LATE with Missing or Mismeasured Treatment

Rossella Calvi[†]   Arthur Lewbel[‡]   Denni Tommasi[§]

August 2021

Abstract

We provide a new estimator, MR-LATE, that consistently estimates local average treatment effects when treatment is missing for some observations, not at random. If instead treatment is mismeasured for some observations, MR-LATE usually has less bias than the standard LATE estimator. We discuss potential applications where an endogenous binary treatment may be unobserved or mismeasured. We apply MR-LATE to study the impact of women's control over household resources on health outcomes in Indian families. This application illustrates the use of MR-LATE when treatment is estimated rather than observed. In these situations, treatment mismeasurement may arise from model misspecification and estimation errors.

JEL Codes: D13, D11, D12, C31, I32.

Keywords: LATE, missing treatment, measurement error, misclassification, collective model, resource shares, health.

[†]Department of Economics, Rice University (e-mail: rossella.calvi@rice.edu).

[‡]Department of Economics, Boston College (e-mail: arthur.lewbel@bc.edu).

[§]Department of Economics, University of Bologna (e-mail: denni.tommasi@unibo.it).

1 Introduction

We consider the estimation of a Local Average Treatment Effect (LATE) when treatment is either missing not at random for some observations, or is generally mismeasured. To deal with these issues, we propose a new estimation method (Mismeasurement Robust LATE or MR-LATE) that can point-identify and consistently estimate LATE when the treatment indicator contains missing values, and can reduce estimation bias when the treatment is sometimes misclassified. The causal effect we focus on identifying and estimating is the LATE of Imbens and Angrist (1994), which is applicable when the true treatment is endogenous, an exogenous binary instrument is available, and treatment effects may be heterogeneous. In this framework, measurement error in the binary treatment variable is, by construction, non-classical (i.e., it is correlated with the true value of the regressor).[1] As a result, the standard instrumental variable approach to mismeasured regressors does not yield unbiased and consistent estimates of LATE.

Some methods exist to deal with misclassification of an endogenous binary treatment variable. Ura (2018) and Tommasi and Zhang (2020a) consider estimation of LATE with mismeasured treatment and standard LATE instrument assumptions (i.e., exclusion restriction and monotonicity), but only obtain set identification bounds. Battistin et al. (2014) use two measures of the misclassified treatment to point-identify LATE, but require re-survey data (i.e., multiple observations of the same individuals). DiTraglia and García-Jimeno (2019) and Yanagi (2018) also obtain point-identification of LATE with mismeasured treatment. The former, however, requires that treatment effects be homogeneous, while the latter requires the availability of two instrumental variables with specific properties, one for the endogenous treatment and the other to deal with the measurement error. We achieve point-identification without these requirements by relying on two misclassified treatment indicators and by imposing restrictions on the misclassification probabilities. Even when these restrictions are violated, under much weaker conditions, our MR-LATE estimator can still reduce the bias in LATE estimation due to measurement errors.

Item nonresponse, misreporting, contamination, or misclassification of key variables are pervasive issues in economic data. For example, O'Connell (2006) argues that up to 20 percent of information on marital histories in the 1996 Survey of Income and Program Participation is missing or inconsistent. Molinari (2010) shows that up to 16 percent of the 1984 National Longitudinal Survey of Youth respondents refuse to state whether they consumed alcohol during pregnancy. Meyer et al. (2015) document dramatic nonresponse rates (up to 40 percent) in five prominent US household surveys for questions regarding the receipt of government transfers such as the Temporary Assistance for Needy Families (TANF) or the Supplemental Nutrition Assistance Program (SNAP). Any empirical analysis of, e.g., the effect of divorce on labor market outcomes or the impact of a mother's drinking during pregnancy on her child's birth weight or the effectiveness of government programs based on such surveys would inevitably rely on strong ignorability or missing at random assumptions. In practice, researchers often resort to dropping observations despite the resulting selection biases, or impute missing responses under a variety of modeling assumptions (see Little

---

[1]A binary treatment can only be misclassified as zero if its true value is one, and vice versa. So, by construction, there must exist a negative correlation between the true value of the treatment and the measurement error.

and Rubin (2019) for an in-depth review of the literature).[2] Such practices, however, usually lead to inconsistent parameter estimates and jeopardize the validity of causal analyses.

Missing or mismeasured treatment is a common problem, especially in applications where treatment is observed rather than directly administered by researchers. However, we would like to emphasize that MR-LATE could also be used in applications where treatment is not observed at all, but can be estimated. There are many examples of potential treatments of clear economic significance that are rarely analyzed causally because the treatments themselves cannot be directly observed. This could include uses of machine learning to estimate treatment (for a survey, see Athey and Imbens (2019)). Examples of such treatments are measures involving expectations, ability, opportunity, utility, risk aversion, or welfare. One might be interested, for example, in quantifying the effects of high discount rates on educational outcomes (like the probability of dropping out of school), or the impact of risk aversion on investment decisions, or the effect of abilities such as cognitive or non-cognitive skills on future earnings. Any such analyses require estimation (typically structural or machine learning-based model estimation) of the treatment. Estimated treatments will generally suffer from measurement errors, because of both model misspecification and unobserved model error terms. To illustrate this point, we apply MR-LATE to study the impact of women's control over household resources on health outcomes in Indian families. In this application, we use a structural model to estimate the treatment (women's resource control), and a LATE framework to estimate the causal effect of this treatment on health. Our MR-LATE estimator deals with errors arising from the structural model, including misspecification error.

In our application, we estimate treatment using the collective household model of Chiappori (1988; 1992), in which a family is characterized as a collection of individuals with separate utility functions, and the allocation of goods is assumed to be Pareto efficient. Using this model, we structurally estimate a measure of individual-level resource control based on observed household-level expenditures. Specifically, we estimate men's and women's resource shares (i.e., the fraction of a household's resources controlled by each decision maker) using a methodology developed in Dunbar et al. (2013). We define a household to be treated if the woman is *empowered*, in the sense of having control over a substantial share of the household's resources.

Applying our MR-LATE estimator, we then estimate a causal effect of the treatment (empowered women) on the health status of family members. Our instrument exploits changes to the Indian inheritance law that, by granting women the right to inherit their natal family property, improved their outside option and strengthened their bargaining power in their marital families (Heath and Tan, 2019; Calvi, 2020). In particular, we focus on the Hindu Succession Act amendments that equalized women's inheritance rights to men's in several Indian states between 1976 and 2005, and only applied to Hindu, Buddhist, Sikh or Jain women who were not yet married at the time of the amendment in their state. We find that accounting for specification, estimation, and/or measurement error in the estimated treatment variable is empirically important, with some substantial differences between the standard 2SLS estimator for LATE, which cannot control for such measurement errors, and our MR-LATE estimator. We estimate that women's control over resources positively affects

---

[2]Missing treatment observations can be ignored when the treatment is missing at random or the distribution of treatments between respondents and nonrespondents is the same. These assumptions, however, are rarely satisfied in empirical research (Molinari, 2010).

women's and children's health outcomes at no cost to men's health outcomes.

This paper relates to the literature on the identification and estimation of treatment effects in the presence of missing data. Measurement (misclassification) error in a key variable such as the treatment status "can be thought of as a special case of missing data" (Bound et al. (2001), p. 3739). Papers empirically documenting substantial misclassification errors in observed treatments include Bollinger (1996), Angrist and Krueger (1999), Kane et al. (1999), Card (2001), Black et al. (2003), and Hernandez et al. (2007). A few previous papers have considered alternative techniques for dealing with misclassification errors in treatment. Homogeneous treatment effects, corresponding to the estimation of constant coefficients of a mismeasured binary regressor, have been estimated using instruments by many authors, including Aigner (1973), Kane et al. (1999), Black et al. (2000), and Frazis and Loewenstein (2003). When treatment is misclassified, point-identification and associated estimators of average treatment effects (without assuming treatment effects are homogeneous) are provided by Mahajan (2006), Lewbel (2007) and Hu (2008). These papers obtain identification exploiting both the assumption that the true treatment is exogenously determined and an assumed instrument to deal with the misclassification error. Under more general conditions, bounds on average treatment effects with misclassified treatment are provided by Klepper (1988), Manski (1990), Bollinger (1996), Kreider and Pepper (2007), Molinari (2010), Imai et al. (2010), and Kreider et al. (2012).

Our estimation problem has the standard LATE structure that a randomized binary instrument is correlated with an endogenous binary treatment, and the true treatment affects an outcome. But in our case, we must overcome the issue that the treatment is missing for some observations or the observed treatment does not equal the true treatment. A similar structure arises in models where outcomes of interest and randomized treatment are not available in the same data set. In these models, a randomized treatment (corresponding to our instrument) affects an intermediate outcome called a statistical surrogate (see Prentice (1989)), corresponding to our mismeasured treatment indicator. The surrogate then affects (or at least strongly correlates with) the outcome of interest. These estimators require that the surrogate satisfies a strong conditional independence assumption (see, e.g., Rosenbaum (1984); Begg and Leung (2000); Frangakis and Rubin (2002); VanderWeele (2015)). Athey et al. (2016) overcome this limitation by observing multiple surrogates, each of which may not satisfy the required conditional independence. They assume that there exists a single latent, unobserved surrogate that has the desired properties, and combine the observed surrogates to model the impact on the outcome of the underlying latent surrogate. In a roughly analogous way, we exploit multiple (two) mismeasures of treatment to model the impact on the outcome of an underlying latent (true) treatment. Beyond this analogy, however, the details of their model, their estimator, and their underlying assumptions differ substantially from ours.

The estimation of the treatment indicator in our empirical application is based on the collective household framework pioneered by Becker (1965, 1981), Apps and Rees (1988) and Chiappori (1988, 1992). In this framework, each household is characterized as a collection of individuals, each of whom has a well-defined utility function, and who interact to generate Pareto efficient allocations. Recent advances in this literature permit the recovery of resource shares (or sharing rule), defined as each member's share of total household consumption (Lewbel and Pendakur (2008), Browning et al.

(2013), Dunbar et al. (2013)). Our specific model is based on Dunbar et al. (2013) and Calvi (2020), who obtain resource shares from estimates of Engel curves (demand equations holding prices constant) of clothing items that are consumed exclusively by women, men, or children.[3] Our analysis using MR-LATE estimates the causal effect of the structurally estimated treatment on family health outcomes. So, from a policy perspective, our empirical analysis adds to the broad literature on the effect of women's empowerment on adults' and children's outcomes. Economic analyses of bargaining power within the household have made use of varying proxies of resource control, such as assets (e.g., in Beegle et al. (2001); Frankenberg et al. (2001); Quisumbing and Maluccio (2003); LaFave and Thomas (2017)), or unearned income (e.g., Thomas (1990); Duflo (2003)).[4] Departing from this literature, our analysis defines treatment based on a structurally estimated measure of resource control, rather than relying on some proxy.

The remainder of the paper is organized as follows. The next section provides a description of the MR-LATE estimator. Then, we present the formal derivation of the MR-LATE estimator and its properties, including an analysis of bias and a Monte Carlo study. We then describe several potential applications, i.e., situations where an endogenous binary treatment may be unobserved or mismeasured. This is followed by our empirical application, and then conclusions. Proofs and additional material are in an online Appendix.

2   LATE Estimation with Missing or Mismeasured Treatment

Before providing formal assumptions and derivations, we here show how to construct our MR-LATE estimator, and illustrate its data requirements. We later give examples of applications containing these types of data.

2.1   Case 1: Correctly measured treatment with no missing observations

The standard data requirements for LATE estimation are that, for individuals $i = 1, ..., n$, we observe an outcome Y_i, a binary treatment indicator D_i, and a binary instrument Z_i. The Imbens and Angrist (1994) LATE estimator is then the coefficient of D_i in a linear instrumental variables regression of Y_i on a constant and on D_i, using as instruments a constant and Z_i. Equivalently, the Imbens and Angrist (1994) LATE estimator is:

$$\widehat{\ell} = \frac{\widehat{cov}\,(Y, Z)}{\widehat{cov}\,(D, Z)} \tag{1}$$

where, for any random variable X, $\widehat{cov}\,(X, Z) = \sum_{i=1}^{n} X_i(Z_i - \overline{Z})/n$ and $\overline{Z} = \sum_{i=1}^{n} Z_i/n$. Under standard conditions, the probability limit of $\widehat{\ell}$ is the LATE ℓ, i.e., the average treatment effect for compliers.

[3]Other applications or extensions of this approach include Penglase (2017), Sokullu and Valente (2017), Bargain et al. (2018), Brown et al. (2018), Tommasi (2019), Calvi et al. (2020), Calvi and Keskar (2021a).

[4]For example, Thomas (1990) documents that child health in Brazil tends to improve if additional non-labor income is in the hands of women rather than men. He estimates that income in the hands of a mother has, on average, twenty times the impact of the same income in the hands of a father with respect to children's survival probabilities. Duflo (2003), studying elderly benefits in South Africa, concludes that the same transfer has drastically different impacts on the health of female grandchildren depending on whether it is paid to the grandmother or to the grandfather.

2.2 Case 2: Missing treatment observations

Now suppose that for individuals $i = 1, ..., n$, we observe an outcome Y_i and a binary instrument Z_i, but we only observe the treatment indicator D_i for some subset of these individuals. If the missing observations of D_i are missing at random (they are *ignorable*), then there would be no inconsistency from simply dropping the individuals with missing treatment from the sample and implementing the LATE estimator $\widehat{\ell}$ above on the remaining data. Note that dropping observations with missing data is very common in empirical practice. In this paper, we do not assume data are missing at random, so dropping the observations where D_i is unobserved may in general lead to inconsistent estimates. We also assume we have no information that can be used to guess or impute what D_i might equal for the individuals i where D_i is not observed.

Given this data, we construct two binary variables, called T_i^a and T_i^b. Both T_i^a and T_i^b are set equal to zero for observations i where D_i is unobserved; for observations where D_i is observed, T_i^a is set equal to D_i while T_i^b is set equal to $1 - D_i$. Our MR-LATE estimator is then:

$$\widehat{\rho} = \frac{\widehat{cov}\left(T^a Y, Z\right)}{\widehat{cov}\left(T^a, Z\right)} - \frac{\widehat{cov}\left(T^b Y, Z\right)}{\widehat{cov}\left(T^b, Z\right)}. \tag{2}$$

As is clear from the above equation, $\widehat{\rho}$ is equivalent to the difference between an instrumental variables regression of $Y_i T_i^a$ on a constant and T_i^a, and an instrumental variables regression of $Y_i T_i^b$ on a constant and T_i^b, using a constant and Z_i as instruments in both cases. What we prove in this paper is that, under some mild conditions, the probability limit of $\widehat{\rho}$ is the same as that of $\widehat{\ell}$, the LATE estimator that would only be feasible if treatment were observed for everyone in the sample.

It is instructive to think of T_i^a and $1 - T_i^b$ as two different treatment indicators, each with different measurement errors. T_i^a differs from the true D_i only for individuals i whose treatment status D_i is unobserved and whose actual treatment status is $D_i = 1$. So, the measure T_i^a has only one kind of measurement error: if $D_i = 0$ then $T_i^a = 0$, but sometimes $D_i = 1$ and $T_i^a = 0$. By contrast, $1 - T_i^b$ as a treatment measure only has the opposite kind of measurement error: if $D_i = 1$ then $1 - T_i^b = 1$, but sometimes $D_i = 0$ and $1 - T_i^b = 1$.

2.3 Case 3: Generally Mismeasured Treatment

Suppose now that, for individuals $i = 1, ..., n$, what we observe is an outcome Y_i, a binary instrument Z_i, and some imperfect information about who is treated and who is not. In particular, suppose we have enough information to sort individuals into at least three groups: those with a high probability that they were treated, those with a low probability, and everyone else (individuals in this third group can either have a middling probability or a completely unknown probability of having been treated). Then, let $T_i^a = 1$ only for those with a high probability of being treated, and set $T_i^a = 0$ for everyone else. Similarly, let $T_i^b = 1$ only for individuals with a low probability of being treated, and set $T_i^b = 0$ for everyone else. We show in this setup that the MR-LATE estimator $\widehat{\rho}$ generally has lower bias than the alternative of just using the ordinary LATE estimator $\widehat{\ell}$, after replacing the unknown D_i with an estimate of D_i. Note that Case 2 above is just a limiting version of Case 3, where those with $T_i^a = 1$ have probability one of being treated, while those with $T_i^b = 1$ have probability

zero of being treated.

When would we have enough information to sort individuals into high, low, and unknown probabilities of being treated? One case would be when multiple proxies of treatment are observed. People where all proxies indicate treatment would be high probability, those where no proxies indicate treatment would be low probability, and everyone else could be assigned to the unknown probabilities group. Other cases with sufficient information are where probability of treatment is estimated (e.g., by a structural or machine learning model).

3 Derivation of the MR-LATE Estimator

Maintaing the notation of the previous section, D is the true binary treatment variable that affects an outcome of interest. Assume D is not observed and cannot be consistently estimated. Let Z be an unconfounded (e.g., randomized) binary instrument that is correlated with D and satisfies the standard (Imbens and Angrist (1994)) assumptions of an instrument for LATE estimation. In what follows, we ignore additional covariates X, as everything immediately extends to conditioning on them.[5]

Let the random binary variables D_0 and D_1 denote the potential treatments $D_z = D(z)$ for possible realizations z of Z. By definition,

$$D = (1 - Z) D_0 + Z D_1. \tag{3}$$

Let Y be an observed outcome of interest and let random variables Y_0 and Y_1 be the potential outcomes $Y_d = Y(d)$ for possible realizations d of D. Then,

$$Y = (1 - D) Y_0 + D Y_1 = Y_0 + (Y_1 - Y_0) [(1 - Z) D_0 + Z D_1]. \tag{4}$$

Assumption 1. *Y and D satisfy the standard Imbens and Angrist (1994) LATE assumptions:*
i. $0 < E(D) < 1$, $0 < E(Z) < 1$ and $Z \perp (Y_1, Y_0, D_1, D_0)$.
ii. (Y_1, Y_0, D_1, D_0, Z) are independent across individuals and have finite means.
iii. There are no defiers, so $\Pr(D_0 = 1 \text{ and } D_1 = 0) = 0$.

Let C denote a complier, i.e., someone who has $D_0 = 0$ and $D_1 = 1$. As discussed earlier, if D was observed, then, under the conditions listed in Assumption 1 above, the Imbens and Angrist (1994) LATE would be identified by the instrumental variable estimand:

$$\text{LATE} = \ell = \frac{cov(Y, Z)}{cov(D, Z)} = E(Y_1 - Y_0 \mid C) \tag{5}$$

Since we do not observe D, however, we cannot implement this standard approach.

Instead of D, consider observing a binary treatment indicator T, which could be a proxy for

---

[5]To extend our results to include exogenous covariates, simply replace all of our expectations and covariances with conditional expectations and covariances, conditional on the covariates. Since MR-LATE is equivalent to a difference in linear instrumental variable estimators, when covariates affect the model linearly, conditioning is equivalent to including the covariates as additional regressors and instruments in the IV (as is common practice in applications of LATE estimators).

or an estimate of D, or could correspond to reported values of D that are mismeasured for some observations (later we will make use of two such T indicators, but just consider one for now).[6]

Define random variables T_0 and T_1 as potential observed treatments (or potential estimated treatments) so $T_d = T(d)$ for possible realizations d of D. Then, by definition,

$$T = (1 - D) T_0 + D T_1. \tag{6}$$

The variables T_0 and T_1 can be interpreted as indicators of whether T correctly measures treatment or not. In particular, if $T_0 = 0$ and $T_1 = 1$, then treatment is not mismeasured. There are two possible types of measurement or classification error in T: if $T_0 = 1$ for some individual, then that means a true $D = 0$ for that individual is misclassified as treated by T, and if $T_1 = 0$, then a true $D = 1$ for that individual is misclassified as untreated.

