Women's Empowerment and Family Health: Estimating LATE with Mismeasured Treatment

Rossella Calvi*[1], Arthur Lewbel†[2], and Denni Tommasi‡[3]

[1]Department of Economics, Rice University

[2]Department of Economics, Boston College

[3]Department of Econometrics and Business Statistics, Monash University

December 2019

Abstract

We study the impact of women's decision making power on family health in India. We define treatment as a woman having primary control over household resources, using changes in inheritance laws as an instrument. Due to measurement difficulties and sharing of goods, treatment cannot be directly observed and must be estimated using a structural model. Treatment mismeasurement may therefore arise from model misspecification and estimation errors. We provide a new estimation method, MR-LATE, that under some conditions can consistently estimate local average treatment effects when treatment is mismeasured. When such conditions are not satisfied, MR-LATE is still effective at reducing estimation bias. We find that women's control of household resources improves their and their children's health at no cost to men's health. We also show that accounting for measurement error in the estimated treatment is empirically important.

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

Keywords: causality, LATE, structural model, collective model, resource shares, bargaining power, health.

*Backer Hall 274, 6100 Main Street, Houston, TX 77005, USA (e-mail: rossella.calvi@rice.edu).

†Maloney Hall 315, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA (e-mail: arthur.lewbel@bc.edu).

‡Level 5, Building H, Caulfield Campus 900 Dandenong Road Melbourne, VIC 3145 (e-mail: denni.tommasi@monash.edu).

We thank Samson Alva, Erich Battistin, Jonathan de Quidt, Bram De Rock, Laurens Cherchye, Flávio Cunha, Frank DiTraglia, Yingying Dong, Jeremy Fox, Michael Gechter, Mette Gørtz, Rachel Heath, Nikolaj Harmon, Søren Leth-Petersen, Benjamin Solow, Takuya Ura, Kenneth Wolpin, Yi Zhang, seminar participants at University of Copenhagen, ECARES, Rice University, Tilburg University, Stockholm University, Penn State, UT Austin, and participants at the Winter Meeting of the Econometric Society, NEUDC, SEA Conference, the Annual Conference on Economic Growth and Development at ISI-Delhi, the Texas Econometrics Camp, and the Economic Demography Workshop for their suggestions. All errors are our own.

1 Introduction

This paper has two goals. One is the estimation of a causal Local Average Treatment Effect (LATE) of women's decision making power regarding household resources on family health outcomes in India. Our second goal is to introduce a new estimator that can reduce and in some cases eliminate the bias in LATE estimation due to treatment mismeasurement.

There exists a large literature stressing the importance of female intra-household decision making power in developing countries. Examples are seminal work by Thomas (1990; 1994; 1997), based on related issues raised by Becker (1965; 1974; 1981) and by Sen (1983; 1988; 1989). Numerous studies document that income or assets accruing to women are more likely than those of men to be allocated to expenditures that benefit children as well as themselves.[1] Actual control over household expenditure, however, is difficult to observe because expenditure data is typically collected at the household level and goods can be shared. As a consequence, most of these studies focus on estimating the effect of randomized treatments or of other proxies that happen to be observed and are believed to increase women's decision making power.[2] Contrary to previous work, we directly study the impact of women's control over a substantial share of household resources on the health outcomes of the members of Indian families. We estimate treatment using a structural model of intra-household resource allocation and use a plausibly exogenous change in inheritance laws as an instrument.

A typical causal analysis might look at the impact of the change in the law itself on health outcomes, but this would tell us nothing about how other changes in women's control over resources might impact health. In contrast, a typical structural analysis of this problem would require not only modeling the intra-household resource allocation process, but also specifying how inheritance laws affect intra-household allocations and how control over household resources affects health. Many might find such models implausible. This is an example of the commonly noted trade-off that analyses based on natural experiments often answer questions that are possibly limited in scope, while structural analyses depend on strong assumptions regarding the underlying true behavior. In our application, we seek to combine the best of both worlds. We know little about how women's control over household expenditure may affect the health outcomes of household members, and so we address that question using LATE estimation. A great deal more is known, both theoretically and empirically, about the economics of household resource allocation. We exploit this knowledge and a structural modeling approach to estimate the (otherwise unobserved) treatment status.

A drawback of this general procedure is that, due to estimation errors (which include possible model misspecification), our estimated treatment indicator is likely to be mismeasured for some households. More generally, even when treatment is observed rather than estimated, it may sometimes be mismeasured due, e.g., to reporting or contamination errors, or to people who for whatever reason choose not to take the treatment they were assigned. To deal with these issues, we propose a new estimation method, which we call MR-LATE (for Mismeasurement Robust LATE), that can identify and consistently estimate LATE even when the endogenous binary treatment indicator contains measurement errors. Unlike Battistin et al. (2014) or DiTraglia and García-Jimeno (2019), MR-LATE does not require re-survey data or homogeneity of treatment effects. Unlike Ura (2018), who obtains bounds, we (under certain conditions) point identify and estimate LATE. The conditions under which MR-LATE consistently estimates LATE are strong, but even when they are not satisfied, MR-LATE can still substantially reduce the bias associated with measurement

[1] See, e.g., Haddad et al. (1997); Duflo (2003); Quisumbing and Maluccio (2003); Smith (2003); Rubalcava et al. (2009); LaFave and Thomas (2017).

[2] See section 2 for more details.

errors, and therefore represents a preferred alternative to traditional LATE estimators in the presence of mismeasurement (Imbens and Angrist (1994)).

We apply MR-LATE to account for the inevitable estimation and specification errors that arise from using a structural intra-household allocation model to estimate treatment. The end result is that instead of asking what the health impact of a particular policy intervention is (e.g., the inheritance law change), we can address a more general question: What is the impact of empowering women (by giving them control over household resources) on family health? Taking the standard caveats about the external validity of LATE into account, the answer will then allow us to assess the potential impacts of a wide variety of policy interventions that increase women's control over household resources.

Our analysis of treatment relies on the *collective household* model (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. The model is used to structurally estimate a measure of individual-level resource control from the observation of household-level consumption expenditures. Specifically, we estimate men's and women's *resource shares* (i.e., the fraction of a household's resources allocated to each decision maker) using a methodology developed in Dunbar et al. (2013), applied to household expenditure data from the 2005-2006 National Sample Survey (NSS) of Consumer Expenditure in India. We then define a household to be treated if the woman has control over a substantial share of the household's resources.[3]

Due to the lack of NSS data on health outcomes, we use the structural estimates to perform an out-of-sample prediction on the 2005-2006 Indian National Family Health Survey (NFHS), which includes the same socio-economic characteristics of individuals and households as the NSS dataset as well as detailed information about women's, men's and children's health status. Using our MR-LATE estimator (which can also account for any additional misclassification error introduced by the out-of-sample prediction exercise), we then study the causal impact of living in families where women have substantial control over household expenditures on the health status of family members. We exploit changes to the Indian inheritance law and NFHS information on women's religion, year of marriage, and state of residence to construct our instrument. In particular, we focus on women's exposure to the Hindu Succession Act amendments that equalized women's inheritance rights to men's in several Indian states between 1976 and 2005. In order to benefit from these reforms, a woman needed to be Hindu, Buddhist, Jain or Sikh, and unmarried at the time of the reform in her state.

We find that accounting for specification, estimation, and/or measurement error in the estimate of treatment is empirically important, with some substantial differences between the standard 2SLS estimator for LATE (Imbens and Angrist (1994)), which cannot control for such measurement errors, and our MR-LATE estimation method. The results of our empirical analysis indicate that women's control over resources positively affects women's and children's health outcomes at no cost to men's health outcomes. The estimated effects are sizable. Our most conservative estimates indicate that, for compliers, the average treatment effect on women's body mass index is 7.7 and that women in treated households are 72 percentage points less likely to be underweight and 52 p.p. less likely to be anemic. Also, mothers' control over resources substantially reduces the occurrence of cough, fever and diarrhea in children, and mildly improves their height-for-age and weight-for-age z-scores. Our MR-LATE estimates indicate that living in households where the mother controls resources reduces children's likelihood to be sick with fever by 43 to 65 percentage points, with cough by 66 to 89 p.p., and with diarrhea by 45 to 65 p.p., depending on

[3]For simplicity, our analysis focuses on nuclear households with children. As explained in section 3.3, the relevant threshold (that is, what defines a *substantial* share of household resources) can vary across households for unobserved reasons.

the degree of misclassification error in the treatment variable. Our structural model of treatment indicates that women control about 60 percent of resources in households where they do control substantial resources, and about 40 percent when they do not. These large differences in resource control may help understand why our estimated treatment effects are so sizable.

Beyond our particular application, MR-LATE should be generally useful for many applications where treatment is observed with error (e.g., when there are reporting, recording, or contamination errors). However, we also want to emphasize its potential for applications like ours in which the treatment itself is 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 and so must be estimated. 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 effects of risk aversion on investment decisions, or the effect of abilities such as non-cognitive skills on future earnings. Our paper is intended to provide a practical method of implementing causal analyses in many such settings, where treatment cannot be directly observed and so must be estimated.

The remainder of the paper is organized as follows. The next section is a literature review. This is followed by our identification results and the derivation of our associated MR-LATE estimator, including a Monte-Carlo study of its properties. We then implement our study of women's intra-household power on family health outcomes, reporting the results of both our structural model of treatment and its causal effects. These are followed by some concluding remarks. Proofs and additional material are in an Appendix.

2 Literature Review

This paper relates to three main strands of literature: the previous research on women's empowerment and its consequences for family outcomes in developing countries, the existing studies on identification and estimation of treatment effects in presence of mismeasured treatment, and recent work highlighting the advantages of combining structural and causal approaches.

Women's Empowerment and Control over Resources. Our structural treatment model is based on the collective household framework pioneered by Becker (1965, 1981), Apps and Rees (1988) and Chiappori (1988, 1992). In these models, 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 (2019), 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.[4] Partly due to data limitations, in our application we focus on estimating resource shares for men and women only, as in Lewbel and Pendakur (2008) and Browning et al. (2013), while treating children as public goods (see, e.g., Blundell et al. (2005)). Consequently, in our setting resource shares are not measures of individual consumption, but of control over resources and decision power.

[4]Applications or extensions of this approach include Penglase (2017), Sokullu and Valente (2017), Bargain et al. (2018), Brown et al. (2018), Tommasi (2019).

From a policy perspective, our results contribute to the general literature on estimating the effect of women's empowerment on adults' and children's outcomes. An application with a similar motivation to ours is Maitra (2004), who uses relative educational attainment of the parents and self-reported measures of autonomy and decision power to show that a woman's control over household resources (her ability to keep money aside) has a significant effect on health care usage and child mortality in Indian households. More broadly, 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., Schultz (1990)). 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.

Legal reforms aimed at improving women's property or inheritance rights have been also used to assess the effects of changes in bargaining power within the household.[5] In the Indian context, Deininger et al. (2013) find evidence of an increase of women's likelihood of inheriting land following the introduction of Hindu Succession Act (HSA) amendments that equalized women's inheritance rights to men's in several Indian states between 1976 and 2005. Roy (2008) documents that women's exposure to the HSA reforms improved women's autonomy within their marital families. Deininger et al. (2013), Roy (2015) and Bose and Das (2015) find that it increased female education, while Heath and Tan (2019) claim that it increased women's labor supply, especially into high-paying jobs. Calvi (2019) shows that women's exposure to the HSA reforms led to improved health outcomes and increased access to household resources.[6] Departing from this literature, our analysis uses the HSA reforms as the instrument for LATE estimation, where an imperfect measure of treatment is obtained by estimation of a structural model of control over household resources.

LATE and Misclassification Error. Papers empirically documenting substantial measurement (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). In our application, these measurement errors in treatment come (in part) from treatment being structurally estimated, but as these references show, measurement errors are also common even in applications where treatment is directly observed.

A few previous papers have considered alternative techniques for dealing with such measurement errors in treatment. Homogeneous treatment effects, corresponding to 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 mismeasured, point-identification and associated estimators of average treatment effects (without

[5]Parallel to these studies, an extensive literature studying the effects of Conditional Cash Transfer programs in developing countries has shown that providing women with a large amount of cash in their hands can significantly increase, e.g., the budget shares of expenditures on clothing for children and lower shares of alcohol (Attanasio and Lechene (2002)), increase children's health and education, and livestock (Rubalcava et al. (2009)) and improve child development (Macours et al. (2012)).

[6]Calvi (2019) also shows that such amendments do not substantially increase total household expenditure or the characteristics of marital matchings. She finds, however, that exposure to HSA amendments increases food budget shares, reduces the share of resources spent on alcohol and tobacco, and increases the share of household budget devoted to children's clothing. These results provide an empirical rejection of the unitary model (Attanasio and Lechene (2002)), and support the hypothesis that the inheritance reforms changed the bargaining power of women. These also provide suggestive evidence in favor of the validity of the exclusion restriction in our setting.

