Overview: this rubric defines decision rules for selecting predictors from (LSMS-style) household surveys for inclusion as predictors when learning transfer policies. Our goal is to select the predictors that are (a) members of categories of covariates that have actually been used to target real-world programs, e.g. via Proxy Means Tests conducted by governments[1], and that (b) we believe could plausibly be verified, either directly by enumerators (e.g. through direct observation) or indirectly by merging survey responses with other records to which governments may have access (e.g., land ownership registries). Two remarks on this:

(a) We look to define categories of predictors, not individual predictors, that have been used. If we see that governments have used the material out of which the house's walls are made, for example, then we take that as reasonable justification for also using the material out of which other parts of the house such as the roof are made as an equally reasonable predictor. These categories (i.e., alpha-labelled sub-bullets (a), (b), (c), etc) are intended to be exhaustive of the categories of variables within each top-level group (e.g., household demographics) that we would include.

(b) We apply some of our own skeptical judgment in excluding variables that governments have actually used if we cannot see a plausible path to verifying them. For example, responses to questions about food security (such as whether household members skipped meals in the past week) have been used in some PMTs, but we cannot see a plausible way to verify such claims and so omit this variable.

After each (sub)category we provide examples from specific real-world PMTs documenting that the category has been used in practice and illustrating its range.

Categories

(1) Special data: This info is required for our analysis, even though it does not function as PMT input.
    (a) Consumption aggregate: This should include *all* consumption, not just spending.
    (b) Household survey weight/expansion factor.
(2) Household demographics
    (a) Counts of household members by type, including the overall number of household members, the number of male and female members, the numbers in various age-based subcategories such as children, adults, elderly adults / senior citizens, and the number of specially ~abled members, etc. If the

[1] Variables included in various PMTs are documented here for posterity, but please use this guide, *not* that document, to determine whether a covariate should be included.

survey does not pre-define age-based categories then we will ourselves count the number of children (17 and under, per UNICEF's definition) as well as the number of children in different age sub-groups (for e.g, no of children aged 0-4 (Alatas et al. 2012); the number of adults (18 and over); and the number of elderly adults (60 and over as in BPL Guidelines or 65 and over (Alatas et al. 2012 and, Hanna and Olken, 2018).

(b) Household head: Age, gender, and marital status (Alatas et al. (2012) include married/unmarried, Fernandez and Hadiwidjaja (2018) include single/married/divorced, while Kidd and Wylde (2011) include widowhood as well)

(c) Ethnicity, tribe, or caste of the household / household head (e.g, BPL Guidelines for India include the scheduled caste or tribe status of households as a deprivation indicator)

(3) Human capital

(a) Educational attainment of the household head. Depending on the source this may be measured in years, as the highest level completed (primary, secondary, etc; Alatas et al. 2012), simply with a literacy indicator (BPL Guidelines for India) etc.

(b) Maximum educational attainment of all adult members (as in Kidd and Wylde, 2011). As for the head this may be measured in years, in level completed, etc.

(i) *Remark: we take the max rather than including the attainment of all members individually since the number of members varies across households, and we do not want the number of predictors to vary.*

(c) Maximum educational attainment of all female adult members (e.g, Kidd and Wylde, 2011) include literacy status)

(d) Number of children currently enrolled in school

(i) *Remark: because we also include the total number of children (above), this is sufficient to let the model consider the ratio as a potential predictor.*

(4) Household assets—presence and number owned, in particular physical assets, with potentially one exception noted below

(a) We do *not* include ownership-structure details beyond what a household owns. This may be relevant for e.g. agricultural land.

(b) Dwelling characteristics such as per capita number of rooms (Kidd and Wylde, 2011) or floor space (Alatas et al. 2012), material used to construct floors, walls, or roof, etc., as well as an indicator for homeownership itself. For e.g, Alatas et al. (2012) use binary indicators for home-ownership (private vs not) and roof material (concrete roof vs not) whereas Hanna and Olken (2018) consider more granular categories - Is the roof made of concrete/asbestos/shingles/zinc?; Is the house privately owned/rented/occupied for free?