Assumption 2. *T is such that the following conditions are satisfied:*
    i. $Z \perp (Y_1, Y_0, D_1, D_0, T_1, T_0)$.
    ii. $(T_1, T_0) \perp (Y_1, Y_0) \mid C$.
    iii. $E(T_1 - T_0 \mid C) \neq 0$ and $Pr(C) = Pr(D_0 = 0$ and $D_1 = 1) > 0$.

Assumption 2-i just combines the LATE unconfoundedness assumption that $Z \perp (Y_1, Y_0, D_1, D_0)$ with the assumption that the instrument is also independent of the potential measurement errors, and hence of (T_1, T_0). The standard assumption that Z is randomized by experimental or quasi-experimental design is sufficient to make 2-i hold.

Assumption 2-ii says that, for compliers, the potential mismeasures (T_1, T_0) are independent of the potential outcomes (Y_1, Y_0). A standard assumption in models with mismeasured regressors is that the measurement errors are unrelated to the true outcomes. In the statistics literature, a condition like this is known as "non-differential measurement error." Assumption 2-ii (when combined with unconfoundedness) is a little weaker, saying that this independence between measurement errors and potential outcomes only needs to hold for compliers. A sufficient condition for Assumption 2-ii to hold is that $(T_1, T_0) \perp (Y_1, Y_0, D_1, D_0)$, meaning that the measurement errors are independent of all potential outcomes and potential treatments. This may be somewhat easier to interpret, since it does not depend on compliers, but is stronger than necessary. In our empirical application, this assumption is plausible because the measurement errors in treatment should be largely due to idiosyncratic estimation error. [7] Assumption 2-iii is a minimal relevance condition saying that, at least for compliers, T provides some information regarding D. This assumption implies that, at least for compliers, the correlation between D and T is nonzero.

Let $p_d = E(T_d \mid C)$. By definition, p_1 is the probability that a complier would have their treatment correctly observed if they were assigned the true treatment $D = 1$. That is, p_1 is the probability that a complier would have $T = 1$ if they were assigned $D = 1$. In contrast, p_0 is the probability that a complier would have their treatment *incorrectly* observed (meaning $T = 1$) if they were assigned the

---

[6]Although we observe T and not D, each individual's behavior is still based on their actual D. This means that introducing measurement error does not change the no defiers assumption. If we had incorrectly assumed behavior was based on the mismeasured T, and estimated LATE using T in place of D, then what would appear to be defiers could exist. That would be just one of multiple sources of bias in LATE estimates that ignore the measurement error.

[7]If this stronger but unnecessary assumption holds, then q defined below also satisfies $q = E(T_0) / E(T_1 - T_0)$.

true treatment $D = 0$. Note that Assumption 2-iii ensures that $p_1 - p_0$ is nonzero and that compliers exist.

Define q and λ as follows:

$$q = \frac{p_1}{p_1 - p_0} \tag{7}$$

$$\lambda = \frac{cov(YT, Z)}{cov(T, Z)} = \frac{E(YT \mid Z = 1) - E(YT \mid Z = 0)}{E(T \mid Z = 1) - E(T \mid Z = 0)}. \tag{8}$$

If we were to ignore measurement error in T, we would estimate LATE as in Imbens and Angrist (1994) by an instrumental variables regression of Y on T using Z as the instrument, which would asymptotically equal $cov(Y, Z) / cov(T, Z)$. Instead, λ equals the limiting value of an instrumental variables regression of YT on T using Z as the instrument.[8] In the following Theorem, we show that λ is a mixture of the potential outcomes for compliers.

Theorem 1. *Let Assumptions 1 and 2 hold. Then:*

$$\lambda = E\left[qY_1 + (1 - q) Y_0 \mid C\right]. \tag{9}$$

Proof. Substituting equation (3) into equation (6) gives

$$T = T_0 + (T_1 - T_0) D = T_0 + (T_1 - T_0) \left[(1 - Z) D_0 + Z D_1\right] \tag{10}$$

Multiplying equation (4) by equation (10) gives

$$YT = \left[Y_0 + (Y_1 - Y_0) \left[(1 - Z) D_0 + Z D_1\right]\right] \left[T_0 + (T_1 - T_0) \left[(1 - Z) D_0 + Z D_1\right]\right]$$

Using assumption 2-i, this makes

$$
\begin{aligned}
E(YT \mid Z = 1) &= E\left[\left[Y_0 + (Y_1 - Y_0) D_1\right] \left[T_0 + (T_1 - T_0) D_1\right]\right] \\
&= E\left[T_0 Y_0 + (Y_1 T_1 - Y_0 T_0) D_1\right]
\end{aligned}
$$

where the last equality uses $D_1^2 = D_1$. Similarly,

$$E(YT \mid Z = 0) = E\left[T_0 Y_0 + (Y_1 T_1 - Y_0 T_0) D_0\right]$$

So,

$$E(YT \mid Z = 1) - E(YT \mid Z = 0) = E\left[(Y_1 T_1 - Y_0 T_0)(D_1 - D_0)\right]$$

Given the no defiers assumption, either $D_1 - D_0 = 0$ or $D_1 - D_0 = 1$, and someone is a complier if and only if they have $D_1 - D_0 = 1$. The probability of being a complier is $\Pr(D_1 - D_0 = 1) = E[D_1 - D_0]$.

---

[8]Abadie (2002) also makes use of this λ, though for a different purpose.

We therefore apply the standard LATE logic:

$$E\,(YT \mid Z = 1) - E\,(YT \mid Z = 0) = E\,[(Y_1 T_1 - Y_0 T_0)\,(D_1 - D_0)]$$
$$= E\,[Y_1 T_1 - Y_0 T_0 \mid D_1 - D_0 = 1]\,\Pr\,(D_1 - D_0 = 1)$$
$$= E\,(Y_1 T_1 - Y_0 T_0 \mid C)\,E\,(D_1 - D_0)\,.$$

Let $p_d = E\,(T_d \mid C)$. Then, using Assumption 2-ii, the above further simplifies to

$$E\,(YT \mid Z = 1) - E\,(YT \mid Z = 0) = E\,(p_1 Y_1 - p_0 Y_0 \mid C)\,E\,(D_1 - D_0)\,.$$

Replacing Y with one gives

$$E\,(T \mid Z = 1) - E\,(T \mid Z = 0) = E\,(p_1 - p_0 \mid C)\,E\,(D_1 - D_0)$$
$$= (p_1 - p_0)\,E\,(D_1 - D_0)\,.$$

And therefore
$$\frac{E\,(YT \mid Z = 1) - E\,(YT \mid Z = 0)}{E\,(T \mid Z = 1) - E\,(T \mid Z = 0)} = \frac{E\,(Y_1 p_1 - Y_0 p_0 \mid C)}{p_1 - p_0}$$

which equals equation (9), thereby proving the Theorem. $\qquad\square$

Elements of Theorem 1, and results related to Theorem 1, appear in some earlier work, including Abadie (2002), Ura (2018), and references therein. Our primary novelty is in how we make use of the relationship given by Theorem 1.

Assume that we observe *two* different mismeasures of treatment, called T^a and T^b. These could be, for instance, two proxies or two different estimates of D. Recalling that $p_d = E\,(T_d \mid C)$, let $p_d^a = E\,(T_d^a \mid C)$ and $p_d^b = E\,(T_d^b \mid C)$, where T_d^a and T_d^b are the potential mismeasured treatments associated with T^a and T^b. Similarly, define q^a, q^b, λ^a, λ^b, and our MR-LATE estimand ρ as follows:

$$q^a = \frac{p_1^a}{p_1^a - p_0^a},\quad q^b = \frac{p_1^b}{p_1^b - p_0^b},$$
$$\lambda^a = \frac{cov\,(T^a Y, Z)}{cov\,(T^a, Z)},\quad \lambda^b = \frac{cov\,(T^b Y, Z)}{cov\,(T^b, Z)},\ \text{and}$$

$$\text{MR-LATE} = \rho = \lambda^a - \lambda^b = \frac{cov\,(T^a Y, Z)}{cov\,(T^a, Z)} - \frac{cov\,(T^b Y, Z)}{cov\,(T^b, Z)} \tag{11}$$

This Corollary follows immediately from Theorem 1:

Corollary 1. *Let Assumption 1 hold, and let Assumption 2 hold with $T = T^a$ and with $T = T^b$. Then,*

$$\rho = \left(q^a - q^b\right) E\,[Y_1 - Y_0 \mid C] = \left(q^a - q^b\right) \ell. \tag{12}$$

Equation (12) in Corollary 1 shows the connection between the true LATE ℓ given by equation (5) (which in our application cannot be estimated because D is not observed) and our MR-LATE estimand ρ. Corollary 1 has some straightforward implications, one of which is the following:

10

Corollary 2. *Let Assumption 1 hold, and let Assumption 2 hold with $T = T^a$ and with $T = T^b$. If $q^a - q^b = 1$, then $\rho = \ell$. A sufficient condition for $\rho = \ell$ is $p_0^a = p_1^b = 0$.*

The fact that the MR-LATE estimand ρ equals the true LATE ℓ when $p_0^a = p_1^b = 0$ follows from equations (7) and (12). Case 2 in the previous section is an example of Corollary 2. This is because, when D is observed for some individuals and not others, by constructing T^a and T^b as described in Case 2 of the previous section, we make $p_0^a = p_1^b = 0$.[9] One way to interpret MR-LATE is that $\lambda^a = q^a \ell + E(Y_0|C)$ and $\lambda^b = q^b \ell + E(Y_0|C)$, where q^a and q^b are certain ratios of probabilities. T^a is chosen to drive q^a up towards one, and T^b is chosen to drive q^b down towards zero, so that $\lambda^a - \lambda^b = (q^a - q^b) \ell$ is driven to equaling LATE ℓ.

Assuming we observe the vector (Y_i, Z_i, T_i^a, T_i^b) for individuals $i = 1, ..., n$, the MR-LATE estimator is $\widehat{\rho}$ given in equation (2). This $\widehat{\rho}$ is a consistent estimator of ρ as long as we can apply a weak law of large numbers to the expectations that define the covariances in equation (2). Similarly, root-n asymptotic normality follows mechanically as long as we can apply an appropriate central limit theorem and the delta method. Independent, identically distributed observations and some finite higher moments are sufficient and stronger than necessary conditions for $\widehat{\rho}$ to be a root-n consistent, asymptotically normal estimator of the MR-LATE estimand ρ.

One way to write the estimator $\widehat{\rho}$ is to consider the following moments:

$$E \left(Y_i T_i^b - \alpha^b - \lambda^b T_i^b \right) = 0$$

$$E \left(\left(Y_i T_i^b - \alpha^b - \lambda^b T_i^b \right) Z_i \right) = 0$$

$$E \left(Y_i T_i^a - \alpha^a - \left(\rho + \lambda^b \right) T_i^a \right) = 0$$

$$E \left(\left(Y_i T_i^a - \alpha^a - \left(\rho + \lambda^b \right) T_i^a \right) Z_i \right) = 0$$

for some constants α^a, α^b, λ^b, and ρ. These moments correspond to the two instrumental variables regressions that comprise MR-LATE. One could therefore estimate the constants α^a, α^b, λ^b, and ρ by applying GMM to the above moments (which would actually just reduce to method of moments estimation), and the standard GMM asymptotic distribution formula would then deliver both $\widehat{\rho}$ and correct standard errors for $\widehat{\rho}$. It would also be straightforward to bootstrap the two instrumental variables regressions that comprise $\widehat{\rho}$.

4 MR-LATE with Generally Mismeasured Treatment

Corollary 2 showed that, with missing treatment, that is, if D is unobserved for a subset of the sample, MR-LATE identifies the true LATE that we would estimate if D were not missing. This was Case 2 earlier. Now we consider the more general situation where treatment is mismeasured. The general assumption here comes from Case 3 earlier, namely, we assume we can construct two different flawed treatment indicators T^a and T^b, where T^a is more likely to misclassify a treated person as untreated, while $1 - T^b$ is more likely to misclassify an untreated person as treated. Equivalently,

---

[9]To see this, observe in Case 2 that if $D = 0$ for any individual, then $T^a = 0$ for that individual. This makes $T_0^a = 0$ for all individuals, so p_0^a, which is an expectation of T_0^a, must equal zero. A similar analysis applies to p_1^b.

we are now considering the case where p_0^a and p_1^b are small but not necessarily equal to zero, as opposed to Case 2 where we could construct T^a and T^b to make these probabilities equal zero.

4.1 Set Identification

When MR-LATE does not point identify LATE, it can still sometimes provide bounds on LATE. The following Corollary of Theorem 1 introduces the possibility of using MR-LATE for set identification:

Corollary 3. *Let Assumption 1 hold, and let Assumption 2 hold with $T = T^a$ and with $T = T^b$. If $q^a - q^b > 0$, then ρ and ℓ have the same sign, and so MR-LATE signs LATE. If $q^a - q^b \geq 1$, then ℓ lies between 0 and ρ, and so a sufficient condition for the identified set of ℓ to be bounded is $p_1^a > p_0^a$ and $p_0^b > p_1^b$.*

Corollary 3 states that if, for compliers, the share of actually treated in T^a is larger than the share of misclassified actually untreated, and analogously, if the share of actually untreated in T^b is larger than the share of misclassified actually treated, then MR-LATE is informative regarding the sign and the magnitude of LATE.

Ura (2018) constructs an identified set for LATE when a single mismeasured treatment indicator is observed, under assumptions similar to our Assumptions 1 and 2. So, one could calculate Ura's identified set using each of the two observed mismeasures T^a and $1 - T^b$, construct the identified set given by Corollary 3, and then take the intersection of all three sets. We do not know if the bounds in Corollary 3 are sharp. Ura (2018) provides sharp bounds, but under assumptions that are not exactly the same as ours.

4.2 MR-LATE for Bias Reduction

In some contexts, such as when we have general measurement error in treatment rather than just some missing observations, the conditions under which MR-LATE equals or bounds the true LATE may be difficult to confirm. So, here we consider a more general empirically relevant question: When does MR-LATE provide a better (less asymptotically biased) estimator of LATE than the standard LATE estimator in the presence of treatment measurement error? As we show below, the answer is that MR-LATE usually has less asymptotic bias, often much less, and is more biased only in unusual situations. Later, using Monte Carlo simulations, we show similar results in finite samples.

Formally, MR-LATE provides a good approximation to LATE when p_0^a and p_1^b are close to zero. Having p_0^a near zero means that, among compliers, the probability that $T_0^a = 1$ must be low. Similarly, having p_1^b near zero means that, among compliers, the probability that $T_1^b = 0$ must be low. The probabilities p_0^a and p_1^b will be close to zero, making MR-LATE close to the true LATE, if T^a is rarely one when D is zero, and if T^b is rarely one when D is one.[10]

Suppose we can construct two different flawed treatment indicators T^a and T^b, as described earlier in Case 3. In short, T^a differs from the unknown true D in having too many zeros, while $1 - T^b$ differs from D in having too many ones. What we do in this section is compare the asymptotic bias in MR-LATE to that of the usual LATE estimator with mismeasured treatment.

---

[10]This is roughly analogous to how one might interpret the usual LATE estimator as being close to but not equal to a true causal effect if the probability of defiers in the population is small but nonzero.

Let the B-LATE (for Biased LATE) estimator denote the standard LATE estimator when we observe and use a mismeasured treatment indicator T in place of the true unobserved treatment indicator D. That is,

$$\text{B-LATE} = \varsigma = \frac{cov\,(Y, Z)}{cov\,(T, Z)}, \qquad \widehat{\varsigma} = \frac{\widehat{cov}\,(Y, Z)}{\widehat{cov}\,(T, Z)}.$$

So, B-LATE is the instrumental variables regression of Y on T using Z as the instrument.

To abstract from small sample issues, instead of comparing B-LATE $\widehat{\varsigma}$ and MR-LATE $\widehat{\rho}$ to the infeasible LATE estimator $\widehat{\ell}$, we will now compare the corresponding estimands, that is, we look at the biases $\varsigma - \ell$ and $\rho - \ell$. These are equivalent to the asymptotic biases in the B-LATE and MR-LATE estimators (relative to the infeasible true LATE estimator), under the minimal conditions needed for a law of large numbers to hold for the sample averges that define these estimators.

It follows from the proofs of Theorem 1 and Corollary 1 that

$$\varsigma = \frac{E\,(Y \mid Z = 1) - E\,(Y \mid Z = 0)}{E\,(T \mid Z = 1) - E\,(T \mid Z = 0)} = \frac{E\,((Y_1 - Y_0)\,(D_1 - D_0))}{(p_1 - p_0)\,E\,(D_1 - D_0)} = \frac{1}{p_1 - p_0}\ell \qquad (13)$$

This shows that the bias in B-LATE, defined by $\varsigma - \ell$, is $\frac{1}{p_1 - p_0} - 1$ times the true LATE ℓ. So, the bias in B-LATE is small only if the probability p_1 is much larger than the probability p_0, and B-LATE is *always* asymptotically biased unless the probability of any misclassification is zero.

Since MR-LATE assumes we have two measures of treatment T^a and T^b, let T be any combination of these two, i.e., for any probability r, let each observation of T equal T^a with probability r and T^b with probability $1 - r$.[11] The B-LATE bias then becomes

$$\varsigma - \ell = \left[\frac{1}{(p_1^a - p_0^a)\,r - (p_1^b - p_0^b)\,(1 - r)} - 1 \right] \ell \qquad (14)$$

By definition, the bias in MR-LATE is $\rho - \ell$. From Theorem 1, this bias is $\rho - \ell = \left[(q^a - q^b) - 1\right]\ell$, so

$$\rho - \ell = \left[\frac{p_0^a}{p_1^a - p_0^a} - \frac{p_1^b}{p_1^b - p_0^b} \right] \ell \qquad (15)$$

These equations show that as long as p_1^a is large relative to p_0^a, and p_0^b is large relative to p_1^b, the bias in MR-LATE will be relatively small. In contrast, the bias in B-LATE is small only if the weighted average of $(p_1^a - p_0^a)$ and $(p_0^b - p_1^b)$ happens to be close to one. To compare these biases more formally, the following lemma gives a sufficient condition for MR-LATE to have smaller bias (in absolute value) than B-LATE, i.e., to have $|\rho - \ell| < |\varsigma - \ell|$.

Corollary 4. *Let Assumption 1 hold, and let Assumption 2 hold with $T = T^a$ and with $T = T^b$. Assume $p_1^a > p_0^a$ and $p_0^b > p_1^b$. If*

$$p_0^a + p_1^b < \left(\frac{1}{\max\{p_1^a - p_0^a,\, p_0^b - p_1^b\}} - 1 \right) \min\{p_1^a - p_0^a,\, p_0^b - p_1^b\}, \qquad (16)$$

---

[11]Hence, if $r = 1$ or $r = 0$ then B-LATE corresponds to just doing the standard IV estimation with either T^a or T^b, respectively. Whereas if, e.g., $r = 1/2$, then T is constructed as half T^a and half T^b.

then $|\rho - \ell| < |\varsigma - \ell|$.

To prove this corollary, observe that the expression in the brackets in equation (15) is less than $\left(p_0^a + p_1^b\right)/\min\{p_1^a - p_0^a, p_0^b - p_1^b\}$, while the expression in the brackets in equation (14) is greater than $\left(1/\max\{p_1^a - p_0^a, p_0^b - p_1^b\}\right) - 1$.