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 measurement 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 (2008), Imai et al. (2010), and Kreider et al. (2012).

The causal effect we focus on identifying and estimating is the local average treatment effect (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. Identification of LATE with misclassified treatment has recently received some attention. Ura (2018), for example, considers estimation of LATE with mismeasured treatment and standard LATE instrument assumptions, but only obtains set identification bounds.[7]

Our MR-LATE approach makes use of two rather than a single misclassified treatment indicator. Battistin et al. (2014) also uses two measures of the misclassified treatment to identify LATE, but they require re-survey data (that is, multiple observations of the same individuals), which are often not available. 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 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.

Our estimation problem has the standard LATE structure that a randomized binary instrument is correlated with a binary treatment indicator, and the true treatment affects an outcome. But in our case we must overcome the issue that 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 satisfy 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.

Structural vs. Causal Analysis. This paper contributes more broadly to the long-standing debate on the relative benefits and limitations of structural modeling vs. reduced form, randomization-based

[7]Specifically, the instrument must satisfy the exclusion restriction and weakly increase the true treatment (Imbens and Angrist (1994); Angrist et al. (1996)). In the case of continuously distributed mismeasured treatment, Lewbel (1998), Song et al. (2015), Hu et al. (2015) and Song (2015) use instruments and further exclusion restrictions to obtain identification and estimation of average marginal effects with classical or nonclassical measurement error in a nonparametric or semiparametric context.

or quasi-experimental analyses. Proponents of randomization question the validity of results obtained from complex structural modeling assumptions (Angrist and Pischke (2010); Imbens (2010)). Advocates of structural models stress the insights that can be obtained when one allows economic theory to guide the empirical work (Wolpin (2013)). Recent contributions in the econometrics literature have started to formally unify the two camps in order to overcome these divisions (Vytlacil (2002), Heckman et al. (2006); Heckman and Vytlacil (2007); Pearl (2009)). Heckman (2010), for example, proposes to combine the best features of both the structural and the causal modeling approaches in what he calls a *third way* of policy analysis. Similarly, Lewbel (2019) argues that combining the strengths of both approaches can often be the best strategy for identification. We contribute to this literature by using structure provided by economic theory to significantly expand the set of causal questions that researchers can answer, when combined with statistical techniques (like our MR-LATE estimator) that account for the inevitable estimation and specification errors of structural models.

3 LATE With Mismeasured or Misspecified Treatment

This section is organized in four parts. First, we describe our theoretical framework and derive the new MR-LATE estimator. Second, we discuss the use of MR-LATE as a bias-reduction method. Third, we provide examples to illustrate the use of our new estimation method. Finally, we carry out a series of Monte Carlo experiments designed to evaluate the properties of MR-LATE.

3.1 Setup and Identification

We first introduce some notation, ignoring additional covariates for now, as everything immediately extends to conditioning on covariates X.

Let D be the true binary treatment variable that affects an outcome of interest. 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. 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,

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

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,

$$Y = (1 - D) Y_0 + D Y_1 = Y_0 + (Y_1 - Y_0) [(1 - Z) D_0 + Z D_1]. \tag{2}$$

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 $D_1 = 1$. 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:

$$\frac{cov(Y,Z)}{cov(D,Z)} = E(Y_1 - Y_0 \mid C) = \text{LATE}. \tag{3}$$

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 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). In our empirical application, for instance, T will be an estimate of D that is based on a structural model. Thus, in our application, T will not equal D for some individuals either because of estimation error or because the structural model may be misspecified. Our goal is to identify and estimate LATE, even though D is not directly observed and cannot be consistently estimated.[8]

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,

$$T = (1-D)T_0 + DT_1. \tag{4}$$

The variables T_0 and T_1 can be interpreted as indicators of whether treatment is correctly measured 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: if $T_0 = 1$, then a true $D = 0$ is misclassified as treated, and if $T_1 = 0$, then a true $D = 1$ is misclassified as untreated. Since there exists a negative correlation between the true treatment and measurement error, the measurement error is non-classical.

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. $E(T_1 - T_0 \mid C) \neq 0$.

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). Combined with unconfoundedness, this corresponds to the standard assumption that measurement errors are unrelated to outcomes.[9] Finally, Assumption 2-iii is a minimal relevance condition saying that, at least for compliers, T provides some information regarding D. This assumption implies that, 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 true treatment $D = 0$. Note that Assumption 2-iii ensures that $p_1 - p_0$ is nonzero.

[8]Although we observe T and not D, individual 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.

[9]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 the potential outcomes and potential treatments. If this stronger assumption holds then q defined below satisifes $q = E(T_0)/E(T_1 - T_0)$. We do not require this plausible but stronger condition.

Define q and λ as follows

$$q = \frac{p_1}{p_1 - p_0} \tag{5}$$

$$\lambda = \frac{cov(YT, Z)}{cov(T, Z)} = \frac{E(YT \mid Z=1) - E(YT \mid Z=0)}{E(T \mid Z=1) - E(T \mid Z=0)}. \tag{6}$$

If one were to ignore measurement error in T, one 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 asymptotic value of an instrumental variables regression of YT on T using Z as the instrument.[10] 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:*

$$\lambda = E[qY_1 + (1-q)Y_0 \mid C]. \tag{7}$$

Proof. See Appendix A.1. □

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 $p_d^a = E\left(T_d^a \mid C\right)$ and $p_d^b = E\left(T_d^b \mid C\right)$, 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, ρ, and our MR-LATE estimand as follows:

$$q^a = \frac{p_1^a}{p_1^a - p_0^a}, \quad q^b = \frac{p_1^b}{p_1^b - p_0^b},$$

$$\lambda^a = \frac{cov(T^aY, Z)}{cov(T^a, Z)}, \quad \lambda^b = \frac{cov\left(T^bY, Z\right)}{cov(T^b, Z)}, \text{ and}$$

$$\text{MR-LATE} = \rho = \lambda^a - \lambda^b.$$

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:*

$$\textit{MR-LATE} = \left(q^a - q^b\right)E[Y_1 - Y_0 \mid C] = \left(q^a - q^b\right)\textit{LATE}, \tag{8}$$

where $LATE = cov(Y, Z)/cov(D, Z)$.

Note that $LATE$ in Corollary 1 refers to the true LATE, which in our application cannot be estimated using Imbens and Angrist (1994), because D is not observed. Corollary 1 has some straightforward implications, one of which is the following:

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

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 MR-LATE $= cov(Y,Z)/cov(D,Z)$, which equals the true $LATE$. A sufficient condition for MR-LATE to equal the true LATE is $p_0^a = p_1^b = 0$.*

The fact that MR-LATE equals the true LATE when $p_0^a = p_1^b = 0$ follows from equations (5) and (8). More generally, MR-LATE provides a good approximation to LATE when p_0^a and p_1^b are close to zero. Having $p_0^a = 0$ means that, among compliers, the probability that $T_0^a = 1$ is zero. In other words, all compliers who have $D = 0$ must also have $T^a = 0$. Therefore, p_0^a equals zero if, among compliers, the treatment measure T^a has only one kind of measurement error, never mistaking the actually untreated as treated (even if it often mistakes the actually treated as untreated). Similarly, having $p_1^b = 0$ requires that all compliers who have $D = 1$ must also have $T^b = 0$. This means that, among compliers, $1 - T^b$ never mistakes the actually treated as untreated (even if it is frequently wrong about mistaking the actually untreated as treated).[11]

The strong restrictions that $p_0^a = p_1^b = 0$ can also be thought of as rough analogs to the standard no defiers assumption, which assumes a zero probability of certain combinations of D and Z realizations. Analogously, the sufficient condition introduced in Corollary 2 requires a zero probability of certain combinations of D and T^a realizations and of D and T^b realizations. Like the probability of defiers, the probabilities p_0^a and p_0^b are not identified without additional assumptions.

A simple summary is as follows: MR-LATE equals the true LATE if, among compliers, when D is zero then T^a is zero, and when D is one then T^b is zero. More generally, p_0^a and p_1^b will be close to zero, making MR-LATE close to the true LATE (meaning $q^a - q^b$ is close to one), if T^a is rarely one when D is zero, and if T^b is rarely one when D is one. This is 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. See the next subsection for more discussion of this point.

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 $q^a - q^b > 0$, MR-LATE signs LATE. If $q^a - q^b \geq 1$, LATE lies between 0 and MR-LATE. A sufficient condition for the identified set 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.

The application of Corollaries 2 and 3 requires having information regarding misclassification probabilities that may be difficult to obtain in practice. Next, we consider more general conditions under which MR-LATE reduces estimation bias.

3.2 MR-LATE for Bias Reduction

Since the conditions under which MR-LATE equals or bounds the true LATE may be difficult to confirm, here we consider a more 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 mea-

[11] It is standard in the measurement error literature to construct measures that are positively correlated with the true object they are measuring (in this case, D). In that sense we should think of $1 - T^b$, not T^b, as being a measure of D. Note, however, that this is just an issue of notation rather than economic content, since T^b contains the exact same information as $1 - T^b$.

surement 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 we show similar results in finite samples.

Let B-LATE (for Biased LATE) denote the standard LATE estimator when we observe a mismeasured treatment indicator T in place of the true unobserved treatment indicator D. B-LATE is the instrumental variables regression of Y on T using Z as the instrument. It follows from the proofs of Theorem 1 and Corollary 1 that

$$\begin{aligned}\text{B-LATE} &= \frac{cov(Y,Z)}{cov(T,Z)} = \frac{E(Y \mid Z=1) - E(Y \mid Z=0)}{E(T \mid Z=1) - E(T \mid Z=0)} \\ &= \frac{E\left[(Y_1 - Y_0)(D_1 - D_0)\right]}{(p_1 - p_0)\, E(D_1 - D_0)} \\ &= \frac{1}{p_1 - p_0}\text{LATE}.\end{aligned} \tag{9}$$

This shows that the bias in B-LATE is $\frac{1}{p_1-p_0} - 1$ times the true LATE. 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, 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 $1-r$.[12] The B-LATE bias then becomes

$$\text{Bias}_{\text{B-LATE}} = \left[\frac{1}{\left(p_1^a - p_0^a\right) r - \left(p_1^b - p_0^b\right)(1-r)} - 1\right]\text{LATE}. \tag{10}$$

From Theorem 1, the bias in MR-LATE is $(q^a - q^b) - 1$ times the true LATE, which equals:

$$\text{Bias}_{\text{MR-LATE}} = \left[\frac{p_0^a}{p_1^a - p_0^a} - \frac{p_1^b}{p_1^b - p_0^b}\right]\text{LATE}. \tag{11}$$

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 given above will be relatively small. In contrast, the bias in B-LATE is small only if the r weighted average of $\left(p_1^a - p_0^a\right)$ and $\left(p_0^b - p_1^b\right)$ 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.

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*

$$p_0^a + p_1^b < \left(\frac{1}{\max\{p_1^a - p_0^a,\, p_0^b - p_1^b\}} - 1\right)\min\{p_1^a - p_0^a,\, p_0^b - p_1^b\} \tag{12}$$

then $|Bias_{MR\text{-}LATE}| < |Bias_{B\text{-}LATE}|$.

To prove this corollary, observe that the expression in the brackets in equation (11) 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 (10) 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. Corollary 4 shows

[12] 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.

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.

that, as long as the probabilities p_0^a and p_1^b are sufficiently small (noting that T^a and T^b are supposed to make these probabilities small), MR-LATE will have smaller bias than B-LATE. Moreover, this result holds no matter what combination of T^a and T^b (i.e., no matter value of r) we use to construct the B-LATE treatment measure T.

To more easily interpret the Corollary, consider the symmetric case where $p_0^a = p_1^b$ and $p_1^a = p_0^b$. In that case, the sufficient condition in Corollary 4 reduces to $p_0^a + p_1^a < 1$ and $p_0^b + p_1^b < 1$, which are very mild restrictions. Satisfying Corollary 4 is similarly mild.

To further illustrate the point, and to assess the magnitude of the usual 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 (10)) and MR-LATE (from equation (11)) for $r = 0.5$.[13] 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, 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 A1 and A2 in the Appendix show the magnitudes of the respective biases that are differenced in

[13]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, 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. We later show similar gains in finite sample bias reduction.

3.3 The MR-LATE Estimator and Examples

The numerical computation of the MR-LATE estimator is extremely easy. Assume we observe the vector $\left(Y_i, Z_i, T_i^a, T_i^b\right)$ for individuals $i = 1, ..., n$. Consider a linear instrumental variables regression of $Y_i T_i^a$ on a constant and on T_i^a, using as instruments a constant and Z_i. Let the estimated coefficient of T_i^a in this regression be $\widehat{\lambda}^a$. Note that $\widehat{\lambda}^a$ does not equal the ordinary Imbens and Angrist (1994) LATE estimator, since $\widehat{\lambda}^a$ regresses $Y_i T_i^a$ on T_i^a instead of regressing Y_i on T_i^a (as noted earlier, Abadie (2002), Ura (2018) among others have similar constructions, though used differently).