    (i) We do *not* include other details as to the ownership/non-ownership arrangement of the home, such as ownership document type or presence, rent, etc.

  (c) Presence and physical characteristics of amenities such as type of latrine, water source, lighting source, drainage system, waste collection, access to electricity and gas, type of cooking fuel, cable connection. For e.g, water source could be recorded as a simple binary indicator for availability of clean water within the house (Alatas et al. 2012) or subject to availability, could be recorded in more granular categories: bottle, tap, pump, protected or unprotected well (Hanna and Olken, 2018).
  (d) Consumer durables such as appliances (e.g., radio, television, refrigerator, generator, cooker, heater, fan, AC etc), transportation (e.g., car, bicycle, motorbike), furniture (e.g., sofa, bed etc), and devices (e.g., computers)
  (e) Productive agricultural assets including land, livestock, irrigation facilities, farm machines (e.g., tractors (BPL Guidelines) or other mechanized instruments)
    (i) We **exclude** cultivation details: crop types, amounts, etc, as they may not be verifiable.
    (ii) We **include** the amount of land owned.
  (f) Productive non-agricultural assets, if these are listed separately from household assets (for example, a sewing machine might be listed as a household asset or as an enterprise asset)
  (g) Financial assets — we generally **exclude** these as we expect the kinds of financial assets held by poor households to be hard to verify, but would include any that they hold as part of a government scheme for which the government might plausibly hold records they could use to verify possession. An example of this is the Kisan credit card scheme in India (BPL Guidelines).
  (h) We **exclude** crop stores, as they may not be easy to verify.

(5) Livelihood activities
  (a) Primary sector of employment of the household head (e.g., agriculture, manufacturing, services)
  (b) Primary occupation of the household head (e.g., self-employed, salaried employee, casual laborer, etc.)
  (c) We do *not* include the occupations of other members of the household
  (d) Ownership of enterprises
  (e) Receipt of other public transfers, including amount, if this could plausibly be verified by merging in other administrative records. For example, social

security transfers (Camacho and Conover, 2011) can, in principle, be verified using such records.

(6) Geographic indicators
    (a) Urban / rural status, including any available classifications such as peri-urban
    (b) All administrative geographic information: Which administrative division of any level a household is in
    (c) Distance to important locations such as district centers (Alatas et al. 2012), markets (Alatas et al. 2012), or public facilities such as a post office (Hanna and Olken, 2018). (Note that availability at "distance 0", i.e. within the community itself, may be captured separately by the facilities variables described below.)
    (d) Environmental conditions such as rainfall history.
        (i) *Remark: this reflects cases in which environmental conditions are used as part of the larger eligibility determining process, e.g. if drought-prone regions are selected and then a PMT is run to select households within them, even if the household-level determination does not depend on the environmental variable. del Ninno & Mills (2015) provide a few case studies that illustrate feasibility.*

(7) Community characteristics
    (a) Presence of publicly or privately provided services such as healthcare, e.g. whether there is a doctor (Alatas et al. 2012) or a midwife (Kidd and Wylde, 2011) who works within the community.
    (b) Presence of publicly or privately provided infrastructure such as paved roads, banking facilities (Kidd and Wylde, 2011 and Alatas et al. 2012) , or regional government offices (Kidd and Wylde, 2011).
    (c) Population characteristics such as headcount or population density (Kidd and Wylde, 2011)

Dropped variables

The following are examples of variables that are included in some real-world Proxy Means Tests but that we omitted on the grounds that we could not see a plausible path to verifying them at large scale:

- Age of dwelling
- Indicators of food security
- Use of fertilizer
- Use of an internet connection

- Ownership of a mobile phone, on the grounds that these can easily be concealed, and related questions about usage of someone else's phone (we may use this information to link CDR, but not as a PMT predictor in itself absent some scheme for verification).
- Income and private transfers (like foreign remittance)