Note that Corollary 4 only provides a sufficient, not a necessary, condition for MR-LATE to have smaller bias than B-LATE. There can also be other situations where MR-LATE is less biased. The conditions required by Corollary 4 are inequalities on the probabilities p_1^a, p_0^a, p_0^b, and p_1^b. These are very mild conditions that we expect would generally hold in practice. The first two inequalities, that $p_1^a > p_0^a$ and $p_0^b > p_1^b$, are equivalent to assuming that T^a and T^b are informative about D, saying essentially that T^a and $1 - T^b$ positively correlate with D, as they would generally be constructed to do. The third inequality, equation (16), is more complicated, but essentially says only that the sum of the probabilities $p_0^a + p_1^b$ must be sufficiently small, noting that T^a and T^b are constructed precisely to make these particular probabilities small (in the case where the MR-LATE bias is zero, these probabilities are zero).

To more easily interpret this Corollary, consider the symmetric case where $p_0^a = p_1^b$ and $p_1^a = p_0^b$. In that case, the sufficient conditions in Corollary 4 reduce to just the two inequalities $p_0^a + p_1^a < 1$ and $p_0^b + p_1^b < 1$, which are mild restrictions. To further illustrate the point, and to assess the magnitude of the advantage of MR-LATE over B-LATE, in Table 1 we report the difference in the absolute values of the biases associated with B-LATE (from equation (14)) and MR-LATE (from equation (15)) for $r = 0.5$.[12] We report these bias differences for sensible departures from the point-identifying condition of $p_0^a = p_1^b = 0$. We also consider varying degrees of informativeness of T^a and T^b: the higher is p_1^a and p_0^b, the more informative are T^a and T^b as measures of treatment and control, respectively.

Almost all the entries in Table 1 are positive, showing that MR-LATE is superior to B-LATE in all but rare cases. Moreover, the positive entries in the table are mostly much larger than the negative ones, showing that MR-LATE usually has much less bias than B-LATE, and in the few cases where B-LATE is superior (the last row and column of Panel A), it is not superior by much. Note that, consistent with the Corollary above, these rare cases of B-LATE having lower bias correspond to $p_0^a + p_1^a > 1$ or $p_0^b + p_1^b > 1$.

Tables A6 and A7 in the Appendix show the magnitudes of the respective biases that are differenced in Table 1, which provide a few more insights. First, MR-LATE is unbiased when $p_0^a = p_1^b = 0$, while B-LATE is always biased if any misclassification is present. Second, the larger are p_1^a and p_0^b, the larger is the bias of both B-LATE and MR-LATE. And third, the advantage of MR-LATE over B-LATE is particularly strong when p_1^a is small and p_0^b is large.

We conclude that, in the presence of treatment measurement error, MR-LATE has, except in rare cases, smaller asymptotic bias than the standard LATE estimator, and the bias reduction it provides is usually substantial.

---

[12]We also evaluated biases using other values of r, but as Corollary 4 suggests, the value of r only rarely affects which estimator has smaller bias.

Table 1: B-LATE vs. MR-LATE Bias Comparisons: $|\text{Bias}_{\text{B-LATE}}| - |\text{Bias}_{\text{MR-LATE}}|$

Panel A: $p_1^a = 0.9$, $p_0^b = 0.9$					
$p_1^b \downarrow$ | $p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	0.111	0.106	0.084	0.051	-0.036
0.01	0.106	0.101	0.079	0.047	-0.039
0.05	0.084	0.079	0.059	0.028	-0.054
0.1	0.051	0.047	0.028	0.000	-0.077
0.2	-0.036	-0.039	-0.054	-0.077	-0.143

Panel B: $p_1^a = 0.8$, $p_0^b = 0.8$					
$p_1^b \downarrow$ | $p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	0.250	0.245	0.224	0.190	0.095
0.01	0.245	0.241	0.219	0.187	0.093
0.05	0.224	0.219	0.200	0.170	0.081
0.1	0.190	0.187	0.170	0.143	0.062
0.2	0.095	0.093	0.081	0.062	0.000

Panel C: $p_1^a = 0.7$, $p_0^b = 0.7$					
$p_1^b \downarrow$ | $p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	0.429	0.424	0.405	0.372	0.267
0.01	0.424	0.420	0.401	0.369	0.266
0.05	0.405	0.401	0.385	0.356	0.262
0.1	0.372	0.369	0.356	0.333	0.252
0.2	0.267	0.266	0.262	0.252	0.200

Notes: Results obtained setting $r = 0.5$. Each cell reports $|\text{Bias}_{\text{B-LATE}}| - |\text{Bias}_{\text{MR-LATE}}|$ under different values of $p_1^a, p_0^a, p_1^b, p_0^b$. Cells are empty if the difference is not finite (one of the two biases equals infinity). The true LATE is normalized to 1.

5 Monte Carlo Simulations

Here we provide results from some Monte Carlo experiments to check the finite sample properties of the MR-LATE estimator.

Setup. For our data generating process (DGP), we construct unobserved potential outcomes Y_0 and Y_1 as follows:

$$Y_0 = S + V_0$$
$$Y_1 = 1 + S + V_1,$$

where S, V_0 and V_1 are random unobserved errors. The corresponding observed outcome equals:

$$Y = (1 - D)Y_0 + DY_1$$

We consider an unobserved true treatment indicator D based on a threshold-crossing model.

So, $D = \mathbb{I}(R^* \geq 0)$, where the unobserved index R^* is defined as

$$R^* = 1 + Z + S + U,$$

with Z being the observed binary instrument and U being an additional error. We let $S \sim \mathcal{N}(0,1)$, $Z = \mathbb{I}(\sim \mathcal{U}(0,1) > 0.5)$, $V_0 \sim \mathcal{N}(0,1)$, $V_1 \sim \mathcal{N}(0,1)$, and $U \sim \mathcal{N}(0,1)$. Each sample consists of 5,000 observations, and the DGP is simulated 200 times.

To simulate a treatment variable that is missing (unobserved) not at random for a subset of the population, we generate:

$$T_0^a = \mathbb{I}[\Phi(U_{a,0}) < p_0^a],$$
$$T_1^a = \mathbb{I}[\Phi(U_{a,1}) \geq 1 - p_1^a],$$
$$T_0^b = \mathbb{I}[\Phi(U_{b,0}) \geq 1 - p_0^b],$$
$$T_1^b = \mathbb{I}[\Phi(U_{b,1}) < p_1^b],$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution, $U_{a,0}$ and $U_{a,1}$ ($U_{b,0}$ and $U_{b,1}$) are jointly normal, and p_0^a, p_1^a, p_0^b and p_1^b are the misclassification probabilities defined in the previous section. Based on this, we construct two treatment mismeasures as follows:

$$T^a = DT_1^a + (1 - D)T_0^a,$$
$$T^b = DT_1^b + (1 - D)T_0^b.$$

We also construct treatment mismeasures based on a variable P, which takes values $\{1, 0, -1\}$, with $P = 1$ if $T^a = 1$, $P = -1$ if $T^b = 1$, and $P = 0$ otherwise. So, P is equal to 0 for those whose treatment is potentially misclassified.

LATE with Missing Treatment. In our first experiment, we maintain the sufficient conditions of Corollary 2 ($p_0^a = p_1^b = 0$) and assume that the proxies T^a and T^b have each one type of misclassification error. We fix $p_1^a = 0.6$, $p_0^b = 0.9$. So, this DGP assumes that 60 percent of true treated are observed treated and 90 percent of true control are observed control. Furthermore, we assume the econometrician observes Y, Z, T^a and T^b (or P).

For this first experiment, we compare the performances of five different estimators. The first two estimators are infeasible since they assume D is observed without error (Case 1 in Section 2). First, we estimate an ordinary least squares (OLS) regression model of Y on a constant and D. Due to the correlation between treatment and potential outcomes (the variable S causes a violation of the unconfoundedness assumption), OLS is an inconsistent estimator of the effect of D on Y. Second, we estimate a two-stage least squares (2SLS) regression model of Y on a constant and D, using Z as an instrument for D. The coefficient of D in this 2SLS regression is the standard LATE estimator, which is consistent but infeasible because it relies on D being observed without error.

The remaining estimators we consider are feasible and correspond to Case 2 in Section 2. First, we estimate a two-stage least squares (2SLS) of Y on a constant and T^a, using Z as an instrument

Table 2: Monte Carlo Simulations

	Case 1		Case 2			Case 3		
	OLS	2SLS	2SLS	2SLS-drop	MR-LATE	2SLS	2SLS-drop	MR-LATE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimate	2.205	1.008	1.704	1.220	0.975	1.872	1.329	1.133
SE	0.048	0.230	0.475	0.225	0.301	0.537	0.250	0.349
Bias	1.205	0.008	0.704	0.220	-0.025	0.872	0.329	0.133
MSE	1.455	0.053	0.721	0.099	0.091	1.049	0.170	0.139

Notes: In each simulation, the true value is set to 1.000. Results are based on 200 simulations with 5,000 observations each. In Case 2, the sufficient conditions hold, that is, $p_0^a = p_1^b = 0$. In Case 3, the sufficient conditions do not hold, that is, $p_0^a = p_1^b = 0.05$.

for T^a.[13] This approach is equivalent to a naïve approach that ignores treatment missingness or misclassification and uses T^a as a measure of D. Second, we estimate a linear regression model using 2SLS of Y on T^a with Z as instrument over a subsample that drops observations with $P = 0$ (2SLS-drop). Third, we construct $T^a = \mathbb{I}(P = 1)$ and $T^b = \mathbb{I}(P = -1)$, and apply the MR-LATE estimator over the full sample. MR-LATE is given by $\widehat{\lambda}^a - \widehat{\lambda}^b$, where, for $j = a, b$, $\widehat{\lambda}^j$ is the 2SLS coefficient of T^j, obtained by regressing YT^j on a constant and T^j, using Z as an instrument for T^j.

The first five columns of Table 2 show results of our first set of simulations. As expected, the OLS estimator is biased due to the correlation between treatment and potential outcomes (Column (1)) and the standard 2SLS estimator, which is infeasible because it uses the true unobserved D, has virtually no bias (Column (2)). Column (2) is the benchmark we wish to compare the feasible estimators to. A 2SLS estimator that uses T^a in place of D results in substantially biased estimates of the LATE or D on Y (70 percent bias) and a large mean square error (Column (3)). This estimator is feasible but inconsistent due to the misclassification error in T^a. Dropping observations with $P = 0$ and using a 2SLS estimator on the remaining subsample yields estimates of the true LATE that are biased by about 22 percent (Column (4)). In line with Corollary 2, our MR-LATE estimator, which is both feasible and consistent, has a near-zero bias (Column (5)).

LATE with Generally Mismeasured Treatment. In our second experiment (which corresponds to Case 3 discussed in Section 2), we relax the sufficient conditions for point-identification of Corollary 2. Specifically, we set $p_0^a = p_1^b = 0.05$. As before, we assume that the proxies T^a and T^b have each one type of misclassification error (with that $p_1^a = 0.6$, $p_0^b = 0.9$) and that the econometrician just observes Y, Z, T^a and T^b (or P). We compare the performance of three estimators: a feasible 2SLS of Y on T^a that uses Z as an instrument for T^a and ignores misclassification errors, a feasible 2SLS of Y on T^a with Z as instrument on the subsample of observations where T^a equals $1 - T^b$ (which is equivalent to dropping observations where $P = 0$), and the MR-LATE estimator, which uses both measures T^a and T^b and considers the full sample. Columns (6)-(8) of Table 2 show the results of the Monte Carlo simulations. As shown in Column (7), dropping potentially mismeasured observations leads to an even larger bias (33 percent) relative to Column (4). Although the MR-LATE estimator is biased, this bias is much smaller. These findings confirm our analysis in Sec-

---

[13]Using T^b in place of T^a would deliver equivalent results.

tion 4.2 and highlight the potential use of MR-LATE as a bias-reduction tool when estimating LATE with a generally mismeasured treatment.

6   Examples of Missing and Mismeasured Treatment

In this section, we discuss the potential usefulness of MR-LATE, by summarizing a variety of economic data sets and empirical applications that suffer from missing or mismeasured treatment. As these examples show, missing or mismeasured binary endogenous variables is a pervasive problem in economics (Bound et al., 2001). This list is far from exhaustive, but it helps clarify the extent of the problem in applied economic research.

As discussed in Molinari (2010), various socio-economic surveys of the US population suffer from severe item nonresponse rates. For example, the National Longitudinal Survey of Youth asks female respondents about their drinking consumption during pregnancy. In the 1984 wave, the nonresponse rate for these questions ranged from 6 and 14 percent, leading to a missing treatment problem when studying, e.g., the effect of drinking during pregnancy on birth outcomes. The Survey of Consumer Finances includes questions regarding ownership of stocks, bonds, businesses, and bank accounts, with nonresponse rates ranging between 6 to 20 percent in the 1995 wave. The 1992 Health and Retirement Study (HRS92) and the 1993 Assets and Health Dynamics Among the Oldest Old display nonresponse rates for similar questions up to 33 and 45 percent, respectively. The HRS92 nonresponse rate for questions about the respondent's children's income was 14 percent. Such data distortions may jeopardize any causal analysis of the effect of, e.g., parental wealth on children's outcomes or of children's economic status on the likelihood that they receive transfers from their parents.

Previous research also suggests that survey-gathered divorce data can be highly inaccurate. Depending on the study, scholars have found that the survey estimates of divorce are between 8 and 25 percent less than the official figures from the Vital Statistics (Mitchell, 2010). O'Connell (2006) provides an extensive commentary on the magnitude of missing information on marital histories in the Survey of Income and Program Participation and shows that up to 20 percent of respondents reported only part or none of the required information. Zhang et al. (2016) observe a similar problem for weight and height-related questions in the Consortium on Safe Labor Survey. Meyer et al. (2015) analyze the item nonresponse rates for questions related to the receipt of government transfers in a few prominent US household surveys. They document a high likelihood of missing treatment when using the Current Population Survey (which has nonresponse rates ranging between 16 and 20 percent, depending on the wave of the survey) or the National Health Interview Survey (with nonresponse rates up to 24 percent). Linking survey data with administrative data on SNAP (the Supplemental Nutrition Assistance Program or Food Stamp Program), Meyer et al. (2018) also show that 23 percent of true food stamp recipients do not report receipt in the Survey of Income and Program Participation (SIPP), 35 percent in the American Community Survey, and 50 percent in the Current Population Survey. A significant fraction of true non-recipients are also recorded as recipients, especially in the SIPP. The empirical analysis of, e.g., the effect of divorce on labor market outcomes or the impact of maternal weight on children's health or the effectiveness of government

programs based on such surveys would inevitably rely on strong ignorability assumptions and possibly lead to incorrect conclusions.[14]

When collecting information on issues that are private, perceived as potentially shameful, or related to illegal or criminal activities, the problems of item nonresponse and misreporting may be particularly severe. So, for example, estimates of the effect of domestic violence on women's outcomes and children's development are likely biased due to selection into reporting or lack of trust to report truthfully (Alderman et al., 2013). Agüero and Frisancho (2020) compare reporting of domestic violence by women in Peru when using direct questions as applied by the Demographic and Health Surveys (a global data collection effort comprising 122 surveys in 61 developing countries) and indirect questions that provide further anonymity to the respondent. They find evidence of measurement error in responses to direct questions as up to 30 percent of women tend to underreport physical and sexual violence by the husbands. Based on data from 24 developing countries, Palermo et al. (2014) also show that only seven percent of women who report being victims of violence made a report that can be captured by administrative records, suggesting that in many contexts both survey and administrative data may be incomplete.

Finally, binary treatment that is defined as a function of multiple valued or continuous data will suffer from mismeasurement of that underlying data. Previous work has applied Benford's law (on the expected frequencies of the first digits of numbers found in many real-world data), to detect misreporting of variables in surveys and other economic data. For example, Carslaw (1988) and Durtschi et al. (2004) use this approach to detect fraud and misreporting of accounting data. Nye and Moul (2007) show that World Bank GDP figures from the developing world often do not conform well to Benford's law. Based on nine household surveys across the globe used extensively by research economists, Judge and Schechter (2009) show that crop data are frequently of low quality; they find this is especially true for the Progresa data from Mexico. Finally, in the Indian context, Calvi and Keskar (2021b) document the existence of some misreporting of marital transfers in the 1999 Rural Economic and Demographic Survey.[15] In all these instances, the estimation of the effect of economic shocks (defined, e.g., as profits, GDP, crop yields, or marital transfers above or below a certain threshold) may be problematic.

Our MR-LATE estimator could be applied in many such applications as the primary identification strategy or as a robustness check. Alternatively, as we illustrate below, MR-LATE can be used in situations where treatment is not observed but can be estimated. In these situations, treatment mismeasurement may arise from model misspecification and estimation errors.

---

[14]Other examples in the literature include the misclassification of union status (Card, 1996), participation to trainings (Barron et al., 1997), coverage of health insurance (Black et al., 2000), language fluency (Dustmann and Soest, 2001), self-evaluation of health-related status (Crossley and Kennedy, 2002), educational attainment (Black et al., 2003), chemical emissions by firms (Marchi and Hamilton, 2006), disability status (Kreider and Pepper, 2007), types of corporate governance structure (Almeida et al., 2010), school meals (Gundersen et al., 2012), dental insurance (Kreider et al., 2015), firm's formality status (Gandelman and Rasteletti, 2017), and technology adoption (Wossen et al., 2018).

[15]Others who have used Benford's law to check the validity of data in the social sciences include Varian (1972), Carslaw (1988), Nigrini (1996), Durtschi, Hillison, and Pacini (2004), Geyer and Williamson (2004), de Marchi and Hamilton (2006), Giles (2007), and Nye and Moul (2007).

7   Women's Empowerment and Family Health in India

We apply the MR-LATE estimator to study the impact of women's control over household resources on family members' health outcomes in India. We define our unobserved true treatment indicator D to equal one if the wife has primary control of resource allocation decisions in the household, and zero otherwise. Formally, for each household i, we define $D_i = \mathbb{I}\left(R_i^* \geq e_i\right)$, where R_i^* is the share of resources of household i that are under the wife's control, and e_i is an unobserved threshold that may vary across households. Here R_i^* and e_i are both percents, and so could range from 0 to 100. So, if we had $e_i = 60$ for a given a household i, then D_i would equal 1 if $R_i^* \geq 60$. Each household may have its own unobserved threshold e_i.

A key difficulty in observing or calculating R_i^* is that survey data typically do not record individual-level expenditures nor how resources are allocated within families. Control over household resources is also hard to observe as most goods in a household can be shared or consumed jointly to some extent by household members. For example, home heating is almost completely shared, while cooking fuel is jointly consumed just among household members who are eating together. Other goods, like food, are consumed individually, but it is difficult to track exactly who eats what within the household. To overcome this limitation, we use a structural collective household model to construct an estimate R_i of the women's actual, unobserved resource control R_i^* for each household i.[16] We then apply MR-LATE to account for both the unobserved heterogeneous threshold e_i and for possible model misspecification and estimation errors in R_i, and hence in D_i.