Similarly, let $\widehat{\lambda}^b$ be the estimated coefficient of T_i^b in a linear instrumental variables regression of $Y_i T_i^b$ on a constant and on T_i^b, again using as instruments a constant and Z_i. The MR-LATE estimator is then given by $\widehat{\rho} = \widehat{\lambda}^a - \widehat{\lambda}^b$. Equivalently,

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

where, for any X, $\widehat{cov}(X, Z) = \sum_{i=1}^n X_i (Z_i - \overline{Z})/n$ and $\overline{Z} = \sum_{i=1}^n Z_i/n$.

With our identifying assumptions, consistency of $\widehat{\rho}$ follows as long as we can apply a law of large numbers to each expectation in these regressions. 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 for $\widehat{\rho}$ to be a root-n consistent, asymptotically normal estimator of the MR-LATE estimand ρ.

One convenient way to write the estimator is to consider the following moments:

$$\begin{aligned} E\left(Y_i T_i^b - \alpha^b - \lambda^b T_i^b\right) &= 0 \\ E\left(\left(Y_i T_i^b - \alpha^b - \lambda^b T_i^b\right) Z_i\right) &= 0 \\ E\left(Y_i T_i^a - \alpha^a - \left(\rho + \lambda^b\right) T_i^a\right) &= 0 \\ E\left(\left(Y_i T_i^a - \alpha^a - \left(\rho + \lambda^b\right) T_i^a\right) Z_i\right) &= 0 \end{aligned}$$

for some constants α^a, α^b, λ^b, and ρ. These moments correspond to the two instrumental variables regressions that comprise MR-LATE and ρ is the MR-LATE estimand. One could therefore estimate the constants α^a, α^b, λ^b, and ρ 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 correct standard errors for the MR-LATE estimator $\widehat{\rho}$. Alternatively, one might bootstrap the two instrumental variables regressions that define $\widehat{\lambda}^a$ and $\widehat{\lambda}^b$ and hence $\widehat{\rho}$.

Before proceeding further, we discuss two examples (related to our later empirical application) that illustrate the use of the MR-LATE estimator.

Example 1: Incompletely Measured Treatment. Due to growing attention regarding the status of women in developing countries, in household surveys of such countries a common type of question to ask is, "Who usually makes decisions about [X] in your household?" For example, the National Family Health Survey used in this paper asks this, with [X] referring to decisions regarding own health care, household purchases, and visits to family or relatives. Possible answers to this type of survey question are "the wife" or "the husband," but other ambiguous responses are also common. These other answers can include no response, or "someone else," or "both." Let $P = 1$ if the answer is "the wife," let $P = -1$ if the answer is "the husband," and let $P = 0$ denote any other answer. Assume we also have some outcome Y and some randomized binary instrument Z that is positively correlated with P.

Suppose we define treatment to be $D = 1$ if the wife makes most of the decisions regarding [X], otherwise $D = 0$ if the husband makes most of these decisions. This means we observe D for households that have $P = 1$ or $P = -1$, but we do not observe who is ultimately deciding [X] when $P = 0$. Thus, for those households with $P = 0$, whatever value we assigned to D would be subject to measurement error. One common procedure with this type of data is to construct a treatment proxy T where $T = 1$ if $P = 1$, otherwise $T = 0$, and apply the usual LATE estimator using T in place of D. Alternatively, one might discard all the observations that have $P = 0$, and apply the standard LATE estimator to the remaining observations. However, either of these estimators will generally be biased in unknown ways, due to misclassification errors or to correlated selection when $P = 0$. Instead of these typical estimators, suppose we let $T^a = 1$ if $P = 1$ ($T^a = 0$ otherwise), and let $T^b = 1$ if $P = -1$ ($T^b = 0$ otherwise). With these definitions of T^a and T^b, if observations having $P = -1$ and $P = 1$ are not mismeasured, MR-LATE will correctly identify and consistently estimate the LATE that corresponds to observing the true D for all households without error. Alternatively, even if some of these values of P are mismeasured (i.e., households incorrectly reporting who actually is making the decisions), MR-LATE can provide a significantly less biased estimate of LATE relative to standard instrumental variable estimation when the treatment is observed with error.[14]

Example 2: Threshold Crossing Model. It is common to combine responses from multiple empowerment related questions into a single index of the wife's relative decision making power within a household (see, e.g., Smith (2003), Roy (2008), Upadhyay et al. (2014) and references therein). Alternatively, as we do in our empirical application, one might estimate a power measure using a structural model of household behavior. One may then use MR-LATE to deal with potential measurement or estimation errors in these power indices or measures, as follows.

Define the indicator function $\mathbb{I}(\cdot)$ to equal one if its argument is true, and zero otherwise. Let R^* be the true measure of a woman's relative decision making power, or control over resources. Suppose that the true treatment D is determined by a typical threshold crossing model, so $D = \mathbb{I}(R^* \geq e)$ for some unobserved random threshold e. This means that the wife has primary decision making power, or controls a majority of household resources, corresponding to $D = 1$, if R^* is sufficiently large.

Assume we cannot observe R^* perfectly. Instead, what we observe, construct, or estimate, is the variable R, which is related to the true R^* by $R = R^* + \varepsilon$, where ε is an unknown error due to mismeasurement, misspecification, or estimation error. For simplicity in exposition, assume $\varepsilon + e$ is independent of Z and R^*.[15]

[14] A drawback of using individual self-reported measures of power, as in this example, is that each response concerns a relatively specific type of decision (e.g., own health care, household purchases, and visits to family or relatives) and is likely not representative of women's power more generally. Thus, in this application the LATE we obtain might not be particularly informative.

[15] The assumption that e and ε are independent of Z and R^* can be relaxed. We only assume this here to simplify the exposition. Specifically, if e or ε correlates with Z or R^*, then the correct expressions for p_0^a and p_1^b will need to condition on compliers, and will therefore be more complicated than

Let κ^a and κ^b be two constants chosen by the researcher, with $\kappa^a > \kappa^b$, and define treatment measures T^a and T^b as follows:

$$T^a = \mathbb{I}(R \geq \kappa^a) \quad \text{and} \quad T^b = \mathbb{I}(R < \kappa^b). \tag{13}$$

Note that $R \geq \kappa^a$ implies $R^* \geq \kappa^a - \varepsilon$. Thus, by construction,

$$p_0^a = \Pr(R^* \geq \kappa^a - \varepsilon \mid R^* < e).$$

Therefore, if κ^a is larger than the maximum value that $\varepsilon + e$ can take on, then $p_0^a = 0$. More generally, p_0^a is near zero if the chance that $\varepsilon + e$ is greater than κ^a is small. The intuition is straightforward: an individual is untreated (having $D = 0$) when their true R^* is sufficiently small. So, if we define T^a to equal one only when the observed or estimated R is very large, then the probability of having $T^a = 1$ when $D = 0$ is very small, meaning that p_0^a is near or equal to zero. In section A.2 in the Appendix, we provide a graphical illustration of this construction.

We could guarantee that p_0^a is zero, as desired, by taking κ^a to be infinite (or, if R is bounded, taking κ^a to be greater than the largest value that R can take on). Then, however, T^a would equal zero for every observation, and so would be useless as a measure of treatment. Using the notation of Theorem 1, if κ^a is infinite, then $p_0^a = p_1^a = \Pr(R^* \geq \kappa^a - \varepsilon \mid R^* \geq e) = 0$, which would lead to a violation of Assumption 2-iii.

In practice, we have a trade-off in the selection of κ^a. Recall $q^a = p_1^a / \left(p_1^a - p_0^a\right)$, and we want q^a to be as close as possible to one. This means choosing κ^a to make the probability p_0^a as small as possible relative to the probability p_1^a. The larger κ^a is, the closer p_0^a is to zero as desired. However, when κ^a is very large, p_1^a becomes small also. This is because the larger κ^a is, the less informative T^a becomes as a measure of treatment (e.g., the lower is the correlation between T^a and the true D). Empirically, to get q^a to be as close to one as possible, we will want to choose a moderate value of κ^a, which makes p_0^a close to zero and p_1^a relatively large (see the Monte Carlo analysis for further analysis of this point).

A comparable construction applies to T^b, where

$$p_1^b = \Pr(R^* < \kappa^b - \varepsilon \mid R^* \geq e),$$

so $p_1^b = 0$ if κ^b is less than the minimum value that $\varepsilon + e$ can take on, and p_1^b is near zero if the chance of $\varepsilon + e$ being less than κ^b is small.

Once κ^a and κ^b are chosen, the MR-LATE estimator is as described above, with $T^a = \mathbb{I}(R \geq \kappa^a)$ and $T^b = \mathbb{I}\left(R < \kappa^b\right)$. It is interesting to contrast this with ordinary LATE estimation. In this context, one would typically construct $T = \mathbb{I}(R \geq c)$, where c is one's best guess of the midpoint of $\varepsilon + e$, thereby constructing T to be as close as possible to the true unknown D. However, as discussed in section 3.2, estimates of LATE using a measure T like this in place of the true unknown D will generally be substantially (asymptotically) biased. The problem with replacing the unknown D with the known T in an ordinary LATE estimation is that compliers who have R close to c are precisely those who are most likely to be misclassified.

Point-identification, where MR-LATE equals the true LATE in this example, requires that $\varepsilon + e$ be bounded from both above and below, and that κ^b and κ^a be chosen to lie outside these bounds (but still within the range of R^* and R). In some contexts, we may have sufficient information to know these bounds: for instance, the threshold e might be an observable policy variable, and the measurement error ε might be rounding errors of known maximum possible magnitude.[16] When these bounds are unknown

the expressions we derive below, though the corresponding intuition regarding identification will remain unchanged.

[16] Even when such information is not available, we can still obtain bounds on the true LATE by selecting κ^a such that $p_1^a > p_0^a$ and κ^b such that $p_0^b > p_1^b$ as in Corollary 3. The selection of κ^b and κ^a as described in the previous paragraph will generally satisfy these inequalities, since they make p_1^a large

or ε is unbounded, the MR-LATE approach will stilly typically provide a much less biased estimate of the true LATE in presence of a misclassified treatment, as discussed in the previous subsection (and as shown in our Monte Carlo analyses below).

To illustrate these points, suppose we observed the answer to m survey questions of the type discussed in Example 1 above. Give each answer a value of one if the response is that the wife controls that decision, a value of minus one if the response is that the husband controls that decision, and a value of zero for any other response. Now define an estimated index R to equal the sum of these responses across all of the questions. If we took $\kappa^a = k$ for some integer k where $m-k$ is small and positive, we would be assuming that it is a rare evernt (p_0^a is close to zero) that the husband has most of the power when the household reports that the wife makes $m-k$ or more of the m decisions. Similarly, taking $\kappa^b = -k$ means assuming it is unlikely (p_1^b is close to zero) that the wife really has most of the power if they report that the husband makes $m-k$ or more of the decisions. Having p_0^a and p_1^b close to zero then ensures that the asymptotic bias in MR-LATE is small.

Similarly, in our later empirical application, where R^* is the share of household resources controlled by a woman, we cannot be certain that our choices of κ^b and κ^a will point identify LATE. However, our MR-LATE estimator is very likely to yield less biased estimates of the true LATE than the standard Imbens and Angrist (1994) LATE estimator in the presence of mismeasured treatment.

3.4 Monte Carlo Simulations

We now implement some Monte Carlo experiments to check the finite sample properties of the MR-LATE estimator. The experiments correspond to Examples 1 and 2 discussed in the previous subsection.

We derived the MR-LATE estimator without additional covariates. All of our results immediately extend to including covariates (assuming they are exogenous), by replacing expectations and covariances with conditional expectations and covariances, conditional on the conditional covariates. Moreover, since MR-LATE is just a function of IV estimators, when covariates only affect the model linearly, this is equivalent to including the covariates as additional regressors and instruments (as is common practice in applications of LATE estimators). However, it is common empirical practice to include covariates in LATE estimation, so we include a covariate X in our Monte Carlo simulations.

In our data generating process (hereafter DGP), we construct unobserved potential outcomes Y_0 and Y_1 and the corresponding observed outcome Y as follows:

$$Y_0 = 0.5 + X + S + V_0$$
$$Y_1 = 1.5 + X + S + V_1$$
$$Y = (1-D)Y_0 + DY_1$$

where X is an observed covariate, while S, V_0 and V_1 are random unobserved errors. The unobserved true treatment indicator D is given by $D = \mathbb{I}(R^* \geq 0)$, where the unobserved index R^* is

$$R^* = 0.1X + 0.1Z + 0.1S + U \tag{14}$$

with Z being the observed binary instrument and U is an additional error. The error U is equivalent to $\varepsilon + e$ in the previous subsection. The exogenous ly determined variables, errors, and parameter values in our simulations are all set to values that resemble our empirical application. In particular, we let $X \sim$

relative to p_0^a and p_0^b large relative to p_1^b.

