to Reach the Poor and Most Vulnerable. The World Bank, Washington DC*, 2015.

P. Premand and P. Schnitzer. Efficiency, legitimacy, and impacts of targeting methods: Evidence from an experiment in Niger. *The World Bank Economic Review*, 35(4):892–920, 2021.

P. Schnitzer and Q. Stoeffler. Targeting social safety nets: Evidence from nine programs in the Sahel. *The Journal of Development Studies*, pages 1–22, 2024.

C. Trachtman, Y. Permana, and G. Sahadewo. How much do our neighbors really know? The limits of community-based targeting. *University of California, Berkeley Working Paper*, 2022.

D. Tweneboah-Koduah. Worthy programme, unworthy implementation: a case of Ghana's livelihood empowerment against poverty. *Asian Journal of Social Sciences and Management Studies*, 7(3):219–228, 2020.

# A Practical Application to the Ghana Living Standards Survey

Open source code for estimating CFE demand systems and households' values of $w$ is available at https://bitbucket.org/ligonresearch/cfedemands. This package is implemented in the `python` programming language. If one has a computer with a functioning `python` installation then installing the package can be accomplished simply with

```
> pip install CFEDemands
```

To estimate a demand system, two sorts of data should be provided, both as `pandas DataFrames`. The first is data on log expenditures. In the application here we use publicly available data from the Ghana Living Standards Survey,[1] and extract data on food expenditures and household characteristics using publicly available open source code from the `LSMS_Library` project.[2] Running this code creates the two necessary dataframes, stored in the open source `parquet` format.

After the creation of these two dataframes, actual estimation using the `CFEDemands` package is as simple as:

```
import cfe.regression as rgsn
import pandas as pd
import numpy as np

y = np.log(pd.read_parquet('food_expenditures.parquet')).squeeze()

z = pd.read_parquet('household_characteristics.parquet')

r = rgsn.Regression(y=y,d=z,alltm=False)

r.graph_beta()  # Estimates Frisch elasticities beta, produces Figure 1

r.get_w()  # Estimates welfare values w, in "nominal" terms
r.get_pi() # Estimates rates of inflation, translates w to "real" terms.
```

[1] https://www2.statsghana.gov.gh/nada/index.php/catalog/97/get_microdata
[2] https://bitbucket.org/eligon/lsms_library/GhanaLSS