Using MR-LATE, we examine the impact of treatment D on a variety of health related outcomes Y. For adults, we consider body mass index (hereafter BMI), defined as weight in kilograms divided by height in meters squared, and indicators for being underweight or anemic. For children, we consider height-for-age and weight-for-age z-scores, which are typically used as indicators for stunting and undernutrition in children, recent occurrences of diarrhea, fever and cough, and an indicator for whether a child has been vaccinated against one or more diseases for children. Our instrument Z is based on inheritance law reforms that equalized women's inheritance rights to men's in several Indian states between 1976 and 2005.[17]

Our analysis assumes that $D_i = \mathbb{I}\left(R_i^* \geq e_i\right)$, rather than some other function of R_i^*, is the relevant measure of wives' bargaining power for decisions involving health outcomes. This is consistent with other models of intra-household power that make use of cutoffs (e.g., models that use the threat of divorce as bargaining leverage are based on whether women's resource shares are above or below a cutoff that is determined by the resources they would have access to after divorce). More generally, to the extent that health outcomes are the results of many health decisions, it is reasonable to assume that a family member with sufficiently high control over resource allocations will determine many of these decisions. This is roughly analogous to voter models, where outcomes are

---

[16]There exists a variety of indicators of women's status and control over resources that might be used to measure treatment, including self-reports of decision making power. However, these measures are quite crude, usually focusing on just a few specific decisions. Nevertheless, in Section B in the Appendix , we compare our model estimates to these measures.

[17]So, our empirical analysis involves two models: one for determining treatment D, and one for how treatment D causally affects the outcome Y. We use a structural model to estimate D (women's control of resources) because we have a great deal of economic theory and data to guide us in the construction of that model, and we have our MR-LATE estimator that can compensate for estimation and specification errors in that structural model. We use a LATE framework to estimate the causal effect of D on Y (health outcomes) because we have far less structural theory to guide us in modeling that connection.

primarily determined by the party with the most power. For example, a woman who has primary control over the household's resources may be able to make timely decisions to treat herself or a sick child after discovering an illness, or to more easily make use of health services that must be paid for and follow through with treatment recommendations. It is important to note, however, that our application does not imply a discontinuity in observed behavior, because the threshold e is random. It also does not imply a winner-take-all system: D will yield a nonzero average treatment effect if it just affects the relative weight of the wife's preferences in household health decisions.

Despite the behavioral arguments above, it is possible that the magnitude of R^* itself, not just D, is more relevant for determining Y. To test this alternative possibility we estimated a model that regresses Y on R (our estimate of R^*) and other covariates, and found limited evidence of a direct continuous effect of women's resource shares R on health outcomes Y.[18]

7.1 Modeling Women's Control over Household Resources

We apply the structural methodology developed by Dunbar et al. (2013) (hereafter DLP) to obtain R, an estimate of R^*, which we then use to construct the treatment indicators T^a and T^b required for the application of MR-LATE.

Define a good to be *private* if it is not shared or consumed jointly. Define a good to be *assignable* if it appears in just one (known) household member's utility function, and so is only consumed by that household member. If all goods were private and assignable, then we could potentially directly observe R^*. The difficulty is that most goods are not assignable or private. Assume we observe the household's total expenditures on all goods, M, and we observe the household's expenditures on (at least) one private assignable good for each decision maker in the household. Let scalars w_w and w_m denote the household's budget shares (fraction of total expenditures M) spent on the observed private assignable goods, which in our data are women's clothes and men's clothes, respectively.

We cannot just use w_w and w_m as measures of R^* and $1 - R^*$, because men and women may have very different tastes for clothing. For example, a wife might control fewer household resources than her husband, but still consume more clothes than him, because she derives more utility from clothing consumption than her husband does. Following DLP, we instead identify and estimate a separate clothing Engel curve for each decision maker. Then, we implicitly invert these Engel curves to solve for R^*. Details on this intra-household model and on the derivation of these Engel curves are provided in Section B.1 in the Appendix.

Let $X = (X_1, ..., X_K)$ denote a vector of observable attributes of households and their members. Household attributes X may affect the preferences of each household member and may also affect the household's bargaining process or social welfare function, and as a result may directly affect resource shares.[19] We employ the commonly used Piglog (price independent generalized logarithmic)

---

[18]We focused on linear regression for these tests. One could also consider flexible non-linear but still continuous models relating Y to R (with or without using Z as an instrument). However, such models will generally be biased due to the errors in R (the same kind of errors that will make MR-LATE useful), particularly since there is no reason to believe the errors in R would satisfy classical measurement error. We do not report these results for brevity, but they are available on request.

[19]In the collective household model literature, covariates that only affect the household's bargaining process but not the tastes of the household members are known as *distribution factors*. A feature of the DLP approach is that it does not require observation of distribution factors. However, if any of our covariates are distribution factors, then they would affect R^* but not the other parameters.

functional form for these Engel curves, which is

$$\begin{cases} w_w & = R^*\delta_w + R^*\beta \ln(R^*M) \\ w_m & = (1 - R^*)\delta_m + (1 - R^*)\beta \ln((1 - R^*)M) \end{cases} \tag{17}$$

where $\beta = \beta(X)$, $\delta_w = \delta_w(X)$, $\delta_m = \delta_m(X)$, and $R^* = R^*(X)$.[20] Note that the demand functions for other goods (those that are not private and assignable) may be more complicated, but are not needed for estimation of resource shares. DLP prove that the functions $\beta(X)$, $\delta_w(X)$, $\delta_m(X)$, and $R^*(X)$ are identified in this model. Identification relies partly on the assumption that $\beta(X)$ is the same for men and women. DLP call this the SAP (similar across people) assumption, and provide empirical evidence supporting this restriction.

For our empirical application, we assume the functions $\beta(X)$, $\delta_w(X)$, $\delta_m(X)$, and $R^*(X)$, are all linear in their arguments. In particular, we specify

$$R^*(X) = \theta_0 + \theta_1 X_1 + \ldots + \theta_K X_K. \tag{18}$$

7.2   Estimation

We employ two different datasets from India. One, the 62nd round of the NSS Consumer Expenditure Survey (hereafter NSS), contains detailed consumption data that we use for estimating the above model of resource shares. The other, the 3rd round of the National Family Health Survey (hereafter NFHS), collects health outcomes of women, men, and children, as well as information about year of marriage which is critical for the construction of our instrument Z. Both surveys were conducted between 2005 and 2006, and the covariate vector X (of attributes of households and their members) is observed in both datasets. Specific details of the surveys and descriptive statistics for the two samples are discussed in Section B.2 in the Appendix.

We append an error term to the equations in system (17), yielding a two equation system that we estimate by non-linear Seemingly Unrelated Regressions (SUR) using the NSS data. The non-linear SUR is iterated until the estimated parameters and the covariance matrix converge, which is asymptotically equivalent to maximum likelihood with multivariate normal errors. Our estimation sample includes 7,703 nuclear families with children, i.e., households consisting of a mother, a father, and their children.[21]

Let $\widehat{\theta}$ denote the estimate of the vector of coefficients θ in equation (18). Then, for each individual i drawn from the NFHS data, we use these estimates to predict the share of resources controlled

---

[20]We do not include Z as an element of X for two reasons. First, doing so could induce spurious correlation between the estimated treatment indicators and the instrument. Second, the NSS expenditure data does not include information on women's year of marriage, which is required to construct an exact measure of exposure to the inheritance law reforms and hence Z in the NSS dataset. However, we acknowledge that this might lead to a violation of Assumption 2. We therefore repeat our analysis by including a measure of women's *eligibility* to the amendments, defined as the interaction between an indicator variable for being Hindu, Buddhist, Sikh or Jain, and an indicator variable equal to one if a woman was 14 or younger at the time of the amendment in her state and to zero if she was 23 or older (see Heath and Tan (2019) and Calvi (2020)). Results are confirmed and available upon request.

[21]We exclude households in the top or bottom 1 percent of expenditure, and we exclude households that report having performed any ceremony during the month prior to the survey, as unusual purchases of clothing items and non-standard expenditure patterns may occur for festivities and ceremonies. Additional details about the estimation sample are discussed in Section B.2 in the Appendix.

by the woman in individual i's household as

$$R_i = \widehat{\theta}_0 + \widehat{\theta}_1 X_{1i} + \ldots + \widehat{\theta}_K X_{Ki}.$$

Using this estimated R_i, our goal is to estimate a LATE of $D_i = \mathbb{I}(R_i^* \geq e_i)$ on a range of health outcomes. We separately consider health outcomes for mothers, fathers, and children. So, e.g., when i is a child and Y_i is a child's height-for-age z-score (a popular indicator for stunting), the treatment effect we wish to estimate is the change in i's measure of stunting if he/she is exposed to highly empowered mothers. We wish to estimate this treatment effect, even though the mother's true resource share R_i^* (and hence D) is unobserved.

We apply our MR-LATE estimator by constructing two mismeasures of treatment, i.e., $T_i^a = \mathbb{I}(R_i \geq \kappa^a)$ and $T_i^b = \mathbb{I}(R_i < \kappa^b)$, where κ^a and κ^b are chosen constant bounds. For $j = a, b$, the estimation procedure consists of regressing $Y_i T_i^j$ on a constant, T_i^j, and X_i using 2SLS (with Z_i being the excluded instrument). Based on the MR-LATE approach, estimates are then obtained as the difference between the estimated coefficients of treatment in these two 2SLS regressions, that is $\widehat{\rho} = \widehat{\lambda}^a - \widehat{\lambda}^b$.

We provide estimates using four different choices of the bounds κ^a and κ^b, as follows. Recall that R^* and R are percentages, ranging from 0 to 100. To construct each pair of bounds κ^a and κ^b we first choose a percentage $\mathcal{K}$ (either 0, 1, 5, or 10). Then, for each $\mathcal{K}$ we let κ^a be the value that makes $\mathcal{K}/2$ percent of the sample have R in the interval $[50, \kappa^a]$ and κ^b is the value that makes $\mathcal{K}/2$ percent of the sample have R in the interval $[\kappa^b, 50]$. By this construction, MR-LATE will be an asymptotically unbiased estimator of the true LATE if all households i that have $R_i \geq \kappa^a$ also have $R_i^* \geq e_i$, and all that have $R \leq \kappa^b$ also have $R_i^* \leq e_i$. So, the larger is $\mathcal{K}$, the further κ^a and κ^b are from 50, and hence the further $R_i - 50$ can be from $R_i^* - e_i$ without biasing MR-LATE. More generally, the larger is $\mathcal{K}$, the fewer households there will be that violate these inequality conditions, and hence the lower will be the asymptotic bias in MR-LATE.

Choice of $\mathcal{K}$ is therefore essentially a mean squared error tradeoff: larger $\mathcal{K}$ means less asymptotic bias, but noisier (higher variance) estimates, due to having fewer informative observations in T^a and T^b. The extreme case of $\mathcal{K} = 0$ makes MR-LATE equal a B-LATE alternative. Specifically, with $\mathcal{K} = 0$, MR-LATE is numerically identical to the standard Imbens and Angrist (1994) 2SLS LATE estimator, using the mismeasured $T_i = \mathbb{I}(R_i \geq 50)$ in place of the unobserved true $D_i = \mathbb{I}(R_i^* \geq e_i)$. More generally, based on Corollary 4, MR-LATE should have less bias than B-LATE for a wide range of possible misclassification rates. In both our Monte Carlo and empirical estimates, we find that our estimates vary little across middling choices of $\mathcal{K}$.

The Hindu Succession Act and its Amendments. To construct a plausibly unconfounded instrumental variable Z, we exploit changes in the Indian inheritance law. A woman's right to inherit land and other property is often claimed to play a significant role in determining women's power within the household (World Bank, 2014). Inheritance rights in India differ by religion and, for most of the population, are governed by the Hindu Succession Act (HSA). The HSA was first introduced in 1956 and only applied to Hindus, Buddhists, Sikhs, and Jains, in all states other than Jammu and

Kashmir. Before then, the traditional systems (Mitakhshara and Dayabhaga) were strongly biased in favor of sons (Agarwal, 1995). Gender inequalities, however, remained even after the introduction of the HSA. On one hand, in the case of a Hindu male dying intestate (without leaving a will) all his separate or self-acquired property devolved equally upon sons, daughters, widow, and mother. On the other hand, the deceased's daughters had no direct inheritance rights to joint family property, whereas sons were given direct right by birth to belong to the coparcenary. In the decades following the introduction of the HSA, state governments passed amendments that equalized inheritance rights for daughters and sons (Kerala in 1976, Andhra Pradesh in 1986, Tamil Nadu in 1989, and Maharashtra and Karnataka in 1994). A national-level ratification of the amendments occurred in 2005. However, these amendments only applied to Hindu, Buddhist, Sikh or Jain women who were not yet married at the time of the amendment in their state.[22]

For each individual in our NFHS sample, we construct our instrumental variable Z as the indicator of whether the inheritance law reform applied to the woman in that individual's household. Whether Z equals one or not depends on the woman's religion, state of residence, and year of marriage, since exposure to the reform varies by these characteristics.[23]

All specifications presented in our analysis include woman's cohort, religion, state, cohort-religion and state-religion fixed effects, together with state specific time trends up to degree four. The exclusion restriction needed for identification is that, once these fixed effects and time trends are included, being Hindu, Buddhist, Jain or Sikh and unmarried at the time of implementation has an effect on health outcomes only through control of household resources. Although other factors and policies may have differentially affected young Indian women, we do not expect these to vary by religion. Importantly, the existence of other changes resulting *exclusively* from women's increased power within their marital families due to their HSA exposure would not invalidate our instrument. In Section B.4 of the Appendix, we explore this issue further and investigate the effects of the HSA on marital sorting, household wealth, and patterns of consumption. Taken together, this analysis does not uncover any evident violation of the exclusion restriction in our context. Nevertheless, as typical in instrumental variable estimation, we acknowledge that the validity of this assumption cannot be directly tested and violations are possible.

7.3 Empirical Results

In this section, we summarize our estimates of the resource share R, associated treatment measures, and the results of our causal analysis of the effect of women's empowerment D on health outcomes Y. Estimates of the Engel curves of women's and men's private assignable clothing, used to construct R, are reported in Table A10 in the Appendix.

Figure I plots the empirical distributions of the predicted resource shares for women R in the NSS and NFHS samples. The average of R in the two samples equals 46.26 and 46.44, with standard

---

[22]Previous works have evaluated the HSA amendments using difference-in-difference methods (see, e.g., Roy (2008, 2015), Deininger et al. (2013), Heath and Tan (2019), Calvi (2020)). This type of analysis considers exposure to the HSA amendments as treatment. Our goal is not to estimate the treatment effect of this particular policy, but more broadly to estimate the health effect of living in a household where a woman controls a substantial fraction of resources, using exposure to these inheritance rights reforms as an instrument.

[23]Z equals one for 18 percent of women in the sample. Due to the gender age gap at marriage (on average 5 years), the percentage of men married to HSA exposed women is larger (28 percent).

Figure I: Estimated Resource Shares for Women (R)

(A) NSS Sample   (B) NFHS Sample
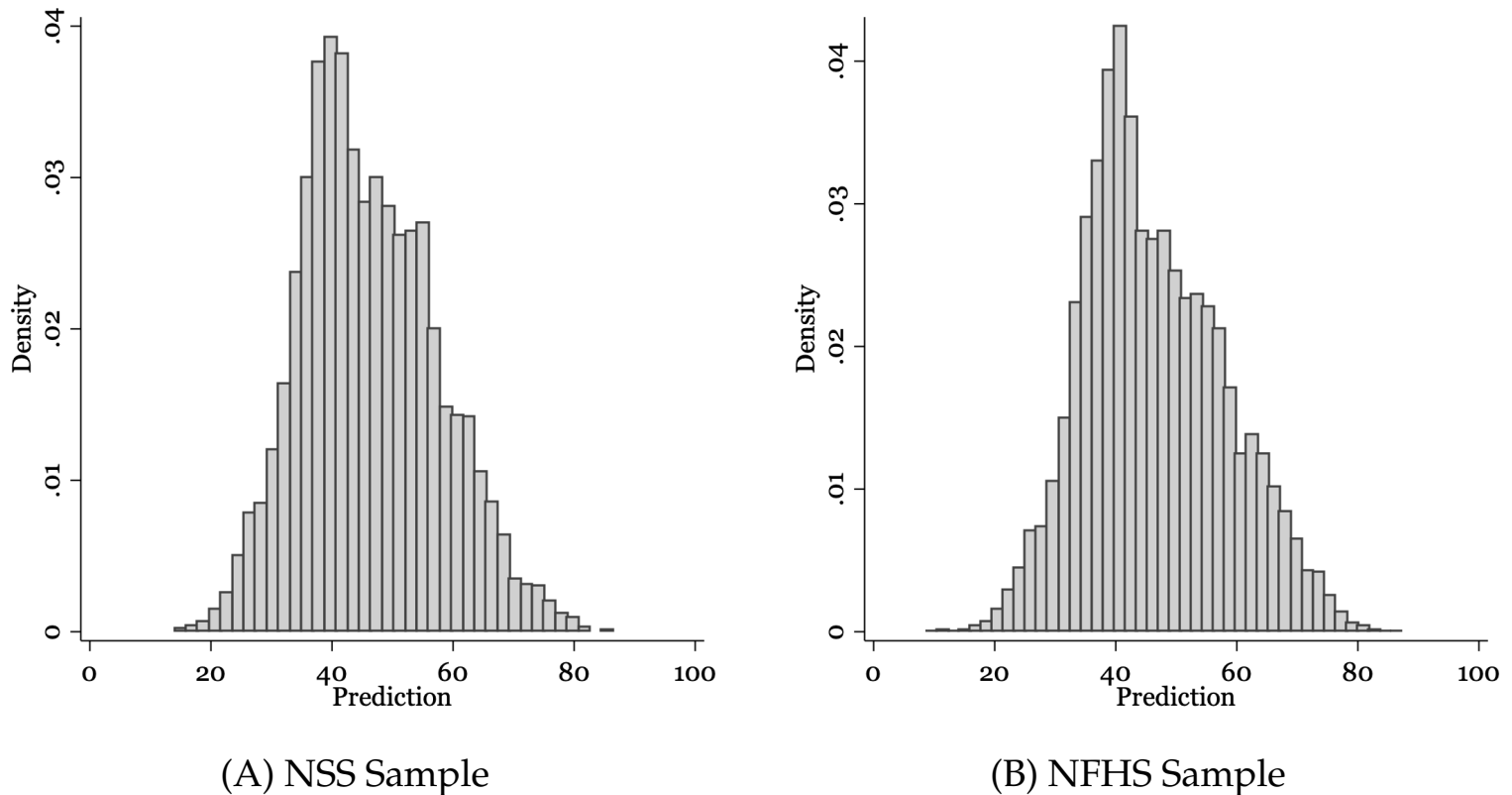
deviations of 11.28 and 11.61, respectively. This means women are estimated to control approximately 46 percent of the household's resources on average, but with considerable variation across households. The summary statistics vary somewhat across the two samples, because they entail averages over the empirical distributions of the covariates ($X_1, ..., X_K$) in each sample. It is therefore reassuring that the estimated distributions of R in the two samples are similar, indicating that the samples are comparable. It is also reassuring that the minima and maxima of the estimated resource shares do not fall outside the zero to 100 percent range for all households, despite them being modeled as linear (and hence not bounded) functions of household characteristics X. Finally, the estimates accord with ex ante expectations. For example, the average R is particularly high in the North-East states (62 percent), which is consistent with the presence of a number of matrilineal societies and cultures in these regions, such as the Khasi and Garo societies. In contrast, North Indian women have much lower control over resources (38 percent on average). Finally, highly educated women (who have completed high school) are found to have a substantially higher command over resources (54 percent) than low educated women (45 percent).