Table 2: Simulations: Example 1

	D is known		D is unknown		
	OLS	2SLS	OLS	2SLS	MR-LATE
D	1.015	1.004	0.712	0.903	1.021
sd	0.014	0.101	0.020	0.096	0.084
Bias	0.015	0.004	-0.288	-0.097	0.021
MSE	0.000	0.010	0.083	0.019	0.007

Notes: In each simulation, the true value is set to 1.000. Results are based on 1,000 simulations for 10,000 observations each.

$\mathcal{N}(0,0.1)$, $S \sim \mathcal{N}(0,0.1)$, $Z = \mathbb{I}(\sim \mathcal{U}(0,1) > 0.9)$, $V_0 \sim \mathcal{N}(0,1)$, $V_1 \sim \mathcal{N}(0,1)$, and $U \sim \mathcal{N}(0,0.04)$. Unless noted otherwise, each sample consists of 10,000 observations, and the DGP is simulated 1,000 times.

In our first experiment (which corresponds to Example 1 in the previous subsection) we assume the econometrician just observes Y, X, Z, and P, where P takes on values of -1, 0, and 1 each with probability 1/3, based on low, medium and high values of R^*. So, $P = D = 1$ in one third of the sample, $D = 0$ and $P = -1$ in another third of the sample, and $P = 0$ in the remaining sample, regardless of the value of D.

We compare the performances of five different estimators. The first two estimators are infeasible since they assume D is observed without error. First, we estimate an ordinary least squares (OLS) regression model of Y on a constant, D, and X. Due to the correlation between treatment and potential outcomes, this OLS is inconsistent estimator of an average treatment effect. In particular, the variable S causes a violation of the unconfoundedness assumption. Second, we estimate a two stage least squares (2SLS) regression model of Y on a constant, D, and X, 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 uses D observed without error.

The remaining estimators we consider are feasible. First, we let $T = \mathbb{I}(P = 1)$ and estimate a linear regression of Y on a constant, T, and X, using OLS. Next, we estimate the same linear model using 2SLS, taking Z to be an instrument for T. The coefficient of T in this 2SLS regression is the standard LATE estimator, using T in place of the unobserved true treatment indicator D. Third, we construct $T^a = \mathbb{I}(P = 1)$ and $T^b = \mathbb{I}(P = -1)$, and apply the MR-LATE estimator. 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, T^j, and X, using Z as an instrument for T^j.

Table 2 shows our first simulation results. As expected, both of the OLS estimators are biased due to the correlation between treatment and potential outcomes, with the second OLS estimator behaving particularly poorly (bias of 29 percent), since it is also biased due to the measurement error in T relative to D. The standard 2SLS LATE estimator, which is infeasible because it uses the true unobserved D, has bias near zero. This is our benchmark for comparison of the feasible estimators. The standard 2SLS LATE estimator that uses T in place of D, which is feasible, has about 10 percent bias and a MSE almost double that of the infeasible LATE. This estimator is inconsistent due to the measurement error in T. Finally, our MR-LATE estimator, which is both feasible and consistent, has a relatively small bias (2 percent). Surprisingly, in this problem MR-LATE performs so well that it has a smaller mean squared error than even the infeasible 2SLS LATE that uses the true D.

In our next experiment, we assume that the practitioner observes multiple measures $P_1, \dots, P_m$ relating to treatment, each constructed analogously to the variable P defined above. Following Example 2 of section 3.3, let $R = \sum_{j=1}^{m} P_j$ be an index intended to approximate the true unobserved R^*. We now only

Table 3: Simulations: Example 2

	Panel A: $R = \sum_{j=1}^{6} P_j$					
			MR-LATE			
	OLS	2SLS	$k = 0$	$k = 1$	$k = 2$	$k = 3$
D	0.758	0.899	0.899	1.012	1.013	1.015
sd	0.019	0.092	0.092	0.083	0.081	0.081
Bias	-0.242	-0.101	-0.101	0.012	0.013	0.015
MSE	0.059	0.019	0.019	0.007	0.007	0.007
	Panel B: $R = R^* + \varepsilon$ (known bounds)					
	$\kappa = 0$			$\kappa = 0.05$		
	OLS	2SLS	MR-LATE	OLS	2SLS	MR-LATE
D	1.015	0.999	0.999	0.827	1.032	1.016
sd	0.014	0.097	0.097	0.018	0.105	0.085
Bias	0.015	-0.001	-0.001	-0.173	0.032	0.016
MSE	0.000	0.009	0.009	0.030	0.012	0.007
	Panel C: $R = R^* + \varepsilon$ (unknown bounds)					
			MR-LATE			
	OLS	2SLS	$\kappa = 0$	$\kappa = 0.01$	$\kappa = 0.05$	$\kappa = 0.1$
D	0.780	1.063	1.063	1.042	1.021	1.031
sd	0.020	0.109	0.109	0.105	0.091	0.088
Bias	-0.220	0.063	0.063	0.042	0.021	0.031
MSE	0.049	0.016	0.016	0.013	0.009	0.009

Notes: Results are based on 1,000 simulations of 10,000 observations each.

consider feasible estimators. Let $T = \mathbb{I}(R \geq 0)$. This T is an unbiased estimate of D given R. We first estimate a linear regression of Y on a constant, T, and X, using OLS. Next, we estimate the same linear model using 2SLS, taking Z to be an instrument for T. The coefficient of T in this 2SLS regression is again the standard LATE estimator that uses T in place of the unobserved true treatment indicator D. Third, we let $T^a = \mathbb{I}(R \geq k)$ and $T^b = \mathbb{I}(R < -k)$, for a few different choices of the integer k, and apply the MR-LATE estimator. Since R can only take on integer values from $-m$ to m, k must be an integer less than m. Specifically, we set $m = 6$ and consider $k = 0, 1, 2$, and 3.[17] Results of this experiment are reported in Panel A of Table 3. Note that for $k = 0$, MR-LATE is numerically identical to the 2SLS estimator. As the table shows, we find that MR-LATE is not too sensitive to the choice of k and performs better than the other estimators for all $k > 0$.

For our last set of experiments, we again assume the econometrician can observe Y, X, Z, and R, but now R is given by $R = R^* + \varepsilon$, where ε is either normal with variance 0.04 (the same variance as U), or ε is distributed as a truncated normal ranging between $-\kappa$ and κ. We may interpret ε either as measurement error in R^*, or as the specification and estimation error in a structural model estimate of R^*.[18] Once again, the alternative feasible estimators include ignoring the measurement error in R and applying OLS or 2SLS using $T = \mathbb{I}(R \geq 0)$ in place of D. We compare these to MR-LATE, using $T^a = \mathbb{I}(R \geq \kappa)$ and $T^b = \mathbb{I}(R < -\kappa)$ for a few different values of κ. In Panel B of Table 3, we report results where ε is bounded, with bounds κ and $-\kappa$ known to the econometrician. In these cases, the MR-LATE estimator consistently estimates

[17] For this particular DGP, R turns out to be highly correlated with R^*, resulting in the proxy T having about a 0.90 correlation with D. So this example represents a case of relatively little measurement error. Our empirical application likely has more measurement error than this.

[18] It is worth pointing out that our choice of variance of the measurement error is not excessively large, since otherwise T^a and T^b would not be informative about D, e.g., T^a is positively correlated with D. This is analogous to the typical real world situation where misclassification of the treatment indicator matters but does not lead to a sample where the share of actual treated in T^a is smaller than the share of misclassified actual untreated, and analogously, where the share of actual untreated in T^b is smaller than the share of misclassified actual treated.

the true LATE. In Panel C of Table 3, ε is normal so bounds are nonexistent. In this case, the MR-LATE estimator does not consistently estimate the true LATE. Both with known bounds and nonexistent bounds, MR-LATE always performs considerably better than OLS and 2SLS (except in the extreme case of $\kappa = 0$, where MR-LATE reduces to being numerically identical to 2SLS by construction).

As discussed in subsection 3.3, in the case of unknown or nonexistent bounds, the practitioner faces a trade-off when choosing κ. The larger the chosen value of κ, the closer p_0^a and p_1^b are to zero, but the farther $\widehat{\lambda}^a$ and $\widehat{\lambda}^b$ are to their limiting values at any given sample size. To further explore this tradeoff, we examine the performance of MR-LATE and 2SLS as the sample size increases (results are in Figure A1 in the Appendix; Table A3 in the Appendix shows similar analysis for known bounds).

Our conclusions are that the bias of MR-LATE estimation is very small compared to the bias in the standard 2SLS LATE estimator using T in place of D (which confirms our analysis of section 3.2). Moreover, when there is a choice in how to define T^a and T^b (that is, when one must choose k or κ), the MR-LATE estimator is relatively insensitive to that choice, and remains superior to alternative estimators for a wide range of such choices.

4 Women's Control Over Resources and Family Health

We apply the MR-LATE estimator to study the impact of women's intra-household empowerment on family members' health outcomes in India. As primary caregivers, the greater is a woman's decision power and control over resources, the more effective her care for herself and her children can be (Smith (2003)). We therefore 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, we define $D = \mathbb{I}(R^* \geq e)$, where R^* is a woman's unobserved decision making power (or the share of household resources that she controls) and e is a threshold that might vary across households for unobserved reasons.

We examine the impact of treatment D on a variety of health related outcomes Y. For adults, we consider body mass index (hereafter BMI), an indicator for being underweight, and an indicator for being anemic. For children, we consider height-for-age and weight-for-age z-scores, recent occurrences of diarrhea, fever and cough, and an indicator for whether a child has been vaccinated against one or more diseases. 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.

As discussed in section 3.3, there exist 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.[19] A typical LATE study might use these survey responses (or an index of them) as direct measures of treatment, despite the fact that they are at best coarse proxies for women's overall decision making power. In fact, in our dataset we find a quite low correlation between these self-reports and the instrument Z.[20] Thus, using these self-reports yields weak and uninformative estimates of treatment effects, regardless of whether they are taken to equal treatment for ordinary LATE estimation, or if the MR-LATE estimator is used (empirical results using these self-reports are in Appendix A.3).

For our primary analysis, we instead employ a structural model that makes use of both economic theory (of collective household decision making), and detailed household expenditure data, to estimate

[19] These self-reports are examples of what we called P_j in the previous section.

[20] This is in line with results in Heath and Tan (2019), who document no significant link between self-reported participation in household decisions and these inheritance reforms.

a measure of women's resource control and hence decision making power. We then employ MR-LATE estimation to account for possible model misspecification and estimation errors in the construction of this treatment measure. In contrast to our structural model of treatment, which is grounded in economic theory, we do not attempt to structurally model health production as a function of control over resources.[21] Our goal is instead to estimate a treatment effect of D on Y using a causal rather than structural model, exploiting plausibly exogenous variation in our instrument Z that correlates with D.

Our analysis assumes that D as we've defined it is a relevant measure of treatment for health outcomes. To the extent that health outcomes are the results of many health decisions, it is reasonable to assume that the family member with the most power over resource allocations will generally determine many of these decisions.[22] 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.[23]

An alternative possibility is that the magnitude of R^* itself, not just D, is more relevant for determining Y. To test this, and for comparison with our main results, we estimate a model that linearly regresses Y on R (our estimate of R^*) and other covariates (estimation results are reported in Appendix A.3). The results provide little evidence of a continuous relationship between women's resource shares R and health outcomes Y. One could also try more flexible non-linear 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, particularly since there is no reason to believe the errors in R would satisfy classical measurement error assumptions. An advantage of discretizing treatment the way we do is to mitigate the impact of these measurement errors.

Note that by definition, LATE averages over all random variables that affect outcomes, conditional on treatment. In our model where the treatment D is defined as $\mathbb{I}(R^* \geq e)$, this means that the average treated outcome includes averaging over R^* for all compliers who have $R^* \geq e$, while the average control outcome includes averaging over R^* for all compliers who have $R^* < e$. As reported below, with e centered at 50 percent, we estimate (based on R, since R^* is unobserved) that the treated group has an average R^* of around 60 percent, while the control has group has an average R^* of approximately 40 percent. So, the average woman in the treatment group has much more control over resources than the average woman in the control group, which is important to recognize for interpreting our results.

4.1 Modeling Women's Control over Household Resources

A key difficulty in observing or calculating the resource shares R^* is that control over household resources in its entirety is not observable. 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.