MR-LATE does not require observation or estimation of D; it instead requires that we construct mismeasures T^a and T^b. However, as discussed in Section 2, the obvious alternative would be to apply the usual Imbens and Angrist (1994) LATE estimator (corresponding to $cov(Y, Z) / cov(D, Z)$), by replacing the the true unknown D with a best guess T. A reasonable estimate of D for this purpose would be T defined by $T = \mathbb{I}(R \geq 50)$. This T replaces the unknown true women's resource share R^* with our structural estimate R, and replaces the unknown household specific power threshold e_i with a simple 50 percent. We construct this T, and find that, in both samples, about 35 percent of families have $T = 1$. In the NFHS sample, women who have $T = 1$ have an average R of 59.27, while those having $T = 0$ have an average R of 39.29. So, while we cannot know the average fraction of resources controlled by the truly treated and untreated, i.e., $E(R^*|D)$, our estimates of $E(R|T)$ indicate that the treated group controls a considerably larger fraction of household resources than

Table 3: Adults' Health: MR-LATE Estimates

	Women			Men		
	BMI	Pr(BMI≤18.5)	Pr(Anemic)	BMI	Pr(BMI≤18.5)	Pr(Anemic)
$\mathcal{K} = 0$	8.7906	-0.8113	-0.4842	1.4278	-0.1559	0.0253
	(2.1282)	(0.2080)	(0.1966)	(2.3739)	(0.2981)	(0.1977)
$\mathcal{K} = 1$	8.1164	-0.7859	-0.4678	1.0064	-0.1124	0.0227
	(2.2227)	(0.2168)	(0.1936)	(2.4839)	(0.3046)	(0.2012)
$\mathcal{K} = 5$	6.1220	-0.5945	-0.3816	0.0960	-0.0605	-0.0202
	(3.6769)	(0.2154)	(0.1733)	(4.2341)	(0.2430)	(0.1699)
$\mathcal{K} = 10$	5.0136	-0.4575	-0.4238	-0.7469	-0.1542	0.0131
	(7.9989)	(0.2557)	(0.2055)	(7.7900)	(0.2509)	(0.1899)

Notes: Estimates are obtained using the NFHS-3 data and the MR-LATE estimator. The women sample includes married women of age 15 to 49 in nuclear households. The men sample includes married men of age 15 to 54 in nuclear households. All specifications include an indicator variables for being Hindu, Buddhist, Sikh or Jain, for region of residency, for number of children, rural areas, for being part of Scheduled Castes, Scheduled Tribes or Other Backward Classes, land ownership, woman's and man's high school completion, the fraction of female children, woman's and man's ages and average age of children 0-14. All specifications include state-religion and cohort-religion fixed effects, and state specific time trends (up to degree four). Anemia includes severe and moderate anemia. Bootstrap standard errors in parentheses.

the control group.[24,25]

Recall that to implement MR-LATE in this application, we construct bounds κ^a and κ^b based on choosing a misclassification percentage $\mathcal{K}$ as described in Section 7.2. As noted there, the percentages $\mathcal{K}$ we consider are $\mathcal{K}$ equal to 0, 1, 5, and 10. Table A8 in the Appendix reports the bounds κ^a and κ^b that correspond to each of these values of $\mathcal{K}$.[26] Recall also that $\mathcal{K} = 0$ makes MR-LATE equal the usual LATE estimator using $T = \mathbb{I}(R \geq 50)$ in place of the true D. Table 3 reports the resulting MR-LATE estimates for adult health outcomes, while estimates for children's health outcomes are reported in Table 4. Bootstrapped standard errors are reported in parentheses. In a few instances, even quite small deviations of $\mathcal{K}$ from zero substantially change the MR-LATE estimates for some outcomes, showing that accounting for misclassification appears to be empirically important.

Overall, our MR-LATE estimates indicate that a woman's control of household resources exerts a positive and significant effect on her own health. Women with high control over household resources have a much higher BMI and face a lower likelihood to be underweight or anemic. The estimated effects are sizable: our most conservative estimates indicate that, for compliers, the average treatment effect on women's body mass index is 5.01 and that women in treated households are 45.75 percentage points less likely to be underweight and 38.16 percentage points less likely to be anemic.

A mother's control over household resources positively affects her children's health, too. A highly empowered mother significantly decreases her children's likelihood of being sick with cough,

---

[24]Note that our estimate R of the true R^* refers to resources controlled by the woman, not necessarily those consumed by the woman. For example, mothers and fathers may value differently the well-being of their children, and so they might allocate a different fraction of the resources they control to children.

[25]If our structural estimates of resource control do in fact provide meaningful measures of decision making power in the household, then we would expect them to positively correlate with the household's own reports of who makes decisions. In Section B of the Appendix, we show that this is indeed the case; there is a clear positive relationship between our estimated R and survey reported measures of decision making power within the household. The same holds for T: e.g., we find that women are more likely to report participating in household decisions in households that have $T = 1$ vs. $T = 0$, even after conditioning on individual and household level controls, fixed effects, and state time trends.

[26]The values of κ^a and κ^b vary across the subsamples of women, men and children in the NFHS due to variation in the distribution of covariates. Table A9 in the Appendix shows results of the first stage of the MR-LATE estimates for the different values of κ^a and κ^b considered above, together with the corresponding F-statistics.

Table 4: Children's Health: MR-LATE Estimates

	Weight-for-age (z-score)	Height-for-age (z-score)	Pr(Cough)	Pr(Fever)	Pr(Diarrhea)	Pr(Any Vaccination)
$\mathcal{K} = 0$	1.1075	1.4914	-0.3370	-0.3284	-0.1102	-0.1191
	(0.7100)	(1.1317)	(0.2626)	(0.2334)	(0.1609)	(0.1854)
$\mathcal{K} = 1$	1.0141	1.4602	-0.4086	-0.3759	-0.1560	-0.1472
	(0.7190)	(1.1376)	(0.2642)	(0.2382)	(0.1649)	(0.2117)
$\mathcal{K} = 5$	0.8408	1.6288	-0.3979	-0.3460	-0.2359	-0.1802
	(0.7439)	(1.1209)	(0.2369)	(0.2148)	(0.1612)	(0.2719)
$\mathcal{K} = 10$	0.6933	2.3322	-0.6014	-0.4464	-0.4304	-0.2629
	(1.1050)	(1.3665)	(0.2853)	(0.2691)	(0.2036)	(0.4312)

Notes: Estimates are obtained using the NFHS-3 data and the MR-LATE estimator. The sample includes children 0 to 5 in nuclear households. All specifications include an indicator variables for being Hindu, Buddhist, Sikh or Jain, for region of residency, for number of children, rural areas, for being part of Scheduled Castes, Scheduled Tribes or Other Backward Classes, land ownership, parents' high school completion, the fraction of female children, parents' ages, the child's age and gender. All specifications include state-religion and cohort-religion fixed effects for the mother, and state specific time trends (up to degree four). Bootstrap standard errors in parentheses.

fever or diarrhea in the two weeks prior to the survey. The magnitude of the estimated effects and their statistical significance are affected by the degree of measurement error accounted for. A highly empowered mother also boosts her children's height-for-age and weight-for-age, though these effects are not significantly different from zero. By contrast, we do not find any positive (or negative) effect of a wife's control of resources on her husband's health.

Some of our estimated treatment effects are quite large.[27] To see why, note that outcomes for the treated average over all compliers who have $R_i^* \geq e_i$, while for the untreated outcomes average over compliers who have $R_i^* < e_i$. As noted above, if we (just for this calculation) ignore the measurement error in R_i and let e_i equal 50, then we can calculate that the treated group has an average R_i of 60 percent, while the control has group has an average R_i of just 40 percent. So, the average woman in the treated group has far more control over household resources than the average woman in the control group, which is important to recognize for interpreting the magnitudes of our estimates.

8 Conclusion

We propose a novel estimator to consistently estimate LATE when the treatment indicator is missing not at random for a subset of the population. We also show that when treatment is observed, but with error, our estimator generally has a lower bias than the standard LATE estimator.

A useful direction for future research would be extending MR-LATE to allow for treatments and for instruments that take more than two values. With multiple instruments and binary treatment, one might average MR-LATEs based on each instrument, or impose homogeneity on misclassification probabilities. It should also be possible to use methods similar to how Angrist and Imbens (1995) extend ordinary LATE, by having instruments take more values than treatments. See Tom-

---

[27]When Y is binary, MR-LATE, like standard LATE, corresponds to a linear probability model, and so can yield fitted probabilities that are outside the zero to one range. In this application, both standard LATE (i.e., B-LATE in this case) and MR-LATE yield predicted probabilities that do occasionally lie outside the unit interval. Despite supporting evidence provided in Appendix B.4, we acknowledge that these magnitudes may also be driven by potential violations of the exclusion restriction.

masi and Zhang (2020b) for some results along these lines.

Returning to our base case of binary treatment and binary instrument, MR-LATE can be applied to a variety of contexts when treatment is either missing not at random for some observations, or generally mismeasured due, e.g., to item nonresponse, misreporting, or data contamination. In addition, our method can be useful when treatment is not observed but can be estimated. In such situations, misclassification error may arise from misspecification in the model used to estimate treatment as well as from estimation error. To illustrate this last point, we apply MR-LATE to study the effects of intra-household women's empowerment on the health status of family members in India, where the indicator of women's empowerment is estimated using a structural model. One might also use MR-LATE when treatment is estimated via other means, such as machine learning methods, or by an index of proxy measures.

The use of estimated treatment can provide the advantages of structural models while mitigating the costs of misspecification. For example, in our application, one could directly estimate the causal impact of inheritance law changes on health outcomes by standard methods. But that would reveal nothing about the effects on health of alternative policies that empower women. By instead estimating the direct effect of women's household resource control on health (using inheritance law changes as an instrument), we can evaluate the impact on health of alternative policies that would empower women. More broadly, by exploiting structure, we can estimate causal effects of substantial economic interest and relevance. This may be particularly useful for constructing causal tests and benchmarks of economic models of behavior, since the researcher can directly focus on treatments that are motivated by theory (in our example, women's control of household resources), instead of only calculating the treatment effects of proxies that happen to be directly observed.

Appendix

Our Appendix, which is available online, contains three main sections. Appendix A provides a graphical illustration of the MR-LATE estimator. Appendix B discusses several details of our empirical application: the derivation of the demand equations for private assignable goods is in Appendix B.1, details on data sources and estimation samples are presented in Section B.2, a validation of our structural estimates is in Appendix B.3, and a discussion of possible violations of the exclusion restriction is in Appendix B.4. Additional figures and tables are in Appendix C.

Bibliography

ABADIE, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292. [9], [10]

AGARWAL, B. (1995): "A Field of One's Own," *Cambridge Books*. [24]

AGÜERO, J. M. AND V. FRISANCHO (2020): "Measuring Violence Against Women with Experimental Methods," Tech. rep. [19]

AIGNER, D. J. (1973): "Regression with a binary independent variable subject to errors of observation," *Journal of Econometrics*, 1, 49 – 59. [4]

ALDERMAN, H., J. DAS, AND V. RAO (2013): "Conducting ethical economic research: complications from the field," *World Bank Policy Research Working Paper*. [19]

ALMEIDA, H., M. CAMPELLO, AND J. GALVAO, ANTONIO F. (2010): "Measurement Errors in Investment Equations," *The Review of Financial Studies*, 23, 3279–3328. [19]

ANGRIST, J. D. AND G. W. IMBENS (1995): "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American statistical Association*, 90, 431–442. [27]

ANGRIST, J. D. AND A. B. KRUEGER (1999): "Chapter 23 - Empirical Strategies in Labor Economics," Elsevier, vol. 3, Part A of *Handbook of Labor Economics*, 1277 – 1366. [4]

APPS, P. F. AND R. REES (1988): "Taxation and the Household," *Journal of Public Economics*, 35, 355 – 369. [4]

ATHEY, S., R. CHETTY, G. IMBENS, AND H. KANG (2016): "Estimating Treatment Effects using Multiple Surrogates: The Role of the Surrogate Score and the Surrogate Index," *arXiv preprint arXiv:1603.09326*. [4]

ATHEY, S. AND G. W. IMBENS (2019): "Machine Learning Methods That Economists Should Know About," *Annual Review of Economics*, 11, 685–725. [3]

BARGAIN, O., G. LACROIX, AND L. TIBERTI (2018): "Validation of Collective Models for Individual Welfare Analysis using Direct Evidence on Sharing," mimeo. [5]

BARRON, J. M., M. C. BERGER, AND D. A. BLACK (1997): "How Well Do We Measure Training?" *Journal of Labor Economics*, 15, 507–528. [19]

BATTISTIN, E., M. D. NADAI, AND B. SIANESI (2014): "Misreported schooling, multiple measures and returns to educational qualifications," *Journal of Econometrics*, 181, 136 – 150. [2]

BECKER, G. S. (1965): "A Theory of the Allocation of Time," *The Economic Journal*, 493–517. [4]

——— (1981): *A Treatise on the Family*, Harvard University Press. [4]

BEEGLE, K., E. FRANKENBERG, AND D. THOMAS (2001): "Bargaining Power Within Couples and Use of Prenatal and Delivery care in Indonesia," *Studies in family planning*, 32, 130–146. [5]

BEGG, C. B. AND D. H. LEUNG (2000): "On the use of surrogate end points in randomized trials," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 163, 15–28. [4]

BLACK, D., S. SANDERS, AND L. TAYLOR (2003): "Measurement of Higher Education in the Census and Current Population Survey," *Journal of the American Statistical Association*, 98, 545–554. [4], [19]

BLACK, D. A., M. C. BERGER, AND F. A. SCOTT (2000): "Bounding Parameter Estimates with Nonclassical Measurement Error," *Journal of the American Statistical Association*, 95, 739–748. [4], [19]

BOLLINGER, C. R. (1996): "Bounding mean regressions when a binary regressor is mismeasured," *Journal of Econometrics*, 73, 387 – 399. [4]

BOUND, J., C. BROWN, AND N. MATHIOWETZ (2001): "Chapter 59 - Measurement Error in Survey Data," Elsevier, vol. 5 of *Handbook of Econometrics*, 3705 – 3843. [4], [18]

BROWN, C., J. PENGLASE, AND R. CALVI (2018): "Sharing the Pie: Undernutrition, Intra-household Allocation, and Poverty," Boston college. [5]

BROWNING, M., P.-A. CHIAPPORI, AND A. LEWBEL (2013): "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power," *Review of Economic Studies*, 80, 1267–1303. [4]

CALVI, R. (2020): "Why are older women missing in India? The age profile of bargaining power and poverty," *Journal of Political Economy*, 128, 2453–2501. [3], [5], [22], [24]

CALVI, R. AND A. KESKAR (2021a): "Dowries, Resource Allocation, and Poverty," *CEPR Discussion Paper*. [5]

——— (2021b): "'Til Dowry Do Us Part: Bargaining and Violence in Indian Families," . [19]

CALVI, R., J. PENGLASE, D. TOMMASI, AND A. WOLF (2020): "The More the Poorer? Resource Sharing and Scale Economies in Large Families," Iza discussion paper. [5]

CARD, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957–979. [19]

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160. [4]

CARSLAW, C. A. (1988): "Anomalies in income numbers: Evidence of goal oriented behavior," *Accounting Review*, 321–327. [19]

CHIAPPORI, P.-A. (1988): "Rational household labor supply," *Econometrica: Journal of the Econometric Society*, 63–90. [3], [4]

——— (1992): "Collective Labor Supply and Welfare," *Journal of Political Economy*, 100, pp. 437–467. [3], [4]

CROSSLEY, T. F. AND S. KENNEDY (2002): "The reliability of self-assessed health status," *Journal of Health Economics*, 21, 643 – 658. [19]

DEININGER, K., A. GOYAL, AND H. NAGARAJAN (2013): "Women's Inheritance Rights and Intergenerational Transmission of Resources in India," *Journal of Human Resources*, 48, 114–141. [24]

DITRAGLIA, F. J. AND C. GARCÍA-JIMENO (2019): "Identifying the effect of a mis-classified, binary, endogenous regressor," *Journal of Econometrics*, 209, 376–390. [2]

DUFLO, E. (2003): "Grandmothers and Granddaughters: Old-age Pensions and Intrahousehold Allocation in South Africa," *The World Bank Economic Review*, 17, 1–25. [5]

DUNBAR, G. R., A. LEWBEL, AND K. PENDAKUR (2013): "Children's Resources in Collective Households: Identification, Estimation, and an Application to Child Poverty in Malawi," *American Economic Review*, 103, 438–71. [3], [5], [21]

DURTSCHI, C., W. HILLISON, AND C. PACINI (2004): "The effective use of Benford's law to assist in detecting fraud in accounting data," *Journal of forensic accounting*, 5, 17–34. [19]

DUSTMANN, C. AND A. V. SOEST (2001): "Language Fluency and Earnings: Estimation with Misclassified Language Indicators," *The Review of Economics and Statistics*, 83, 663–674. [19]

FRANGAKIS, C. E. AND D. B. RUBIN (2002): "Principal stratification in causal inference," *Biometrics*, 58, 21–29. [4]

FRANKENBERG, E., D. THOMAS, ET AL. (2001): *Measuring Power*, Citeseer. [5]

FRAZIS, H. AND M. A. LOEWENSTEIN (2003): "Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables," *Journal of Econometrics*, 117, 151 – 178. [4]

GANDELMAN, N. AND A. RASTELETTI (2017): "Credit constraints, sector informality and firm investments: evidence from a panel of Uruguayan firms," *Journal of Applied Economics*, 20, 351 – 372. [19]

GUNDERSEN, C., B. KREIDER, AND J. PEPPER (2012): "The impact of the National School Lunch Program on child health: A nonparametric bounds analysis," *Journal of Econometrics*, 166, 79 – 91, annals Issue on "Identification and Decisions", in Honor of Chuck Manski's 60th Birthday. [19]

HEATH, R. AND X. TAN (2019): "Intrahousehold Bargaining, Female Autonomy, and Labor Supply: Theory and Evidence from India," *Journal of the European Economic Association*, Forthcoming. [3], [22], [24]

HERNANDEZ, M., S. PUDNEY, AND R. HANCOCK (2007): "The welfare cost of means-testing: pensioner participation in income support," *Journal of Applied Econometrics*, 22, 581–598. [4]

HU, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144, 27 – 61. [4]

IMAI, K., L. KEELE, AND T. YAMAMOTO (2010): "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statist. Sci.*, 25, 51–71. [4]

IMBENS, G. W. AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [2], [5], [7], [9], [23], [25]

JUDGE, G. AND L. SCHECHTER (2009): "Detecting problems in survey data using Benford's Law," *Journal of Human Resources*, 44, 1–24. [19]

KANE, T. J., C. E. ROUSE, AND D. STAIGER (1999): "Estimating Returns to Schooling When Schooling is Misreported," NBER working paper 7235. [4]

KLEPPER, S. (1988): "Bounding the effects of measurement error in regressions involving dichotomous variables," *Journal of Econometrics*, 37, 343 – 359. [4]

KREIDER, B., C. GUNDERSEN, AND D. JOLLIFFE (2012): "Identifying the effects of food stamps on child health outcomes when participation is endogenous and misreported," . [4]

KREIDER, B., R. J. MANSKI, J. MOELLER, AND J. PEPPER (2015): "The Effect of Dental Insurance on the Use of Dental Care for Older Adults: A Partial Identification Analysis," *Health Economics*, 24, 840–858. [19]

KREIDER, B. AND J. V. PEPPER (2007): "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of the American Statistical Association*, 102, 432–441. [4], [19]

LAFAVE, D. AND D. THOMAS (2017): "Extended Families and Child Well-being," *Journal of Development Economics*, 126, 52–65. [5]

LEWBEL, A. (2007): "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75, 537–551. [4]

LEWBEL, A. AND K. PENDAKUR (2008): "Estimation of Collective Household Models with Engel Curves," *Journal of Econometrics*, 147, 350–358. [4]

LITTLE, R. J. A. AND D. B. RUBIN (2019): *Statistical Analysis with Missing Data*, USA: John Wiley & Sons, Inc. [2]

MAHAJAN, A. (2006): "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631–665. [4]

MANSKI, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80, 319–323. [4]

MARCHI, S. D. AND J. T. HAMILTON (2006): "Assessing the Accuracy of Self-Reported Data: an Evaluation of the Toxics Release Inventory," *Journal of Risk and Uncertainty*, 32, 57–76. [19]

MEYER, B. D., N. MITTAG, AND R. M. GOERGE (2018): "Errors in Survey Reporting and Imputation and their Effects on Estimates of Food Stamp Program Participation," Working Paper 25143, National Bureau of Economic Research. [18]

MEYER, B. D., W. K. C. MOK, AND J. X. SULLIVAN (2015): "The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences," Working Paper 15181, National Bureau of Economic Research. [2], [18]

MITCHELL, C. (2010): "Are divorce studies trustworthy? The effects of survey nonresponse and response errors," *Journal of Marriage and Family*, 72, 893–905. [18]

MOLINARI, F. (2010): "Missing Treatments," *Journal of Business & Economic Statistics*, 28, 82–95. [2], [3], [4], [18]

NYE, J. AND C. MOUL (2007): "The political economy of numbers: on the application of Benford's law to international macroeconomic statistics," *The BE Journal of Macroeconomics*, 7. [19]

O'CONNELL, M. (2006): "The Visible Hand: Editing Marital-History Data from Census Bureau Surveys," in *Handbook of Measurement Issues in Family Research*, ed. by S. L. Hofferth and L. M. Casper, The address of the publisher: Mahwah, NJ: Lawrence Erlbaum, chap. 9. [2], [18]

PALERMO, T., J. BLECK, AND A. PETERMAN (2014): "Tip of the iceberg: reporting and gender-based violence in developing countries," *American journal of epidemiology*, 179, 602–612. [19]

PENGLASE, J. (2017): "Consumption Inequality among Children: Evidence from Child Fostering in Malawi," Boston college. [5]

PRENTICE, R. L. (1989): "Surrogate endpoints in clinical trials: definition and operational criteria," *Statistics in medicine*, 8, 431–440. [4]

QUISUMBING, A. R. AND J. A. MALUCCIO (2003): "Resources at Marriage and Intrahousehold Allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa," *Oxford Bulletin of Economics and Statistics*, 65, 283–327. [5]

ROSENBAUM, P. R. (1984): "The consequences of adjustment for a concomitant variable that has been affected by the treatment," *Journal of the Royal Statistical Society. Series A (General)*, 656–666. [4]

ROY, S. (2008): "Female Empowerment Through Inheritance Rights: Evidence from India," Mimeo, London School of Economics, London. [24]

——— (2015): "Empowering women? Inheritance rights, female education and dowry payments in India," *Journal of Development Economics*, 114, 233–251. [24]

SOKULLU, S. AND C. VALENTE (2017): "Individual Consumption in Collective Households: Identification Using Panel Data with an Application to PROGRESA," mimeo. [5]

THOMAS, D. (1990): "Intra-household Resource Allocation: An Inferential Approach," *Journal of human resources*, 635–664. [5]

TOMMASI, D. (2019): "Control of Resources, Bargaining Power and the Demand of Food: Evidence from PROGRESA," *Journal of Economic Behavior & Organization*, 161, 265–286. [5]

TOMMASI, D. AND L. ZHANG (2020a): "Bounding Program Benefits When Participation Is Misreported," Iza discussion paper series no. 13430. [2]

——— (2020b): "Identification and Estimation of Program Benefits When Participation is Misreported," mimeo. [27]

URA, T. (2018): "Heterogeneous treatment effects with mismeasured endogenous treatment," *Quantitative Economics*, 9, 1335–1370. [2], [10], [12]

VANDERWEELE, T. (2015): *Explanation in causal inference: methods for mediation and interaction*, Oxford University Press. [4]

WORLD BANK (2014): "Women, Business and the Law 2014. Removing Restrictions to Enhance Gender Equality," *World Bank Report*. [23]

WOSSEN, T., T. ABDOULAYE, A. ALENE, P. NGUIMKEU, S. FELEKE, I. Y. RABBI, M. G. HAILE, AND V. MANYONG (2018): "Estimating the Productivity Impacts of Technology Adoption in the Presence of Misclassification," *American Journal of Agricultural Economics*, 101, 1–16. [19]

YANAGI, T. (2018): "Inference on local average treatment effects for misclassified treatment," *Econometric Reviews*, 0, 1–23. [2]

ZHANG, Z., W. LIU, B. ZHANG, L. TANG, AND J. ZHANG (2016): "Causal inference with missing exposure information: Methods and applications to an obstetric study," *Statistical methods in medical research*, 25. [18]

Online Appendix for

LATE with Missing or Mismeasured Treatment

Rossella Calvi, Arthur Lewbel, and Denni Tommasi

Contents

A   A Graphical Illustration of MR-LATE for Generally Mismeasured Treatment

Assume that $supp\,(\varepsilon) \subset [\kappa^b - e, \kappa^a - e)$. Then it follows that for $T = T^a$ we have $p_1^a = 1$ with $p_0^a = 0$, and for $T = T^b$ we have $p_1^b = 0$ and $p_0^b = 1$, and so $\lambda^a - \lambda^b = E[Y_1 - Y_0 \mid C]$. Given corollary 2, LATE can be point identified. Figure A1 provides a graphical representation of this. If there was no measurement error, the true treatment and control groups would coincide with the respective observed groups. All individuals on the black line on the right hand side of e, would have a R^* larger than the threshold value; otherwise, they would be on the black line on the left hand side of e. One could construct a treatment proxy $T = \mathbb{I}(R \geq c)$, where R is an estimate of R^* and c is one's best guess of the midpoint between $\varepsilon + e$. This approach, however, will not identify the treatment effect of interest. To achieve point identification of LATE in presence of measurement error or misclassification error, we need to have two treatment indicators, T^a and T^b, such that $q^a = p_1^a/(p_1^a - p_0^a) = 1$ and $q^b = p_1^b/(p_1^b - p_0^b) = 0$. By knowing the bounds κ^a and κ^b, we are able to define a T^a such that for all individuals on the red line on the left hand side of κ^a, $p_0^a = 0$. That is, with probability 0, these individuals, who are observed in the control group, belong to the true treatment group. Analogously, we are able to define also a T^b such that for all individuals on the blue line on the right hand side of κ^b, $p_1^b = 0$. That is, with probability 0, these individuals, who are observed in the treatment group, belong to the true control group.
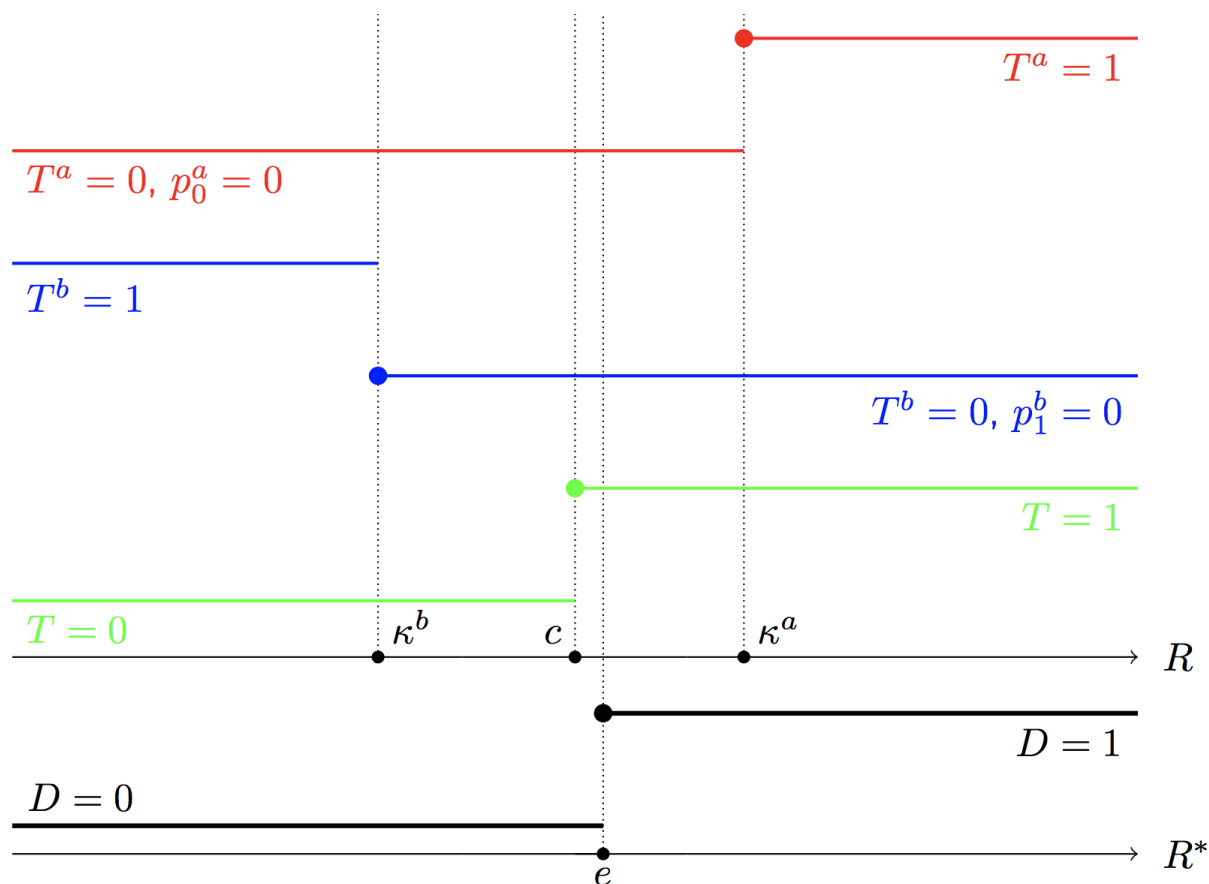

Figure A1: Illustrative Example: MR-LATE and Threshold Crossing Treatment Model

B   Women's Empowerment and Family Health: Details

B.1   Derivation of Household Demand Equations of Private Assignable Goods

Here we summarize the derivation of our structural model, based on Browning et al. (2013) (BCL) and Dunbar et al. (2013) (DLP), for estimating resource shares from the demand equations of private assignable goods. Consider a household comprised of T types of individuals indexed $t = 1, ..., T$. Recall M is the total expenditures of the household, i.e., the household's total budget, X denotes a vector of observable attributes of households and their members, $\widetilde{Z}$ denotes a vector of possible distribution factors (if any), and $Q_1, ..., Q_T$ are quantities of each private assignable good consumed by household member t. Let S be a vector of quantities of all other goods the household consumes. Unlike $Q_1, ..., Q_T$, the goods S may be shared and hence jointly consumed to some extent. In particular, $S = \sum_{t=1}^{T} S_t$ where S_t is the vector of quantities of these goods consumed by member t. The purchased quantities of these goods are given by $A(X)S$, where the matrix $A(X)$ summarizes the extent to which these goods are shared.

Let $P_1, ..., P_T$ be the market prices of the private assignable goods, let P_S be the vector of market prices of goods S, and let P denote the vector of all of these prices.

The household chooses what to consume using the program

$$\max_{Q_1, ..., Q_T, S_1, ... S_T} \widetilde{V} \left[V_1(Q_1, S_1, X), ..., V_T(Q_T, S_T, X) \mid \widetilde{Z}, X, P/M \right] \tag{A1}$$

$$\text{such that} \quad S = \sum_{t=1}^{T} S_t \quad \text{and} \quad M = P_S' A(X) S + \sum_{t=1}^{T} P_t Q_t$$

where $V_t(Q_t, S_t, X)$ for $t = 1, ..., T$ is the utility function of household member t, and the function $\widetilde{V}$ describes the social welfare function or bargaining process of the household. A function $\widetilde{V}$ exists because the household is Pareto efficient.

What makes $Q_1, ..., Q_T$ be private is that they are not shared. What makes them assignable is that the econometrician can observe who consumes each. In particular, each member t has quantity Q_t in his or her utility function, and does not have Q_ℓ for all $\ell \neq t$ in his or her utility function. The square matrix $A(X)$ is what is called by BCL a linear consumption technology function over goods. Having $A(X)$ differ from the identity matrix is what allows goods in S to be partly shared and/or consumed jointly. In particular, $A(X)S$ equals the quantity vector of these goods that the household actually purchases, while $S = \sum_{t=1}^{T} S_t$ is total quantity vector of these goods that the household consumes. These quantities are not the same due to sharing and joint consumption. The smaller an element of $A(X)S$ is relative to the corresponding element of S, the more that good is shared or jointly consumed. See BCL for details.

Household attributes X may affect preferences, and so appear inside the utility functions V_t. These X variables can also affect the extent to which goods are shared through $A(X)$, and they can directly affect the bargaining process or social welfare function given by $\widetilde{V}$ (by, e.g., affecting the relative bargaining power of members). As a result, resource shares may also depend on X. The difference between X and distribution factors $\widetilde{Z}$ is that the vector $\widetilde{Z}$ appears in the model only as arguments of $\widetilde{V}$, and so only directly affects the allocation of resources within the household, but not the tastes of the individual household members or the jointness of consumption.

Applying duality theory and decentralization welfare theorems, it follows from BCL that the household's program above is equivalent to a program where each household member t chooses what to consume using the program

$$\max_{Q_t, S_t} V_t\left(Q_t, S_t, X\right) \quad \text{such that} \quad \eta_t(P, M, X, \widetilde{Z})M = P_S' A\left(X\right) S_t + P_t Q_t \tag{A2}$$

where $\eta_t = \eta_t(P, M, X, \widetilde{Z})$ is the resource share of member t, that is, η_t is the fraction of total household resources M that are allocated to member t. This member then chooses quantities Q_t and the vector S_t subject to a linear budget constraint. The vector $P_S A\left(X\right)$ equals the vector of shadow prices of goods S. These shadow prices for the household may be lower than market prices, due to sharing. Being private and assignable, the shadow price of each Q_t equals its market price P_t. Let $\widetilde{M}_t = \eta_t M$ denote the shadow budget for member t. As shown in BCL, the resource share functions $\eta_t(P, M, X, \widetilde{Z})$ for each member t in general depend on the function $\widetilde{V}$ and on the utility functions $V_1 ..., V_T$.

BCL show that the more bargaining power a household member has (i.e., the greater is the weight of his or her utility function in $\widetilde{V}$), the larger is their resource share η_t. Resource shares η_t all lie between zero and one, and resource shares sum to one, that is, $\sum_{t=1}^{T} \eta_t = 1$.

As in DLP, we will not work with the household demand functions of all goods (which, as shown in BCL, are rather complicated). Instead, we only make use of the demand functions of the private assignable goods Q_t, which are simpler. Since equation (A2) is an ordinary utility function maximized under a linear budget constraint (linear in shadow prices and a shadow budget), the solution to equation (A2) is a set of Marshallian demand equations for Q_t and S_t.

Let $h_t\left(\widetilde{M}_t, P, X\right)$ be the Marshallian demand function of person t' for their private assignable good, that is, $h_t\left(\widetilde{M}_t, P, X\right)$ is the quantity person t in a household with member attributes X would demand of their assignable good if they had a budget equal to their shadow budget $\widetilde{M}_t$ and faced the within-household shadow price vector that corresponds to the market price vector P. Since each Q_t is private and assignable, the quantity Q_t that member t chooses to consume equals the quantity of this good that the household buys. It therefore follows from the above that the household's quantity demand of each private assignable good Q_t is given by

$$Q_t = h_t\left(\eta_t(P, M, X, \widetilde{Z})M, P, X\right) \quad \text{for} \quad t = 1, ..., T. \tag{A3}$$

The interpretation of this equation is that the total resources allocated to member t are $\eta_t M$ (the share η_t of total household budget M) and the function h_t is that member's Marshallian demand function for this good. Since the good is private and assignable, the household's demand for the good just equals that member's own demand for the good. It is important to note that only private assignable goods have the simple form given by equation (A3). The demand functions for other goods are much more complicated, as in BCL.

Let $\widetilde{h}_t\left(\widetilde{M}_t, P, X\right) = P_t h_t\left(\widetilde{M}_t, P, X\right)/\widetilde{M}_t$ denote the Marshallian demand function written in budget share form. That is, $\widetilde{h}_t\left(\widetilde{M}_t, P, X\right)$ is the fraction of the total budget $\widetilde{M}_t$ that is spent on the good t. DLP assume data are drawn from single price regime (that is, Engel curve data), so P is a fixed constant that can be dropped from the model. They provide empirical and the-

oretical evidence that η_t does not depend on M.[1] This allows them to rewrite equation (A3) as $w^t = \eta_t(X, \widetilde{Z})\widetilde{h}_t\left(\eta_t(X, \widetilde{Z})M, X\right)$ for $t = 1, ..., T$, where $w^t = P_t Q_t / M$ is the household's budget share of good t, that is, the fraction of the household's total budget M that is spend on buying Q_t. DLP provide a class of functional forms for the utility functions $\widetilde{V}$ that make $\widetilde{h}_t$ linear in the log of its first argument, so $w^t = \eta_t(X, \widetilde{Z})[\delta^t(X) + (lnM + ln\eta_t(X, \widetilde{Z}))\beta(X)]$ for some functions $\delta^t(X)$ and $\beta(X)$. The assumption that $\beta(X)$ does not depend on t is what DLP call the SAP (similar across people) assumption.

B.2 Survey Data and Descriptive Statistics

For our empirical analysis, we employ two different datasets from India. One, the 62nd round of the NSS Consumer Expenditure Survey (NSS), contains detailed consumption data that we use to structurally estimate resource shares. The other, the 3rd round of the National Family Health Survey (NFHS), collects the health outcomes for women, men, and children, as well as the information needed for the construction of our instrument. Both surveys were conducted between 2005 and 2006.

NSS data. The 2005-2006 NSS Consumer Expenditure Survey contains detailed data on household expenditures, socio-economic characteristics, and other particulars of household members. We select households consisting of a mother, a father, and one to four children. More precisely, we select households with one woman and one man above age 15 (with one of these designated as the head of household), and children under 15. We exclude households in the top or bottom 1 percent of expenditure, and we exclude households that report having performed any ceremony during the month prior to the survey, as unusual purchases of clothing items and non-standard expenditure patterns may occur for festivities and ceremonies. Among other items, households are asked to report how much they spent on clothing and footwear. Given the detailed breakdown of clothing expenditure, it is possible to identify the expenditures on some items of clothing that can be specifically assigned to women and to men, thereby allowing us to construct expenditures on private assignable clothing for each decision maker. We define expenditure on women's assignable clothing as the sum of expenditures on saree, chaddar, dupatta, and shawl. For men's assignable clothing, we combine expenditure on dhoti, lungi, salwar, pajamas, and shirts. Notice that Tommasi and Wolf (2018) shows that if the data exhibit relatively flat Engel curves in the consumption of the private assignable goods, then the DLP model can be weakly identified. However, households in our dataset display a large variation in the consumption of private assignable goods (see Figure A2). Hence, we do not appear to have a weak identification problem with our data.