Define a good to be *private* if it is not shared or consumed jointly. Define a good to be *assignable*

[21] One could imagine many possible mechanisms linking the unobserved resource control treatment indicator D to health outcomes Y. For example, mothers may have different priorities than fathers regarding expenditures on health related goods, or circumstances that contribute to women having greater power might also affect the health care of family members. Previous research (discussed in section 2) provides indirect evidence of linkages between D and Y, by establishing, e.g., how variables that affect women's power correlate with health outcomes of wives and children.

[22] This is roughly analogous to voter models, where outcomes are primarily determined by the party with the most power.

[23] Moreover, changes in behavior may occur when women control a majority of household resources. Bertrand et al. (2015), for instance, study the consequences of relative income within households in the United States and document substantial changes in social and economic outcomes following the violation of the prescription that "a man should earn more than his wife."

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.

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. Assume we observe the household's total expenditure 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. The Engel curve for person $t = w, m$ is defined to be that person's demand for clothing, expressed as a function of the total resources person t controls: R^*M for women and $(1-R^*)M$ for men. Then, given w_w, w_m and M, 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 A.4 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.[24] We employ the commonly used Piglog (price independent generalized logarithmic) functional form for these Engel curves, which is

$$\begin{cases} w_w &= R^*\delta_w + R^*\beta \ln(R^*M) \\ w_m &= (1-R^*)\delta_m + (1-R^*)\beta \ln((1-R^*)M) \end{cases} \tag{15}$$

where $\beta = \beta(X)$, $\delta_w = \delta_w(X)$, $\delta_m = \delta_m(X)$, and $R^* = R^*(X)$.[25] Note that the demand functions for other goods (those that are not private and assignable) are more complicated, but are not required to estimate the resource shares. DLP prove that the functions $\beta(X)$, $\delta_w(X)$, $\delta_m(X)$, and $R^*(X)$ are identified in this model.[26] For our empirical application, we assume the functions $\beta(X)$, $\delta_w(X)$, $\delta_m(X)$, and $R^*(X)$, are all

[24] 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.

[25] Jorgenson et al. (1982) Translog demand system and the Deaton and Muellbauer (1980) Almost Ideal Demand System have Engel curves of the piglog form, and piglog Engel curves were also used in empirical collective household models estimates by DLP. DLP estimate resource shares for children as well as the adults in the household. Due to data availability, we can only estimate resource shares of the mother and father. Our framework still allows for caring preferences and for the possibility that mothers and fathers may value differently the well-being of their children.

[26] The identification depends 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.

linear in their arguments.[27] In particular, we specify

$$R^*(X) = \theta_0 + \theta_1 X_1 + \ldots + \theta_K X_K. \tag{16}$$

4.2 Estimation Strategy

For our empirical analysis, we employ two different datasets from India. One, the 62^{nd} 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 third round of the National Family Health Survey (hereafter NFHS), reports the health outcomes we use for our causal treatment effects estimation. 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.

We append an error term to the equations in system (15), yielding a two equation system that we estimate using non-linear Seemingly Unrelated Regressions (SUR) and the NSS data.[28] Let $\widehat{\theta}$ denote the estimate of θ in equation (16). Then, for each individual i drawn from the NFHS data, we use these estimates to predict the share of resources controlled by the woman in individual i's household as

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

Our goal is to estimate a LATE of $D = \mathbb{I}(R^* \geq e)$ for a range of health outcomes Y. We separately consider health outcomes for mothers, fathers, and children. So, e.g., when i is a child and Y_i is an indicator of whether the child has been vaccinated, the treatment effect we wish to estimate is the change in i's probability of being vaccinated if he/she is exposed to highly empowered mothers, corresponding to $D = 1$. We wish to estimate this treatment effect, even though the mother's true resource share R_i^* is unobserved.

As discussed in Example 2 in section 3.3, 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 $T_i^b = \mathbb{I}\left(R_i < \kappa^b\right)$, where κ_a and κ_b are chosen bounds with $\kappa_a > \kappa_b$. 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 $\widehat{\rho} = \widehat{\lambda}^a - \widehat{\lambda}^b$.

The way we choose bounds κ_a and κ_b is as follows. We consider a few different percentages $\mathscr{K}$, and for each we let κ_a be the value such that $\mathscr{K}/2$ percent of the sample has R in the interval $[50, \kappa_a]$ and κ_b is the value such that $\mathscr{K}/2$ percent of the sample has R in the interval $[\kappa_b, 50]$. This is consistent with (but does not require) having e being centered around 50 percent, implying that households with $D_i = 1$ are usually ones in which the mother has control over a majority of household resources.[29] Essentially, $\mathscr{K}$ corresponds to how much misclassification error T^a and T^b would contain if R was exactly equal to R^*.

[27] We do not include Z as an element of X (particularly in the equation for R^*) for two reasons. First, doing so could induce spurious correlation between the estimated treatment indicator 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 (2019)). Results are confirmed and available upon request.

[28] The non-linear SUR is iterated until the estimated parameters and the covariance matrix converge. The result is asymptotically equivalent to maximum likelihood with multivariate normal errors.

[29] Recall from section 3.3 that the relevant threshold e_i can vary across households for unobserved reasons. In effect, random variation in e is observationally equivalent to measurement error and is therefore addressed by our estimation method.

We apply this procedure separately for a few different health outcomes for men, women, and children, using a few different values of $\mathscr{K}$.[30] The choice of $\mathscr{K}$, and hence of the bounds κ_a and κ_b, that minimizes the bias or mean squared error of MR-LATE depends on unknown features of the data generating process. However, we found in the previous sections that MR-LATE has less bias than the standard LATE estimator for a wide range of chosen bounds, and gave estimates that were relatively insensitive to the exact choice of bounds over a range of moderate bound choices. Our empirical results below similarly vary little across the middling values of $\mathscr{K}$ that we consider.

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.[31] 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.[32] Table A5 in the Appendix 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 (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 one to four children.[33] 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-

[30] In the special case of $\mathscr{K} = 0$, the MR-LATE estimator becomes 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 $D_i = \mathbb{I}(R_i^* \geq e_i)$. Generally, this $\mathscr{K} = 0$ estimator will only be consistent if there is no measurement or estimation error in R_i, and if e_i exactly equals 50 percent for all households.

[31] 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 from 1 to 4 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. The final estimation sample contains 7,480 households.

[32] 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 in the Appendix). Hence, we do not appear to have a weak identification problem with our data.

[33] As shown in table A6, 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.

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.[34]

The Hindu Succession Act and its Amendments. We exploit changes in the Indian inheritance law to construct a plausibly unconfounded instrumental variable Z. A woman's right to inherit land and other property is often claimed to play a significant role in determining women's position 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.[35] 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.

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: 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). 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 women's higher control of household resources.[36]

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 (2019)). That type of analysis

[34]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).

[35]The HSA did not apply to individuals of other religions, such as Muslims, Christians, Parsis, Jews, and other minority communities. Most laws for Christians formally granted them equal rights as of 1986. However, gender equality for Christian women was and is not the practice, as the Synod of Christian Churches has being arranging legal counsel to help draft wills to disinherit female heirs. The inheritance rights of Muslim women in India are governed by the Muslim Personal Law (Shariat) Application Act of 1937, under which daughters inherit only a fraction of what sons inherit (Agarwal, 1995).

[36]This assumption is plausible. Despite other factors and policies may have differentially affected young Indian women, we do not expect these to vary by religion. Moreover, Calvi (2019) demonstrates that the potential endogeneity of women's time of marriage is not a concern. She also shows that the HSA amendments changed neither sorting in the marriage market nor total household expenditures. Finally, the existence of other changes *resulting from* women's increased power within their marital families due to their HSA exposure (e.g., higher labor force participation; see Heath and Tan (2019)) does not determine a violation of the exclusion restriction.

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.

4.3 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 A14 in the Appendix.

Structural Estimates. Table 4 contains descriptive statistics of the predicted resource shares R in the NSS and NFHS samples. Our tables present R in percentage form, so e.g., the reported mean of R of 48 means the wife is estimated to control 48 percent of the household's resources. The reported 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 means and standard deviations of R in the two samples are very similar, indicating that the samples are highly 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 our ex ante expectations. For instance, the average R is particularly high in the North-East states (61 percent), which is consistent with the presence of a number of matrilineal societies and cultures in these regions (Khasi and Garo societies, for example). In contrast, North Indian women seem to have a much lower control over resources (40 percent). Finally, highly educated women (who have completed high school) are found to have a substantially higher command over resources (55 percent) relative to low educated women (47 percent).

Define $T = \mathbb{I}(R \geq 50)$. In the NFHS sample, women who have $T = 1$ have an average R of 60, while those having $T = 0$ have an average R of 40. So, while we cannot know the average fraction of resources controlled by the truly treated and untreated, i.e., $E(R^* \mid D)$, our estimates of $E(R \mid T)$ indicate that the treated group controls a considerably larger fraction of household resources than the control group.[37] 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.[38]

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 A.3 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.

[37] Figure A3 in the Appendix shows the empirical distribution of R in the NFHS sample (the distribution in the NSS is very similar).

[38] Using a more recent round of the NSS of Consumer Expenditure that includes a richer set of assignable goods, Calvi (2019) estimates separate resource shares for men, women and children. For households with children Calvi estimates that women' resource shares are on average only 67 percent of men's, which suggests that, relative to fathers, a higher fraction of resources controlled by mothers is diverted to their children.

Table 4: Estimated Resource Shares and Woman's Power

	Obs.	Mean	St. Dev.	Min.	Max.
NSS Sample :					
Woman's Resource Share (R)	7,440	48.27	11.64	13.54	86.92
$T = \mathbb{I}(R \geq 50)$	7,440	0.41	0.49	0.00	1.00
NFHS Sample:					
Woman's Resource Share (R)	22,767	48.13	11.81	6.50	86.97
$T = \mathbb{I}(R \geq 50)$	22,767	0.40	0.49	0.00	1.00

Note: Household level data. R is in percentage form.

Causal Estimates. As discussed in section 3.3, if we ignored specification and estimation error in R, we would 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 our best guess $T = \mathbb{I}(R \geq 50)$. However, this may lead to large biases due to measurement, estimation, and specification errors in R. We therefore apply our MR-LATE estimator to account for these errors and to reduce estimation bias. To do so, we construct bounds κ_a and κ_b based on choosing a misclassification percentage $\mathscr{K}$ as described in section 4.2. The percentages $\mathscr{K}$ we consider are $\mathscr{K}$ equal to 0, 1, 5, 10, and 20. Table A7 in the Appendix reports the bounds κ^a and κ^b that correspond to each of these values of $\mathscr{K}$. 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 5 reports the resulting MR-LATE estimates for adult health outcomes, while estimates for children's health outcomes are reported in Table 6. Bootstrapped standard errors are reported in parentheses. The MR-LATE estimates for $\mathscr{K} = 0$ are, by construction, numerically identical to the standard LATE estimator that ignores errors in R. In a few instances, even quite small deviations of $\mathscr{K}$ from zero substantially change the MR-LATE estimates for some outcomes, showing that accounting for errors in R 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 7.7 and that women in treated households are 72 percentage points less likely to be underweight and 52 p.p. less likely to be anemic. The large magnitudes are consonant with us comparing women who on average have control of about 60 percent of household resources with women who on average control about 40 percent of household resources.[39]

A mother's control over household resources positively affects her children's health, too. A highly empowered mother decreases her children's likelihood of being sick with cough, fever or diarrhea in the two weeks prior to the survey by 66, 43 and 45 percentage points, respectively. She 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.

Table A8 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. We include household level and individual level characteristics, fixed effects and state specific time trends in all specifications. Even conditioning on several sources of unobserved heterogeneity, the instrument Z is positively and

[39]We acknowledge that these magnitudes are also impacted by the size of the *unobserved* misclassification probabilities (Corollary 1). In our context, however, we do not have compelling reasons to believe $q^a - q^b$ is significantly larger than one.

Table 5: Adult's Health: Estimated LATE of $D = (R^* \geq e)$ on Y

	Women			Men		
	BMI	Pr(BMI≤18.5)	Pr(Anemic)	BMI	Pr(BMI≤18.5)	Pr(Anemic)
$\mathscr{K} = 0$	9.7989	-0.9175	-0.5572	2.4074	-0.2778	0.0860
	(1.9869)	(0.2288)	(0.2531)	(2.4452)	(0.2971)	(0.2318)
$\mathscr{K} = 1$	9.2903	-0.8836	-0.5194	2.0658	-0.2345	0.0556
	(2.1346)	(0.2200)	(0.2420)	(2.7973)	(0.2785)	(0.2181)
$\mathscr{K} = 5$	9.1945	-0.8482	-0.5239	3.0378	-0.2629	0.0925
	(4.1294)	(0.2225)	(0.2487)	(4.3572)	(0.2926)	(0.2195)
$\mathscr{K} = 10$	12.4103	-1.1476	-0.6254	3.7843	-0.3852	0.0558
	(6.2875)	(0.2715)	(0.3124)	(7.4517)	(0.3633)	(0.2630)
$\mathscr{K} = 20$	7.7153	-0.7232	-0.5151	-1.0725	-0.5217	-0.1373
	(9.1580)	(0.2915)	(0.3730)	(15.7390)	(0.4454)	(0.3185)

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 with up to 4 children. The men sample includes married men of age 15 to 54 in nuclear households with up to 4 children. 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.

Table 6: Children's Health: Estimated LATE of $D = (R^* \geq e)$ on Y

	Weight-for-age (z-score)	Height-for-age (z-score)	Pr(Cough)	Pr(Fever)	Pr(Diarrhea)	Pr(Any Vaccination)
$\mathscr{K} = 0$	2.0547	2.8140	-0.6649	-0.6114	-0.4549	-0.1296
	(1.4246)	(1.8384)	(0.2973)	(0.4423)	(0.2185)	(0.2797)
$\mathscr{K} = 1$	2.0093	2.7298	-0.7264	-0.6565	-0.5153	-0.2065
	(1.3814)	(1.7865)	(0.3176)	(0.4551)	(0.2255)	(0.2883)
$\mathscr{K} = 5$	1.7328	3.0385	-0.7306	-0.6273	-0.4923	-0.2159
	(1.3947)	(1.6276)	(0.2868)	(0.4259)	(0.1981)	(0.3159)
$\mathscr{K} = 10$	2.4247	3.0848	-0.8890	-0.6361	-0.5121	-0.1029
	(1.7819)	(1.8369)	(0.3115)	(0.4158)	(0.2023)	(0.4178)
$\mathscr{K} = 20$	2.2458	2.8007	-0.6878	-0.4312	-0.6457	0.0141
	(1.6342)	(1.8806)	(0.3053)	(0.3345)	(0.1789)	(0.5104)

Notes: Estimates are obtained using the NFHS-3 data and the MR-LATE estimator. The sample includes children 0 to 5 in nuclear households with up to 4 children. 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.

significantly correlated with T^a and T^b. The first stage F-test statistics are largely above 10 for $\mathscr{K} = 0$, 1, 5, 10, which is consistent with not having a weak instrument problem. The F-test statistics, however, do fall below 10 for our largest bounds, corresponding to $\mathscr{K} = 20$. This is consistent with our predictions, since T^a and T^b become less correlated with the true D and $1-D$ once the bounds κ^b to κ^a become overly large. We also find in Tables 5 and Table 6 that our estimates are, reassuringly, mostly insensitive to the exact choice of $\mathscr{K}$, excluding extreme values of $\mathscr{K}$.

4.4 External Validity

Instead of going to the trouble of estimating resource shares, we could have simply calculated the ATE of the change in inheritance laws Z on the outcomes Y. However, our interest is not in these particular inheritance policies. Rather, we wish to learn about the likely impact of any policy that changes women's power within the household, as measured by control over resources. This is the value of defining treatment D the way we do.

Given our assumptions, MR-LATE either consistently (or at least reduces the bias in) estimates the average treatment effect for compliers. As with ordinary LATE estimation, the question then remains: How representative are compliers of the general population and, hence, how close are our estimates to the population ATE of empowering women (by giving them control of household resources) on family health?

In Appendix A.5, we discuss the conditions under which our compliers will be close to representative of the general population. Exploiting the fact that our treatment is defined according to a threshold crossing model, we can cast our problem within the marginal treatment effect framework of Heckman and Vytlacil (1999), where the relevant threshold e_i is a source of unobserved heterogeneity across households. We then make use of results in Kowalski (2016) (and references therein) to test the external validity of our LATE estimates. While this analysis abstracts from the issues of measurement and estimation error in R, and therefore needs to be interpreted with caution, the results in Appendix A.5 suggest that our estimated treatment effects are unlikely to vary much with the choice of instrument, lending some empirical support for the external validity of our estimates (or more precisely, providing no evidence that our empirical results lack external validity).

5 Conclusion

We propose a novel approach to study the effects of intra-household women's empowerment on the health status of family members in India. Our causal model looks at the effect on health outcomes Y of a treatment D, defined as a woman having relatively high control over household resources. The treatment is based on an unobserved continuous variable R^* and therefore is itself unobserved. We rely on a structural model of intra-household decision making to obtain R, an estimate of R^*, and use R to estimate treatment. Due to potential measurement, estimation, and specification errors in the structural model for R, our estimated treatment indicator may not equal or consistently estimate the true treatment indicator D. To account for these several possible sources of error, we propose a new mismeasurement-robust LATE estimator (called MR-LATE) which uses two estimated treatment indicators, T^a and T^b, along with an outcome Y and an instrument Z, to estimate the same LATE that would be obtained if we could observe the true treatment indicator D. Under some relatively strong conditions, MR-LATE consistently estimates the same LATE that would be obtained if the true D were observed. Under much more general conditions, MR-LATE provides estimates that usually have far less asymptotic bias than the standard LATE estimator in the presence of measurement or estimation error in D.

Taken together, our empirical results indicate that policies aimed at empowering women within households, such as strengthening their inheritance and property rights, tend to increase their control over household resources. This increased control over resources leads to improvements in women's and children's overall health, while having little effect on men's health.

Our empirical application emphasizes the potential use of MR-LATE in situations where treatment is not observed and must be estimated. However, we wish to stress that MR-LATE can also be used to reduce bias in other applications where a binary treatment indicator is simply observed with error, due for instance to misreporting or contamination.

More broadly, our analysis highlights potential advantages of combining both structural and causal (or quasi-experimental) methodologies in conducting empirical analyses. The MR-LATE estimator specifically accounts for the fact that structural estimation generally suffers from multiple errors, including specification errors. But 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, since the researcher can directly focus on treatments that are motivated by theory (in our example, women's decision power or control of household resources), instead of only calculating the treatment effects of less relevant proxies that happen to be directly observed.

References

ABADIE, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292. [9], [13]

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

AIGNER, D. J. (1973): "Regression with a binary independent variable subject to errors of observation," *Journal of Econometrics*, 1, 49 – 59. [5]

ANGRIST, J. D. (2004): "Treatment Effect Heterogeneity in Theory and Practice," *The Economic Journal*, 114, C52–C83. [41]

ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455. [6]

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. [5]

ANGRIST, J. D. AND J.-S. PISCHKE (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30. [7]

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*. [6]

ATTANASIO, O. AND V. LECHENE (2002): "Tests of Income Pooling in Household Decisions," *Review of Economic Dynamics*, 5, 720–748. [5]

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. [40]

BARGAIN, O., O. DONNI, AND P. KWENDA (2014): "Intrahousehold distribution and poverty: Evidence from Cote d'Ivoire," *Journal of Development Economics*, 107, 262–276. [40]

BARGAIN, O., G. LACROIX, AND L. TIBERTI (2018): "Validation of Collective Models for Individual Welfare Analysis using Direct Evidence on Sharing," mimeo. [4]

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], [6]

BECKER, G. S. (1965): "A Theory of the Allocation of Time," *The Economic Journal*, 493–517. [2], [4]

——— (1974): "A Theory of Social Interactions," *Journal of Political Economy*, 82, 1063–1093. [2]

——— (1981): *A Treatise on the Family*, Harvard University Press. [2], [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. [6]

BERTANHA, M. AND G. W. IMBENS (2014): "External Validity in Fuzzy Regression Discontinuity Designs," Tech. rep., National Bureau of Economic Research. [41]

BERTRAND, M., E. KAMENICA, AND J. PAN (2015): "Gender Identity and Relative Income Within Households," *The Quarterly Journal of Economics*, 130, 571–614. [20]

BJÖRKLUND, A. AND R. MOFFITT (1987): "The Estimation of Wage Gains and Welfare Gains in Self-selection Models," *The Review of Economics and Statistics*, 42–49. [40]

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. [5]

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. [5]

BLUNDELL, R., P.-[U+FFFD] CHIAPPORI, AND C. MEGHIR (2005): "Collective Labor Supply with Children," *Journal of Political Economy*, 113, 1277–1306. [4]

BOLLINGER, C. R. (1996): "Bounding mean regressions when a binary regressor is mismeasured," *Journal of Econometrics*, 73, 387 – 399. [5], [6]

BOSE, N. AND S. DAS (2015): "Women's Inheritance Rights, Household Allocation and Gender Bias," *Unpublished*. [5]

BRINCH, C., M. MOGSTAD, AND M. WISWALL (forthcoming): "Beyond LATE With a Discrete Instrument," *Journal of Political Economy*. [40], [41]

BROWN, C., J. PENGLASE, AND R. CALVI (2018): "Sharing the Pie: Undernutrition, Intra-household Allocation, and Poverty," Boston college. [4]

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], [38]

CALVI, R. (2019): "Why Are Older Women Missing in India? The Age Profile of Bargaining Power and Poverty," *Journal of Political Economy*, Forthcoming. [4], [5], [22], [24], [25] CARD, D. (2001): “Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems,” *Econometrica*, 69, 1127–1160. [5]

CARNEIRO, P., J. J. HECKMAN, AND E. J. VYTLACIL (2011): “Estimating Marginal Returns to Education,” *The American economic review*, 101, 2754–2781. [40]

CHERCHYE, L., B. DE ROCK, A. LEWBEL, AND F. VERMEULEN (2015): “Sharing Rule Identification for General Collective Consumption Models,” *Econometrica*, 83, 2001–2041. [40]

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]

DEATON, A. AND J. MUELLBAUER (1980): “An Almost Ideal Demand System,” *American Economic Review*, 70, 312–26. [21]

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. [5], [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], [6]

DUFLO, E. (2003): “Grandmothers and Granddaughters: Old-age Pensions and Intrahousehold Allocation in South Africa,” *The World Bank Economic Review*, 17, 1–25. [2], [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], [4], [21], [38]

FRANGAKIS, C. E. AND D. B. RUBIN (2002): “Principal stratification in causal inference,” *Biometrics*, 58, 21–29. [6]

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. [5]

HADDAD, L. J., J. HODDINOTT, H. ALDERMAN, ET AL. (1997): *Intrahousehold Resource Allocation in Developing countries*, Johns Hopkins University Press. [2]

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. [5], [19], [22], [24]

HECKMAN, J., S. URZUA, AND E. VYTLACIL (2006): “Understanding Instrumental Variables in Models with Essential Heterogeneity,” *Review of Economics and Statistics*, 88, 389–432. [7]

HECKMAN, J. J. (2010): “Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy,” *Journal of Economic Literature*, 48, 356–98. [7]

HECKMAN, J. J. AND E. VYTLACIL (2005): “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, 73, 669–738. [40]

HECKMAN, J. J. AND E. J. VYTLACIL (1999): “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects,” *Proceedings of the national Academy of Sciences*, 96, 4730–4734. [28], [40], [41]

——— (2007): "Econometric Evaluation of Social Programs, part I: Causal Models, Structural Models and Econometric Policy Evaluation," *Handbook of econometrics*, 6, 4779–4874. [7], [40]

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. [5]

HU, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, 144, 27 – 61. [6]

HU, Y., J.-L. SHIU, AND T. WOUTERSEN (2015): "Identification and estimation of single-index models with measurement error and endogeneity," *The Econometrics Journal*, 18, 347–362. [6]

IMAI, K., L. KEELE, AND T. YAMAMOTO (2010): "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statist. Sci.*, 25, 51–71. [6]

IMBENS, G. W. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399–423. [7]

IMBENS, G. W. AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [3], [6], [7], [9], [13], [16], [23], [26]

JORGENSON, D., L. LAU, AND T. M. STOKER (1982): *The Transcendental Logarithmic Model of Aggregate Consumer Behavior*, Greenwich: JAI Press, 97–238, welfare 1, ch. 8, pp. 203-356. [21]

KANE, T. J., C. E. ROUSE, AND D. STAIGER (1999): "Estimating Returns to Schooling When Schooling is Misreported," NBER working paper 7235. [5]

KLEPPER, S. (1988): "Bounding the effects of measurement error in regressions involving dichotomous variables," *Journal of Econometrics*, 37, 343 – 359. [6]

KOWALSKI, A. (2016): "How to Examine External Validity Within an Experiment," *Mimeo*. [28], [40], [41], [58]

KOWALSKI, A., Y. TRAN, AND L. RISTOVSKA (2016): "MTEBINARY: Stata module to compute Marginal Treatment Effects (MTE) With a Binary Instrument," Statistical Software Components, Boston College Department of Economics. [41], [58]

KREIDER, B., C. GUNDERSEN, AND D. JOLLIFFE (2012): "Identifying the effects of food stamps on child health outcomes when participation is endogenous and misreported," . [6]

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. [6]

LAFAVE, D. AND D. THOMAS (2017): "Extended Families and Child Well-being," *Journal of Development Economics*, 126, 52–65. [2], [5]

LEWBEL, A. (1998): "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," *Econometrica*, 66, 105–121. [6]