Table A1 contains some descriptive statistics. For clothing items, the NSS reports expenditures that occurred in the past 365 days. For simplicity and consistency with other data, we convert these annual expenditures into monthly figures. We consider observable attributes that characterize each individual, the household, and the environment of the household. Specifically, these attributes include the gender composition of children, the wife's age, the age gap between spouses, the average age of children, and indicator variables for the number of children, geographic region, religion

---

[1] Lise and Seitz (2011), Lewbel and Pendakur (2008), Bargain and Donni (2012), Bargain et al. (2014) and DLP all use this restriction in their identification results, and supply some theoretical arguments for it. Cherchye et al. (2015) and Menon et al. (2012) provide empirical support for this restriction.

(Hindu, Buddhist, Sikh or Jain), for living in rural areas, for female and male higher education, and for belonging to a Scheduled Caste, Scheduled Tribe, or other backward classes.

NFHS data. The 2005-2006 National Family Health Survey provides a range of health indicators for women aged 15 to 49, for men aged 15 to 54, and for children born in the 5 years prior to the date of interview. The survey also contains many demographic and socio-economic attributes, comparable to those we observe in the NSS data. As above, we select households consisting of a mother, a father, and their children. We consider women, men and children datasets separately, observing a few different health measures for each individual. The health measures for adults include body mass index or BMI (weight in kilograms divided by height in meters squared) and measures of anemia. A BMI cut-off point of 18.5 is used to define undernutrition. Anemia is a condition in which the number of red blood cells, or their oxygen-carrying capacity, is insufficient. Although its primary cause is iron deficiency, it often coexists with (and hence serves as an indicator of) a number of other health issues such as malaria, parasitic infection, and nutritional deficiencies.

For children, the health related measures we observe include weight-for-age and height-for-age z-scores (standard deviations from the reference median based on the 2006 WHO Child Growth Standards). A z-score greater than 2 indicates over-nourishment with respect to the corresponding anthropometric measurements. Deficits on these indicators (measured by their values less than -2 standard deviations below the median) are known as underweight and stunting, respectively. Another child health measure we observe is mothers' reports of whether a child was sick with fever, cough or diarrhea in the past two weeks. Finally, we observe child vaccination records, which we use to construct an additional indicator variable equal to one if a child has ever received any vaccine to prevent diseases. Specifically, we observe whether a child received a BCG vaccine (against tuberculosis), one to three DPT vaccines (against diphtheria, pertussis, and tetanus), and one to four polio vaccines (at birth and one to three years after).

With the exception of a few variables, the household socio-economic characteristics are on average quite similar in the two samples. The main differences are related to the definitions of completed schooling and land ownership in the two surveys. Moreover, the NFHS covers the 29 states in India, while the NSS includes both the 29 states and the 7 union territories of India. Any errors introduced by the use of two different samples will take the form of estimation error in R, and so should be accounted for by the MR-LATE estimator.

6

Table A1: NSS Consumer Expenditure Data and NFHS Household Data

|	2005-2006 NSS Sample				2005-2006 NFHS Sample			
	Obs.	Mean	Median	St. Dev.	Obs.	Mean	Median	St. Dev.
Woman's Assignable Clothing Budget Share	7,744	0.88	0.75	0.69				
Man's Assignable Clothing Budget Share	7,744	0.93	0.71	0.87				
Total Expenditure (Rupees)	7,744	5995.56	4211.56	5027.59				
Number of Children	7,744	2.31	2.00	1.13	25,218	2.39	2.00	1.21
Fraction of Female Children	7,744	0.45	0.50	0.36	25,218	0.47	0.50	0.36
Woman's Age	7,741	31.08	30.00	6.38	25,215	30.33	30.00	7.20
Age Gap (Man - Woman)	7,737	5.17	5.00	4.25	25,180	5.81	5.00	4.99
Children's Avg. Age	7,744	6.91	7.00	3.48	25,218	6.50	6.50	3.50
$\mathbb{I}$(Hindu, Buddhist, Jain, Sikh)	7,744	0.79	1.00	0.41	25,214	0.76	1.00	0.43
$\mathbb{I}$(Sch. Caste, Sch. Tribe, Oth. Back. Caste)	7,744	0.65	1.00	0.48	24,280	0.68	1.00	0.47
$\mathbb{I}$(Own Land)	7,710	0.69	1.00	0.46	25,212	0.36	0.00	0.48
$\mathbb{I}$(Woman Completed High School)	7,744	0.14	0.00	0.35	25,218	0.08	0.00	0.27
$\mathbb{I}$(Man Completed High School)	7,744	0.22	0.00	0.41	25,218	0.13	0.00	0.33
$\mathbb{I}$(Rural)	7,744	0.48	0.00	0.50	25,218	0.55	1.00	0.50
$\mathbb{I}$(North)	7,744	0.31	0.00	0.46	25,218	0.31	0.00	0.46
$\mathbb{I}$(East)	7,744	0.21	0.00	0.40	25,218	0.16	0.00	0.37
$\mathbb{I}$(North-East)	7,744	0.14	0.00	0.35	25,218	0.20	0.00	0.40
$\mathbb{I}$(South)	7,744	0.21	0.00	0.41	25,218	0.20	0.00	0.40
$\mathbb{I}$(West)	7,744	0.12	0.00	0.33	25,218	0.13	0.00	0.33

Notes: Budget shares are multiplied by 100. Woman's assignable clothing includes expenditures on saree, shawls, chaddar, and dupatta; man's assignable clothing includes expenditures on dhoti, lungi, pajamas, salwar, and shirts. Age variables are divided by 10. North India includes Jammu & Kashmir, Himachal Pradesh, Punjab, Uttaranchal, Haryana, Delhi, Rajasthan, Uttar Pradesh, and Madhya Pradesh. East India includes West Bengal, Bihar, Jharkhand, Orissa, A & N Islands, and Chattisgarh. North-East India includes Sikkim, Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, and Tripura. South India includes Karnataka, Tamil Nadu, Andhra Pradesh, Kerala, Lakshadweep, and Pondicherry. West India includes Gujarat, Goa, Maharashtra, Daman & Diu, and D & N Haveli.
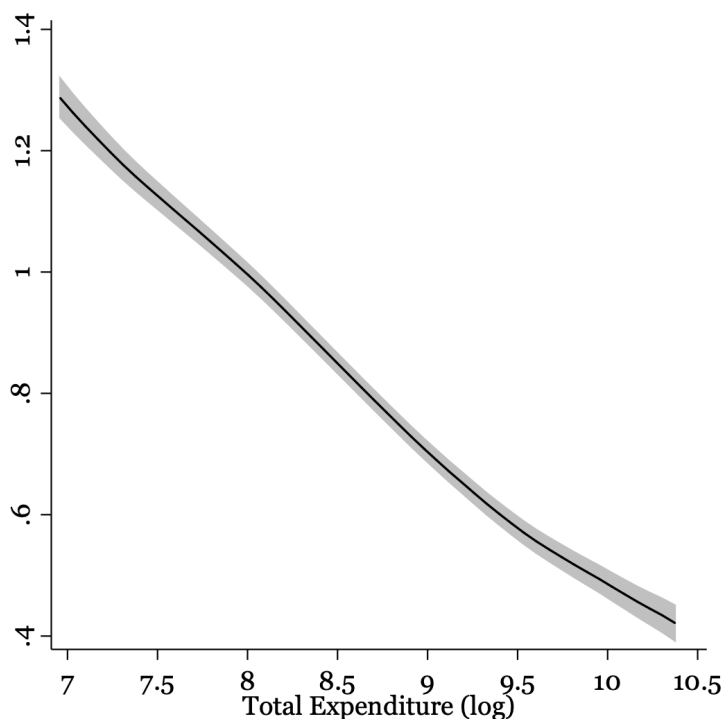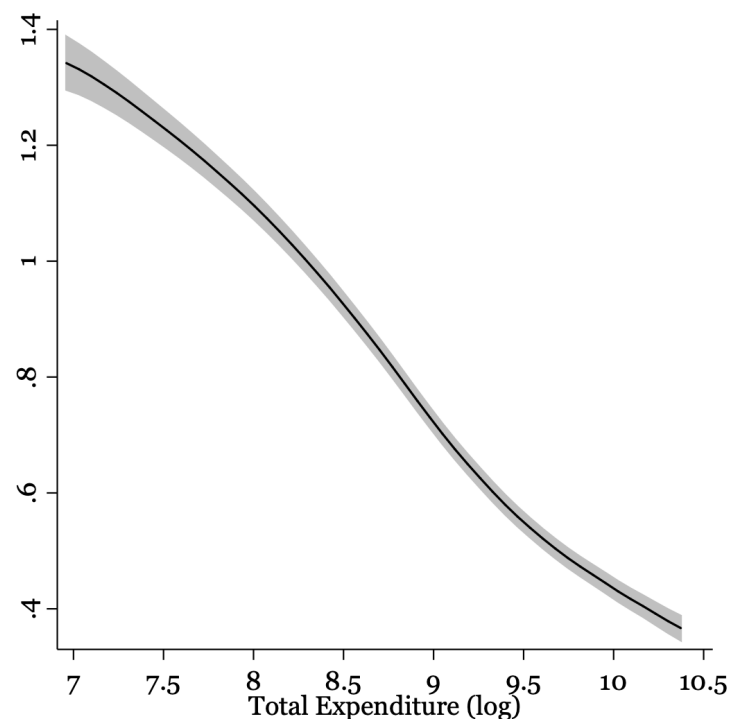
(A) Woman's Assignable Clothing   (B) Man's Assignable Clothing

Figure A2: Non-parametric Engel Curves

Table A2: 2005-2006 NFHS Individual Data

	Women (N = 21,057)			Men (N = 11,516)			Children (N = 18,411)		
	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
Body Mass Index (BMI)I	20.93	20.18	3.93	21.20	20.66	3.52			
$\mathbb{I}$(BMI$\leq$18.5)	0.30	0.00	0.46	0.24	0.00	0.43			
$\mathbb{I}$(Anemic)	0.15	0.00	0.36	0.09	0.00	0.29			
Weight for Age (z-score)							-1.80	-1.88	1.20
Height for Age (z-score)							-1.67	-1.69	1.60
$\mathbb{I}$(Cough in last 2 weeks)							0.17	0.00	0.38
$\mathbb{I}$(Fever in last 2 weeks)							0.15	0.00	0.35
$\mathbb{I}$(Diarrhea in last 2 weeks)							0.09	0.00	0.29
$\mathbb{I}$(Any Vaccination)							0.90	1.00	0.29
$\mathbb{I}$(HSA Exposed)	0.17	0.00	0.38	0.26	0.00	0.44	0.15	0.00	0.36
Number of Children	2.43	2.00	1.21	2.43	2.00	1.20	2.80	3.00	1.37
Fraction of Female Children	0.47	0.50	0.36	0.48	0.50	0.36	0.51	0.50	0.34
Women's Age	0.30	0.30	0.06	0.30	0.29	0.06	0.27	0.27	0.05
Age Gap (Men - Women)	0.06	0.05	0.04	0.06	0.05	0.04	0.06	0.05	0.04
Children's Avg. Age	0.06	0.06	0.03	0.06	0.06	0.03	0.04	0.04	0.02
$\mathbb{I}$(Hindu, Buddhist, Jain, Sikh)	0.78	1.00	0.42	0.76	1.00	0.42	0.71	1.00	0.45
$\mathbb{I}$(Sch. Caste, Sch. Tribe, Oth. Back. Caste)	0.69	1.00	0.46	0.72	1.00	0.45	0.73	1.00	0.45
$\mathbb{I}$(Own Land)	0.36	0.00	0.48	0.35	0.00	0.48	0.36	0.00	0.48
$\mathbb{I}$(Woman Completed High School)	0.08	0.00	0.27	0.08	0.00	0.27	0.05	0.00	0.23
$\mathbb{I}$(Man Completed High School)	0.12	0.00	0.33	0.13	0.00	0.33	0.09	0.00	0.29
$\mathbb{I}$(Rural)	0.55	1.00	0.50	0.53	1.00	0.50	0.59	1.00	0.49
$\mathbb{I}$(North)	0.31	0.00	0.46	0.27	0.00	0.44	0.35	0.00	0.48
$\mathbb{I}$(East)	0.16	0.00	0.37	0.10	0.00	0.30	0.17	0.00	0.37
$\mathbb{I}$(North-East)	0.20	0.00	0.40	0.20	0.00	0.40	0.23	0.00	0.42
$\mathbb{I}$(South)	0.20	0.00	0.40	0.30	0.00	0.46	0.15	0.00	0.36
$\mathbb{I}$(West)	0.12	0.00	0.33	0.13	0.00	0.34	0.10	0.00	0.31
$\mathbb{I}$(Child is Female)							0.48	0.00	0.50
Child's Age							0.21	2.00	1.39

Notes: Age variables are divided by 10. $\mathbb{I}$(Anemic) includes moderate anemia (7.0-9.9 g/dl for women and 9.0-11.9 g/dl for men) or severe anemia (less than 7.0 g/dl for women and less than 9.0 g/dl for men). $\mathbb{I}$(Any Vaccination) includes vaccinations against polio, measles, DPT or BCG. Women of age 15 to 49, men of age 15 to 54 and children of age 0 to 5.

## B.3 Resource Shares and Self-reported Decision-making Power

In this section, we compare our structurally-motivated measure of bargaining power R, estimated from household expenditure data, with some more typical proxies of power, namely, women's self-reports of control over various household decisions and mobility.

The NFHS data contains questions of the form, "Who usually makes decisions about [X] in your household?". Specifically, women are asked to report who has the final say over their own health care, household purchases, and visits to family or relatives. We construct indicator variables equal to 1 if the answer to these questions is "respondent alone" or "respondent and husband/partner jointly" and 0 if the answer is "husband/partner." We exclude women who answer "other/someone else" (less than 1 percent in any question). Several women in our sample report having no say in household decisions: 29 percent of women say they do not participate in decisions over their own health, 25 percent report having no say in determining visits to family and friends, and 33 percent claim to have no say in large household purchases. In addition, women are asked whether they are allowed to go alone to places outside the village, to the health facility or to the market. Many women report an inability to go places alone, especially to places outside the village or community (51 percent). One out of three women report not being allowed to go to the market or to a health facility alone. We combine the above information (three questions on women's mobility) with the responses to the three questions on women's participation in household decisions to construct an index of women's autonomy. Specifically, we give each answer a value of one if the response is that wife controls that decision (or if she can go alone to different places), a value of minus one if the response is that the husband controls that decision (or if she cannot go alone), and a value of zero for any other response. We then define our estimated index to equal the sum of the responses across all of the questions. This index is based on the answers to three questions about women's participation in household decision and three questions on women's mobility, for a total of $m = 6$ questions. This index therefore takes on integer values in the range -6 to 6.

Panels (a) to (c) of Figure A3 display the results of non-parametric regressions of women's reported participation in household decisions on our estimated resource share R. Panel (d) shows the non-parametric relationship between our index of women's autonomy and R. In all cases, the presence of positive relationships emerges clearly. We also examine the link between the self-reported decision making and our binary structural treatment variable $T = \mathbb{I}(R \geq 50)$ conditional on individual and household level controls, fixed effects, and state time trends. The estimation results are in Table A3. Overall, women are significantly more likely to report participating in decisions in treated households, i.e., in households where we estimate, based on expenditures, that they have substantial control over resources. Thus, these results corroborate the theory underlying our structural model of treatment: the larger is R (meaning the higher is the likelihood that a woman has control over household resources), the higher are her self reported decision making and bargaining powers within the household.

B.4 Exclusion Restriction

The exclusion restriction needed for identification is that, once fixed effects and time trends are included, being Hindu, Buddhist, Jain or Sikh and unmarried at the time of implementation has
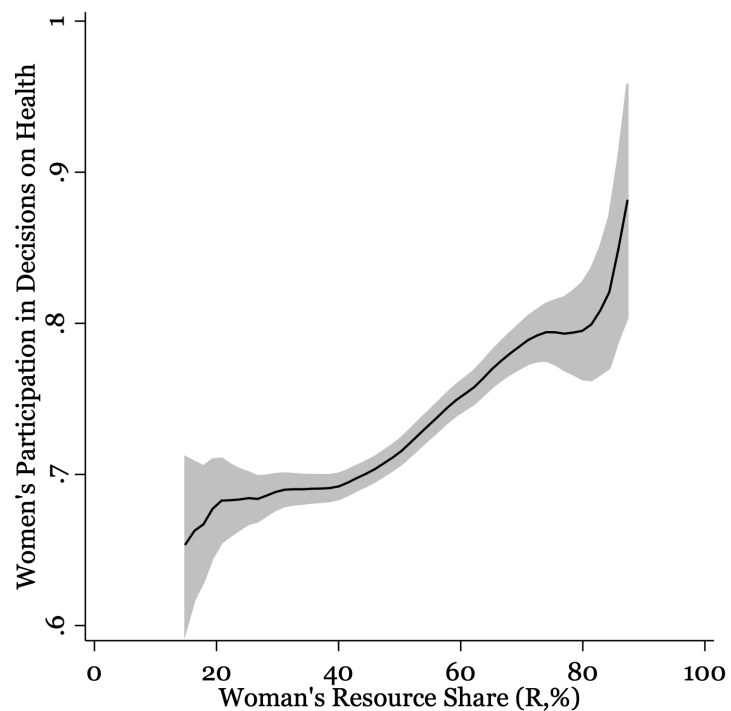
(A) Woman's Health
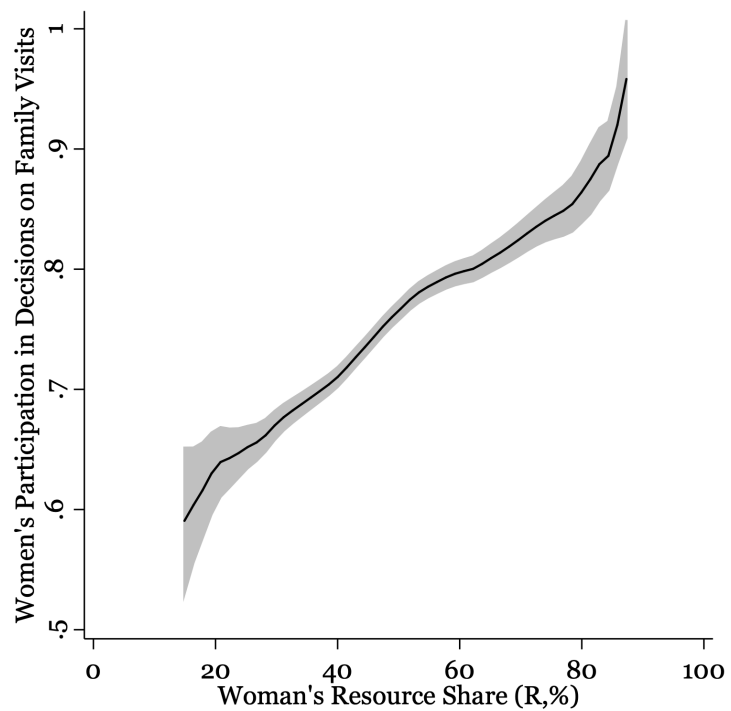
(B) Visits to Family and Relatives
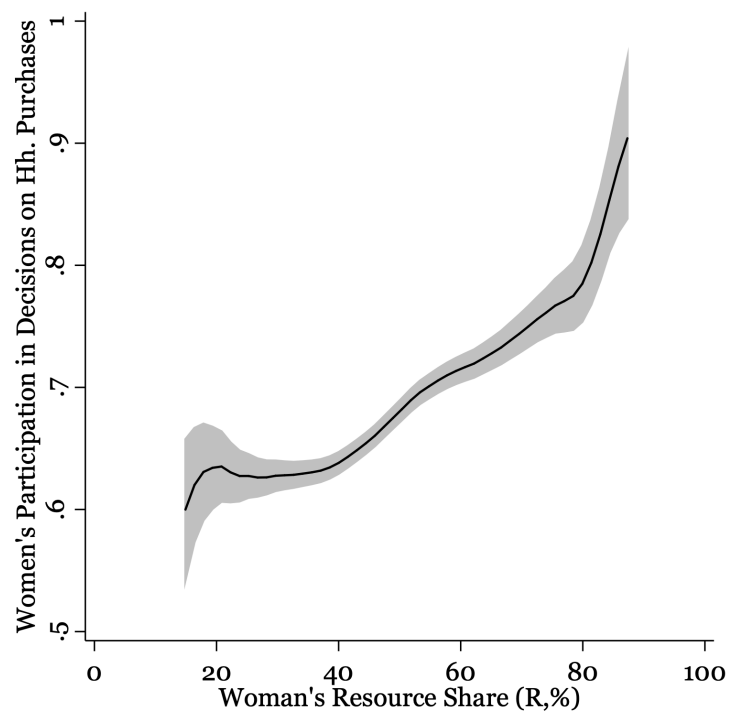
(C) Household Purchases
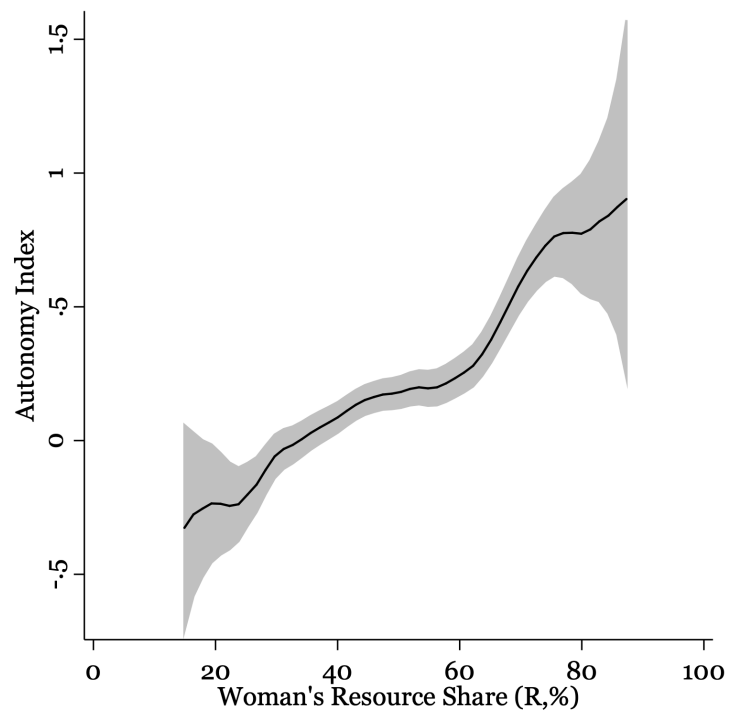
(D) Autonomy Index