——— (2007): "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75, 537–551. [6]

——— (2019): "The Identification Zoo - Meanings of Identification in Econometrics," *Journal of Economic Literature*, Forthcoming. [7]

LEWBEL, A. AND K. PENDAKUR (2008): "Estimation of Collective Household Models with Engel Curves," *Journal of Econometrics*, 147, 350–358. [4], [40] LISE, J. AND S. SEITZ (2011): "Consumption Inequality and Intra-household Allocations," *Review of Economic Studies*, 78, 328–355. [40]

MACOURS, K., N. SCHADY, AND R. VAKIS (2012): "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment," *American Economic Journal: Applied Economics*, 4, 247–73. [5]

MAHAJAN, A. (2006): "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631–665. [6]

MAITRA, P. (2004): "Parental Bargaining, Health Inputs and Child Mortality in India," *Journal of Health Economics*, 23, 259 – 291. [5]

MANSKI, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *The American Economic Review*, 80, 319–323. [6]

MENON, M., K. PENDAKUR, AND F. PERALI (2012): "On the Expenditure-dependence of Children's Resource Shares," *Economics Letters*, 117, 739–742. [40]

MOLINARI, F. (2008): "Partial identification of probability distributions with misclassified data," *Journal of Econometrics*, 144, 81 – 117. [6]

PEARL, J. (2009): "Causal inference in statistics: An Overview," *Statistics Surveys*, 3, 96–146. [7]

PENGLASE, J. (2017): "Consumption Inequality among Children: Evidence from Child Fostering in Malawi," Boston college. [4]

PRENTICE, R. L. (1989): "Surrogate endpoints in clinical trials: definition and operational criteria," *Statistics in medicine*, 8, 431–440. [6]

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. [2], [5]

ROSENBAUM, P. R. (1984): "The consquences of adjustment for a concomitant variable that has been affected by the treatment," *Journal of the Royal Statistical Society. Series A (General)*, 656–666. [6]

ROY, S. (2008): "Female Empowerment Through Inheritance Rights: Evidence from India," Mimeo, London School of Economics, London. [5], [14], [24]

——— (2015): "Empowering women? Inheritance rights, female education and dowry payments in India," *Journal of Development Economics*, 114, 233–251. [5], [24]

RUBALCAVA, L., G. TERUEL, AND D. THOMAS (2009): "Investments, Time Preferences, and Public Transfers Paid to Women," *Economic Development and cultural change*, 57, 507–538. [2], [5]

SCHULTZ, T. P. (1990): "Testing the Neoclassical Model of Family Labor Supply and Fertility," *Journal of Human resources*, 599–634. [5]

SEN, A. (1983): "Economics and the Family," *Asian Development Review*, 1. [2]

——— (1988): *Family and Food: Sex-Bias in Poverty*, New York: Columbia University Press. [2]

——— (1989): "Cooperation, Inequality, and the Family," *Population and Development Review*, 15, 61–76. [2]

SMITH, L. C. (2003): *The Importance of Women's Status for Child Nutrition in Developing Countries*, vol. 131, Intl Food Policy Res Inst. [2], [14], [19] SOKULLU, S. AND C. VALENTE (2017): "Individual Consumption in Collective Households: Identification Using Panel Data with an Application to PROGRESA," mimeo. [4]

SONG, S. (2015): "Semiparametric estimation of models with conditional moment restrictions in the presence of nonclassical measurement errors," *Journal of Econometrics*, 185, 95 – 109. [6]

SONG, S., S. M. SCHENNACH, AND H. WHITE (2015): "Estimating nonseparable models with mismeasured endogenous variables," *Quantitative Economics*, 6, 749–794. [6]

THOMAS, D. (1990): "Intra-household Resource Allocation: An Inferential Approach," *Journal of human resources*, 635–664. [2], [5]

——— (1994): "Like Father, like Son; Like Mother, like Daughter: Parental Resources and Child Height," *The Journal of Human Resources*, 29, 950–988. [2]

——— (1997): "Incomes, Expenditures, and Health outcomes: Evidence on Intrahousehold Resource Allocation," *Intrahousehold Resource Allocation in Developing Countries*, 142–64. [2]

TOMMASI, D. (2019): "Control of Resources, Bargaining Power and the Demand of Food: Evidence from PROGRESA," *Journal of Economic Behavior & Organization*, 161, 265–286. [4]

TOMMASI, D. AND A. WOLF (2018): "Estimating household resource shares: A shrinkage approach," *Economics Letters*, 163, 75 – 78. [23]

UPADHYAY, U. D., J. D. GIPSON, M. WITHERS, S. LEWIS, E. J. CIARALDI, A. FRASER, M. J. HUCHKO, AND N. PRATA (2014): "Women's Empowerment and Fertility: A Review of the Literature," *Social Science & Medicine*, 115, 111–120. [14]

URA, T. (2018): "Heterogeneous treatment effects with mismeasured endogenous treatment," *Quantitative Economics*, 9, 1335–1370. [2], [6], [9], [13]

VANDERWEELE, T. (2015): *Explanation in causal inference: methods for mediation and interaction*, Oxford University Press. [6]

VYTLACIL, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. [7]

WOLPIN, K. I. (2013): *The limits of inference without theory*, MIT Press. [7]

WORLD BANK (2014): "Women, Business and the Law 2014. Removing Restrictions to Enhance Gender Equality," *World Bank Report*. [24]

YANAGI, T. (2018): "Inference on local average treatment effects for misclassified treatment," *Econometric Reviews*, 0, 1–23. [6]

A Appendix

This Appendix contains five main sections. The proof of Theorem 1 is in Appendix A.1. Appendix A.2 provides a graphical illustration of the MR-LATE estimator. Appendix A.3 presents some additional results and robustness checks. Appendix A.4 summarizes the derivation of our model presented in section 4.1 for estimating resource shares from demand equations for private assignable goods. In Appendix A.5, we present a brief discussion of the external validity of our results. Additional figures and tables are in Appendix A.6.

A.1 Proof of Theorem 1

Substituting equation (1) into equation (4) gives

$$T = T_0 + (T_1 - T_0)D = T_0 + (T_1 - T_0)[(1-Z)D_0 + ZD_1] \tag{A1}$$

Multiplying equation (2) by equation (A1) gives

$$YT = [Y_0 + (Y_1 - Y_0)[(1-Z)D_0 + ZD_1]][T_0 + (T_1 - T_0)[(1-Z)D_0 + ZD_1]]$$

Using assumption 2-i, this makes

$$\begin{aligned} E(YT \mid Z=1) &= E[[Y_0 + (Y_1 - Y_0)D_1][T_0 + (T_1 - T_0)D_1]] \\ &= E[T_0Y_0 + (Y_1T_1 - Y_0T_0)D_1] \end{aligned}$$

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

$$E(YT \mid Z=0) = E[T_0Y_0 + (Y_1T_1 - Y_0T_0)D_0]$$

So,

$$E(YT \mid Z=1) - E(YT \mid Z=0) = E[(Y_1T_1 - Y_0T_0)(D_1 - D_0)]$$

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 $\Pr(D_1 - D_0 = 1) = E[D_1 - D_0]$. We therefore apply the standard LATE logic:

$$\begin{aligned} E(YT \mid Z=1) - E(YT \mid Z=0) &= E[(Y_1T_1 - Y_0T_0)(D_1 - D_0)] \\ &= E[Y_1T_1 - Y_0T_0 \mid D_1 - D_0 = 1]\Pr(D_1 - D_0 = 1) \\ &= E(Y_1T_1 - Y_0T_0 \mid C)E(D_1 - D_0). \end{aligned}$$

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

$$E(YT \mid Z=1) - E(YT \mid Z=0) = E(p_1Y_1 - p_0Y_0 \mid C)E(D_1 - D_0).$$

Replacing Y with one gives

$$\begin{aligned} E(T \mid Z=1) - E(T \mid Z=0) &= E(p_1 - p_0 \mid C)E(D_1 - D_0) \\ &= (p_1 - p_0)E(D_1 - D_0). \end{aligned}$$

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 (7), thereby proving the Theorem.

A.2 A Graphical Illustration

Assume that $supp(\varepsilon) \subset \left[\kappa^b - e, \kappa^a - e\right)$. 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 A4 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 $q^a = p_1^a / \left(p_1^a - p_0^a\right) = 1$ and $q^b = p_1^b / \left(p_1^b - p_0^b\right) = 0$. 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.

A.3 Additional Results: Alternative Measures of Power

In this section, we first empirically 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. We then present MR-LATE estimates that are instead based on an index of power constructed from these self-reported proxies, which are available in the NFHS dataset. Finally, we explore the possibility of a continuous relationship between R and the health outcomes of family members

Comparing Self-reported and Structural Measures of Power. 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".[40] 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

[40] We exclude women who answer "other/someone else" (less than 1 percent in any question).

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. Figure A6 in the Appendix shows the distribution of this index.

Panels (a) to (c) of Figure A5 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 A9 (in the Appendix). 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.[41] 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.

MR-LATE Estimates Based on Self-reported Measures of Power. Here we consider estimation of treatment effects where the treatments are women's self-reported measures of power. As discussed in Example 1 in section 3.3, we could discard all the observations where responses to power questions are ambiguous, and apply standard LATE estimation to the remaining observations. But if giving other answers is correlated with the instrument or with potential outcomes, the resulting LATE estimates could then be biased in unknown ways. And even if a resulting LATE estimate is not biased, discarding observations would entail a loss of efficiency. So instead of discarding any observations, we use MR-LATE estimation to make use of these incompletely observed treatments. This is done by setting $T^a = 1$ if the answer is "the wife," otherwise $T^a = 0$, and similarly letting $T^b = 1$ if the answer is "the husband," otherwise $T^b = 0$. Thus the difference between T^a and $1 - T^b$ is in how the ambiguous responses are handled.

Following Example 2 of section 3.3, we can also construct MR-LATE estimates based on an index constructed by summing the self-reported power responses. 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.

Using the notation of Example 2, if we took $\kappa^a = k$ for some integer k, with $k < 6$, we would be assuming that it is extremely unlikely that the husband really has most of the power if the household reports that the wife makes $6 - k$ or more of the decisions (meaning p_0^a is very close to zero). Similarly, taking $\kappa^b = -k$ for some small integer k means assuming it is extremely unlikely that the wife has most of the power if they report that the husband makes $6 - k$ or more of the decisions (meaning p_1^b is very close to zero).

Table A10 and A11 report the MR-LATE estimates for models where T^a and T^b are constructed as described in the previous paragraphs, following Examples 1 and 2 of section 3.3, respectively.[42] The empirical results of these alternative estimates appear unreliable, with peculiar magnitudes and some

[41] We also repeated this exercise using principal components of the self reported responses. The findings do not change. Results are available upon request.

[42] For save space, we report results for the NFHS women's sample only. Results for men's and children's samples are available upon request.

large standard errors. The problem with these estimates are that self reports of power appear to be rather noisy indicators of true power, and as a result, the treatment indicators T, T^a, and T^b based on these self reports are not significantly correlated with the instrument Z. This can be seen in the first-stage MR-LATE F-statistics, which are almost all below 10. So, while the results here illustrate alternative ways that MR-LATE estimates can be constructed and used, we find that more stable and reliable treatment effect estimates are obtained when treatment is defined using our structural model estimates of women's control over resources.

Linear Model. In section 4 we discussed why we expect Y to depend on D. Here we consider the alternative possibility of Y depending continuously on R^*. We do this by linearly regressing each health status measure Y_i on a constant, R_i, and X_i, using 2SLS where Z_i is the excluded instrument for R_i. These estimates should be interpreted cautiously, since we have no reason to expect that the true relationship of Y to R^* and X is linear, or that estimation errors in R relative to R^* satisfy the classical measurement error assumptions needed for validity of linear 2SLS estimation. But if these assumptions do hold, then the estimated coefficient of R in this regression will be a consistent estimate of the average marginal effect of R^* on Y.

Tables A12 and A13 in the Appendix contain the estimation results of these linear regressions for women's and children's health outcomes.[43] For women, we restrict the sample to those living in households where the woman has control of 40 to 60 percent of the resources, which is where we see a positive and significant correlation between R and Z (see columns 1 and 2). In general, while the estimated coefficients of these regressions have the expected signs, we do not find marginal effects that are statistically different from zero. While the correlation between R and Z is statistically significant, based on low first-stage F-statistics we cannot rule out the possibility of a weak instrument problem. The lower significance of these estimates relative to our main MR-LATE results also suggests that the linear model may be misspecified.

A.4 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

$$\max_{Q_1,...,Q_T,S_1,...S_T} \widetilde{V}\left[V_1(Q_1,S_1,X),...,V_T(Q_T,S_T,X) \mid \widetilde{Z},X,P/M\right] \tag{A2}$$

[43] We do not report results for the men's sample, because (possibly due to the smaller sample size), no significant first stage estimates could be obtained using the men's sample.