Figure A3: Structurally Recovered Bargaining Power and Household Decision Making

Table A3: Self-reported Decision Making and Woman's Control of Resource

	$\mathbb{I}$(Woman Participates in Final Decisions on)			Autonomy
	Household Purchases	Visits to Family and Relatives	Own Health	Index
$T = \mathbb{I}(R \geq 50)$	0.0496	0.0570	0.0456	0.624
	(0.0121)	(0.0109)	(0.0122)	(0.0857)
Observations	21,690	21,706	21,773	21,910

Note: NFHS data. The sample includes married women of age 15 to 49 in nuclear households. Bootstrap standard errors in parentheses. All specifications include individuals and household controls, state-religion fixed effects, mother's cohort-religion fixed effects and state specific time trends (up to degree four).

an effect on health outcomes only through women's intra-household bargaining power. Note that the existence of other changes *resulting from* women's increased power within their marital families due to their HSA exposure would not necessarily invalidate our instrument. Nevertheless, we now explore this issue further. Specifically, we check for the existence of changes in marital sorting and household wealth. We also check for changes in the consumption patterns following the HSA amendments. The NFHS does not include consumption data and the NSS does not record year of marriage. We instead use the 2005 India Human Development Survey (IHDS), which contains both information as well as detailed women's marital histories, to study the effect of the HSA reforms on household wealth, consumption, and marital sorting. Taken together, this analysis does not uncover any evident violation of the exclusion restriction in our context. Nevertheless, as is typical in instrumental variable estimation, we acknowledge that the validity of this assumption cannot be directly tested.

Household Wealth and Spending. The improvements in women's ability to inherit property introduced by the HSA amendments, and the consequent increase in their lifetime unearned income, could have an effect both on women's bargaining power and on the household's wealth and expenditure patterns. If this were true, the exclusion restriction required for identification would be violated. In Table A4, we show that women's exposure to the HSA amendments does not impact the number of assets owned by the household (ranging from 0 to 30) nor a household's total or per-capita expenditure. However, consistent with a collective model where men and women have different preferences over consumption goods, the spending composition changes after the amendments. In line with previous works (Attanasio and Lechene, 2014; Tommasi, 2019; Calvi, 2019), we find that households with more empowered women (i.e., HSAA-exposed) spend more on food and less on tobacco, alcohol, and other intoxicants. Finally, consistent with family members being healthier in households with more empowered women, expenditure on healthcare (in and out-patient care) is lower in households with women who benefited from the amendments.

Table A4: HSA Amendments, Household Wealth and Spending

	Hh. Assets	Hh. Income	Pr(Below Pov. Line)	Per-capita Expend.	Total Expend.	Food Expend.	Medical Expend.	Intoxicants Expend.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HSAA	0.643	4739.1	-0.0115	27.79	109.4	58.48**	-674.2**	-30.00***
	(0.440)	(3454.1)	(0.0289)	(47.12)	(194.0)	(24.60)	(261.5)	(6.101)
N	8,458	8,458	8,458	8,458	8,458	8,458	8,458	8,458

Note: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. IHDS data. Married women of age 15 to 49 in nuclear families are included in the sample. HSAA is equal to one if woman i got married after the amendment in state s and is Hindu, Buddhist, Sikh or Jain. Distance from natal family is in hours of travel. Individual and household controls defined as in table 1. Robust standard errors in parentheses. Standard errors clustered at state-religion level.

Marital Sorting. The HSA amendments might have changed sorting in the marriage market, too. Women affected by the amendments, for example, may be able to find better spouses and, hence, to have better health outcomes. Once again, this may cast doubt on the validity of our identification strategy. We study the effect of the HSA reforms on the education age gap between husband and wife, the likelihood of arranged marriages and caste exogamy, women's age at marriage, their

likelihood to marry up (i.e., to marry someone of a higher socio-economic status),[2] and on the extent of their migration upon marriage. As shown in Table A5, we do not find significant changes in marital sorting or marriage characteristics after the introduction of the amendments (with the exception of an increase in women's age at marriage). Importantly, all our results presented in Section 7.3 are robust to including women's age at marriage as an additional covariate.

It is important to note that marriages in India are characterized by religious endogamy and rarely occur across states. The HSA amendments applied to all Hindu, Buddhist, Jain, and Sikh women who were unmarried at the time of the amendment in their state. As a consequence, within a marriage market either all women were exposed to the reforms or none was. In addition, since marriage is nearly universal in India, it is unlikely that the reforms impacted one's decision of whether or not to get married. For these reasons, the influence of the HSA reforms on marital sorting may be limited.

Table A5: HSA Amendments and Marital Sorting

	Educ. Gap	Pr(Marry Up)	Pr(Same Caste)	Pr(Arranged Marriage)	Dist. from Natal Fam.	Pr(Same Village)	Age at Marriage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HSAA	-0.119	-0.0129	0.0154	0.00459	-0.0273	-0.0581	4.602***
	(0.275)	(0.0167)	(0.0117)	(0.0124)	(0.463)	(0.0377)	(0.313)
N	8,449	8,038	8,429	8,389	8,379	8,424	8,458

Note: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. IHDS data. Married women of age 15 to 49 in nuclear families are included in the sample. HSAA is equal to one if woman i got married after the amendment in state s and is Hindu, Buddhist, Sikh or Jain. Distance from natal family is in hours of travel. Individual and household controls defined as in table 1. Robust standard errors in parentheses. Standard errors clustered at state-religion level.

---

[2]The IHDS directly asks *"At the time of your marriage, if you compared the economic status of your natal family with your husband's family, would you say your natal family was... (i) same, (ii) better off, (iii) worse off?"* Based on the answer to this question, I construct an indicator variable equal to one if the answer is (ii) and zero otherwise.

C   Additional Tables

Table A6: $\left|\text{Bias}_{\text{B-LATE}}\right|$

Panel A: $p_1^a = 0.9, p_0^b = 0.9$					
$p_1^b \downarrow \quad \mid \quad p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	0.111	0.117	0.143	0.176	0.250
0.01	0.117	0.124	0.149	0.183	0.258
0.05	0.143	0.149	0.176	0.212	0.290
0.1	0.176	0.183	0.212	0.250	0.333
0.2	0.250	0.258	0.290	0.333	0.429

Panel B: $p_1^a = 0.8, p_0^b = 0.8$					
$p_1^b \downarrow \quad \mid \quad p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	0.250	0.258	0.290	0.333	0.429
0.01	0.258	0.266	0.299	0.342	0.439
0.05	0.290	0.299	0.333	0.379	0.481
0.1	0.333	0.342	0.379	0.429	0.538
0.2	0.429	0.439	0.481	0.538	0.667

Panel C: $p_1^a = 0.7, p_0^b = 0.7$					
$p_1^b \downarrow \quad \mid \quad p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	0.429	0.439	0.481	0.538	0.667
0.01	0.439	0.449	0.493	0.550	0.681
0.05	0.481	0.493	0.538	0.600	0.739
0.1	0.538	0.550	0.600	0.667	0.818
0.2	0.667	0.681	0.739	0.818	1.000

<u>Notes</u>: Results obtained setting $r = 0.5$. Each cell reports $\left|\text{Bias}_{\text{B-LATE}}\right|$ under different values of $p_1^a, p_0^a, p_1^b, p_0^b$. The true LATE is normalized to 1.

Table A7: $|\text{Bias}_{\text{MR-LATE}}|$

Panel A: $p_1^a = 0.9$, $p_0^b = 0.9$					
$p_1^b \downarrow$	$p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	**0.000**	0.011	0.059	0.125	0.286
0.01	0.011	0.022	0.070	0.136	0.297
0.05	0.059	0.070	0.118	0.184	0.345
0.1	0.125	0.136	0.184	0.250	0.411
0.2	0.286	0.297	0.345	0.411	0.571

Panel B: $p_1^a = 0.8$, $p_0^b = 0.8$					
$p_1^b \downarrow$	$p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	**0.000**	0.013	0.067	0.143	0.333
0.01	0.013	0.025	0.079	0.156	0.346
0.05	0.067	0.079	0.133	0.210	0.400
0.1	0.143	0.156	0.210	0.286	0.476
0.2	0.333	0.346	0.400	0.476	0.667

Panel C: $p_1^a = 0.7$, $p_0^b = 0.7$					
$p_1^b \downarrow$	$p_0^a \rightarrow$	0	0.01	0.05	0.1	0.2
0	**0.000**	0.014	0.077	0.167	0.400
0.01	0.014	0.029	0.091	0.181	0.414
0.05	0.077	0.091	0.154	0.244	0.477
0.1	0.167	0.181	0.244	0.333	0.567
0.2	0.400	0.414	0.477	0.567	0.800

Notes: Results obtained setting $r = 0.5$. Each cell reports $|\text{Bias}_{\text{MR-LATE}}|$ under different values of $p_1^a, p_0^a, p_1^b, p_0^b$. Cells are empty if the bias is not finite. The true LATE is normalized to 1.

Table A8: Bounds κ^a and κ^b

	Women		Men		Children	
	κ^a	κ^b	κ^a	κ^b	κ^a	κ^b
$\mathcal{K} = 0$	50.00	50.00	50.00	50.00	50.00	50.00
$\mathcal{K} = 1$	50.21	49.79	50.21	49.77	50.21	49.79
$\mathcal{K} = 5$	51.00	49.05	50.99	49.01	50.89	49.13
$\mathcal{K} = 10$	52.00	48.13	51.98	48.13	51.75	48.36

Note: NFHS data.

Table A9: First Stage Estimates

	Women		Men		Children	
	T^a	T^b	T^a	T^b	T^a	T^b
	Panel A: $\mathcal{K} = 0$					
$\mathbb{I}$(HSA)	0.0867	-0.0867	0.0779	-0.0779	0.0930	-0.0930
	(0.0114)	(0.0114)	(0.0126)	(0.0126)	(0.0181)	(0.0181)
First Stage F-stat.	76.7245	76.7245	48.0541	48.0541	33.3953	33.3953
	Panel B: $\mathcal{K} = 1$					
$\mathbb{I}$(HSA)	0.0756	-0.0876	0.0681	-0.0775	0.0755	-0.0975
	(0.0111)	(0.0114)	(0.0124)	(0.0126)	(0.0177)	(0.0182)
First Stage F-stat.	58.6617	78.5649	37.2418	47.7399	22.5004	34.2528
	Panel C: $\mathcal{K} = 5$					
$\mathbb{I}$(HSA)	0.0581	-0.1078	0.0560	-0.0998	0.0588	-0.1320
	(0.0106)	(0.0120)	(0.0119)	(0.0133)	(0.0167)	(0.0189)
First Stage F-stat.	34.3853	117.3670	25.5154	77.5716	15.2348	62.0974
	Panel D: $\mathcal{K} = 10$					
$\mathbb{I}$(HSA)	0.0341	-0.1203	0.0349	-0.1059	0.0356	-0.1520
	(0.0100)	(0.0128)	(0.0110)	(0.0143)	(0.0162)	(0.0204)
First Stage F-stat.	11.5444	144.9936	9.8665	84.2524	7.6620	80.8510

Note: NFHS data. Robust standard errors in parenthesis. All specifications include individuals and household controls, state-religion fixed effects, mother's cohort-religion fixed effects and state specific time trends (up to degree four).

Table A10: Engel Curves Estimation Results (NSS Sample)

	$R(X)$	$\widehat{\delta}_w(X)$	$\widehat{\delta}_m(X)$	$\widehat{\beta}(X)$
$\mathbb{1}$(1 child)	0.0692	0.000550	-0.244	0.0299
	(0.0228)	(0.476)	(0.479)	(0.0393)
$\mathbb{1}$(2 children)	0.0250	-0.0777	-0.242	-0.0223
	(0.0210)	(0.477)	(0.479)	(0.0585)
$\mathbb{1}$(3 children)	0.0404	-0.451	-0.452	0.00759
	(0.0242)	(0.513)	(0.511)	(0.0596)
Fraction of Female Children	0.00217	-0.319	-0.351	0.0379
	(0.0177)	(0.343)	(0.341)	(0.0416)
Gender Age Gap (Man - Woman)	0.0298	2.552	1.998	-0.221
	(0.143)	(2.560)	(2.707)	(0.327)
Woman's Age	-0.452	1.865	1.519	-0.125
	(0.123)	(2.899)	(2.901)	(0.355)
Children's Avg. Age	-0.277	0.511	0.881	-0.206
	(0.237)	(4.846)	(4.822)	(0.587)
$\mathbb{1}$(Hindu, Buddhist, Jain, Sikh)	0.104	1.260	0.803	-0.113
	(0.0193)	(0.341)	(0.345)	(0.0413)
$\mathbb{1}$(Sch. Caste, Sch. Tribe, Oth. Back. Caste)	-0.0209	0.0666	0.103	-0.0310
	(0.0158)	(0.278)	(0.282)	(0.0338)
$\mathbb{1}$(Own Land)	-0.000975	0.00307	0.0821	0.0135
	(0.0165)	(0.310)	(0.308)	(0.0372)
$\mathbb{1}$(Woman Completed High School)	0.0550	-0.390	-0.388	0.0536
	(0.0263)	(0.488)	(0.489)	(0.0566)
$\mathbb{1}$(Man Completed High School)	0.0437	0.113	-0.0934	0.0139
	(0.0210)	(0.401)	(0.404)	(0.0474)
$\mathbb{1}$(Rural)	-0.00804	1.202	1.409	-0.152
	(0.0155)	(0.307)	(0.306)	(0.0373)
$\mathbb{1}$(North)	-0.0772	0.00410	0.814	-0.0253
	(0.0257)	(0.511)	(0.510)	(0.0615)
$\mathbb{1}$(East)	0.0896	-0.129	-0.445	0.0431
	(0.0267)	(0.529)	(0.518)	(0.0638)
$\mathbb{1}$(North-East)	0.197	-1.668	-2.212	0.168
	(0.0311)	(0.575)	(0.558)	(0.0686)
$\mathbb{1}$(South)	-0.0586	1.026	0.746	-0.0817
	(0.0253)	(0.531)	(0.537)	(0.0648)
Constant	0.494	6.723	7.065	-0.699
	(0.0503)	(1.031)	(1.031)	(0.113)

Note: NSS data. Robust standard errors in parenthesis. Age variables are are divided by 100 to ease computation. Number of children excluded category is 4 children and above.

References

ATTANASIO, O. P. AND V. LECHENE (2014): "Efficient responses to targeted cash transfers," *Journal of political Economy*, 122, 178–222. [11]

BARGAIN, O. AND O. DONNI (2012): "Expenditure on Children: A Rothbarth-type Method Consistent with Scale Economies and Parents' Bargaining," *European Economic Review*, 56, 792–813. [5]

BARGAIN, O., O. DONNI, AND P. KWENDA (2014): "Intrahousehold distribution and poverty: Evidence from Cote d'Ivoire," *Journal of Development Economics*, 107, 262–276. [5]

BROWNING, M., P.-A. CHIAPPORI, AND A. LEWBEL (2013): "Estimating Consumption Economies of Scale, Adult Equivalence Scales, and Household Bargaining Power," *Review of Economic Studies*, 80, 1267–1303. [3]

CALVI, R. (2019): "Why Are Older Women Missing in India? The Age Profile of Bargaining Power and Poverty," *Journal of Political Economy*, Forthcoming. [11]

CHERCHYE, L., B. DE ROCK, A. LEWBEL, AND F. VERMEULEN (2015): "Sharing Rule Identification for General Collective Consumption Models," *Econometrica*, 83, 2001–2041. [5]

DUNBAR, G. R., A. LEWBEL, AND K. PENDAKUR (2013): "Children's Resources in Collective Households: Identification, Estimation, and an Application to Child Poverty in Malawi," *American Economic Review*, 103, 438–71. [3]

LEWBEL, A. AND K. PENDAKUR (2008): "Estimation of Collective Household Models with Engel Curves," *Journal of Econometrics*, 147, 350–358. [5]

LISE, J. AND S. SEITZ (2011): "Consumption Inequality and Intra-household Allocations," *Review of Economic Studies*, 78, 328–355. [5]

MENON, M., K. PENDAKUR, AND F. PERALI (2012): "On the Expenditure-dependence of Children's Resource Shares," *Economics Letters*, 117, 739–742. [5]

TOMMASI, D. (2019): "Control of resources, bargaining power and the demand of food: Evidence from PROGRESA," *Journal of Economic Behavior & Organization*, 161, 265–286. [11]

TOMMASI, D. AND A. WOLF (2018): "Estimating household resource shares: A shrinkage approach," *Economics Letters*, 163, 75 – 78. [5]