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

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

$$\max_{Q_t, S_t} V_t(Q_t, S_t, X) \quad \text{such that} \quad \eta_t(P, M, X, \widetilde{Z})M = P_S' A(X) S_t + P_t Q_t \tag{A3}$$

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 $P_S A(X)$ 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 (A3) is an ordinary utility function maximized under a linear budget constraint (linear in shadow prices and a shadow budget), the solution to equation (A3) 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

$$Q_t = h_t\left(\eta_t(P,M,X,\widetilde{Z})M,P,X\right) \quad \text{for} \quad t = 1,...,T. \tag{A4}$$

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 (A4). 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 theoretical evidence that η_t does not depend on M.[44] This allows them to rewrite equation (A4) 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.

A.5 External Validity of LATE

Here we use a latent variable model and the marginal treatment effects framework developed by Björklund and Moffitt (1987), Heckman and Vytlacil (1999, 2005, 2007), Carneiro et al. (2011), Brinch et al. (forthcoming) and Kowalski (2016) to shed light on the external validity of our results.[45] Specifically, we wish to clarify the relationship between the local average treatment effect (LATE) and the average treatment effect (ATE) in our empirical application.

In our application, we are interested in estimating the treatment effect of D on individual health outcomes Y, with covariates (observable individual and household characteristics) X. Given potential outcomes Y_1 and Y_0, define the functions h_1 and h_0, and corresponding errors U_0 and U_1, by $h_0(X) = E(Y_0 \mid X)$, $h_1(X) = E(Y_1 \mid X)$, $U_0 = Y_0 - h_0(X)$, and $U_1 = Y_1 - h_1(X)$.

Under the standard monotonicity assumption for LATE estimation, the determination of treatment D can be represented by standard threshold crossing model

$$D = \mathbb{I}(R^* - e \geq 0)$$

where R^* is an underlying latent variable and e is an unobserved threshold that can vary across households.

[44] 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.

[45] Brinch et al. (forthcoming) in particular show how a discrete (binary) instrument can be used to identify the MTE under functional structure that allows for treatment heterogeneity among individuals with the same observed characteristics and self-selection based on the unobserved gain from treatment.

It is assumed that $(U_0, U_1, e) \perp Z \mid X$, which implies validity of the instrument Z. Assume also that e and R^* are continuously distributed. Instrument relevance requires that $E(R^* \mid X, Z)$ varies with Z. In our application, R^* is a continuous measure of women's control of household resources and Z is a binary variable capturing women's exposure to inheritance law reforms that improved their ability to inherit property.

The threshold crossing model for D means that households are treated if their unobserved threshold e is less than or equal to R^*. Variation in e can be interpreted as meaning that different households have different levels of R^* that are needed for the wife to have substantial control, making $D = 1$. Under monotonicity, in households where e is low enough ($e \leq R^*_L$), wives control substantial resources even if they are not exposed to the plausibly exogenous changes in women's inheritance rights (that is, they are *always-takers*, with $D = 1$ and $Z = 0$). In households where e is high enough ($e > R^*_H$), husbands control substantial resources even if their wives are exposed to the reforms (that is, they are *never-takers*, with $D = 0$ and $Z = 1$). Under the standard no-defiers assumption, the remaining households correspond to the group of *compliers* ($R^*_L < e \leq R^*_H$), whose treatment status is determined by women's exposure to the inheritance law amendments ($D = Z$).

As in Heckman and Vytlacil (1999), define the marginal treatment effect as the effect on the marginal individual entering treatment. That is, MTE is the average impact for the marginal individual receiving treatment among those with $e = R^*$, so

$$MTE(R^*) = E(Y_1 - Y_0 \mid e = R^*) \tag{A5}$$

We may without loss of generality normalize e to be uniformly distributed, by taking a suitable monotonic transform of R^* and e. Under the normalization $e \sim U(0, 1)$, Heckman and Vytacil show that LATE equals the weighted average of MTE over the interval $(R^*_L, R^*_H]$, with weights equal to $\frac{1}{R^*_H - R^*_L}$. A sufficient condition for LATE to be externally globally valid (so LATE equals ATE) is if $MTE(R^*)$ is constant for all R^*.

Following Kowalski (2016) (analogous tests are proposed in Angrist (2004), Bertanha and Imbens (2014), and Brinch et al. (forthcoming)), we implement a test of global external validity using a difference-in-difference regression.[46] Since we cannot directly observe R^* and D, we perform the test using R and $T = \mathbb{I}(R \geq 50)$ instead. This test must therefore be interpreted with caution, since it does not account for the measurement errors that our MR-LATE estimator is designed for. To implement the test, we regress individuals' health outcomes Y on the covariates X in the sample of households where women are not exposed to the inheritance law reforms (i.e., $Z = 0$). Then, using the estimated coefficients, we predict outcomes $\widehat{Y}$ for all individuals with $Z = 1$ and $Z = 0$, and estimate the following model:

$$\widehat{Y} = \lambda_{TZ} TZ + \lambda_T T + \lambda_Z Z + \lambda \tag{A6}$$

Ignoring the measurement error from using R and T in place of R^* and D, if $MTE(R^*)$ is constant for all R^* (implying external validity) then $\lambda_{TZ} = 0$. To implement the test we assume $MTE(R^*)$ is linear in R^* for simplicity, and we compute standard errors based on 200 bootstrap replications. Table A15 reports the estimates for λ_{TZ} and the corresponding bootstrap standard errors. For all health outcomes Y, in the three NFHS samples (women, men, and children), we cannot reject $\lambda_{ZT} = 0$. So under the caveat that the test is based on T, we do not reject the hypothesis that our LATE is externally valid, meaning that the treatment effects we identify in our empirical application equal ATE, and do not depend on our specific choice of instrument.

[46] We implement the test using the `mtebinary` Stata routine (Kowalski et al. (2016)).

A.6 Additional Figures and Tables

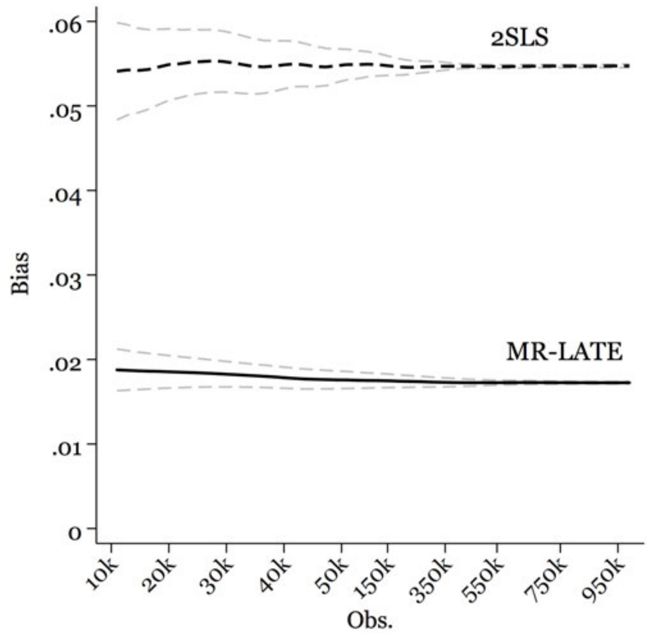
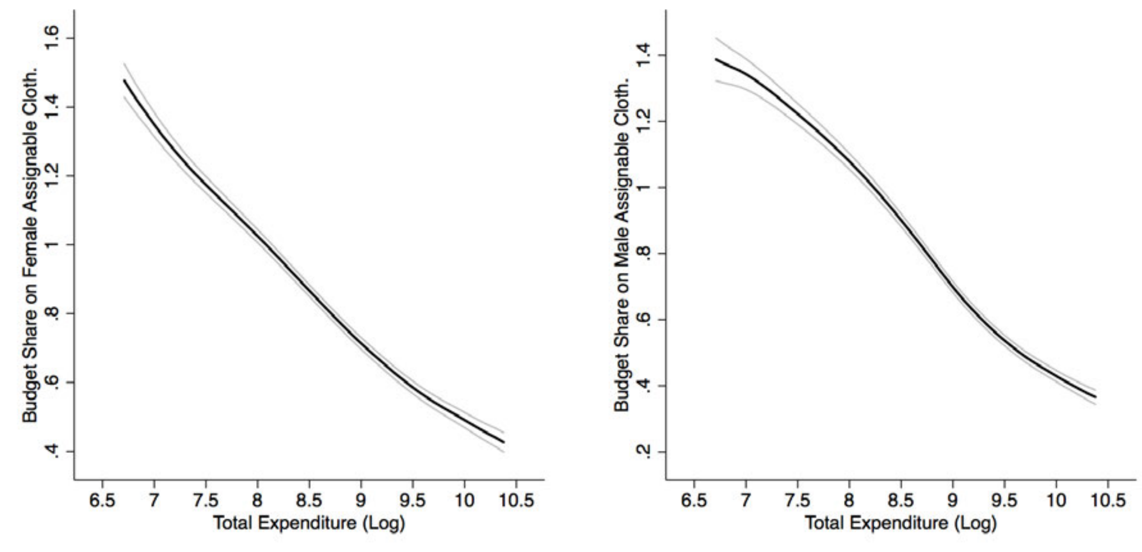


Figure A1: 2SLS vs. MR-LATE when ϵ is Unbounded and $\kappa = 0.05$



(A) Woman's Assignable Clothing

(B) Man's Assignable Clothing

Figure A2: Non-parametric Engel Curves

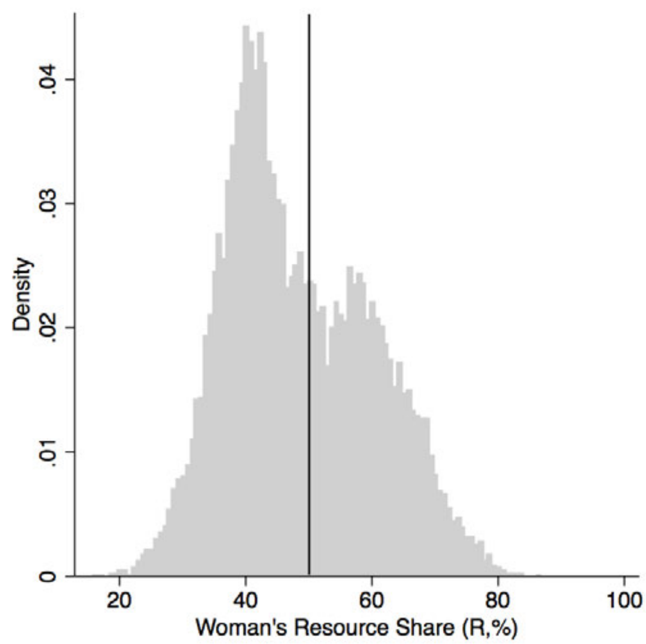


Figure A3: Estimated of Women's Resource Shares (NFHS Sample)

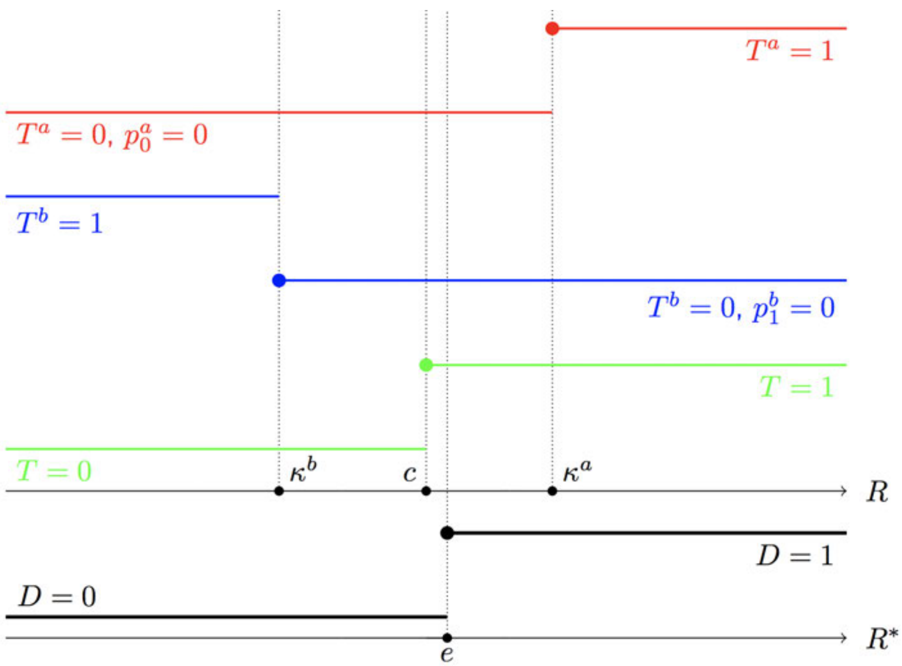
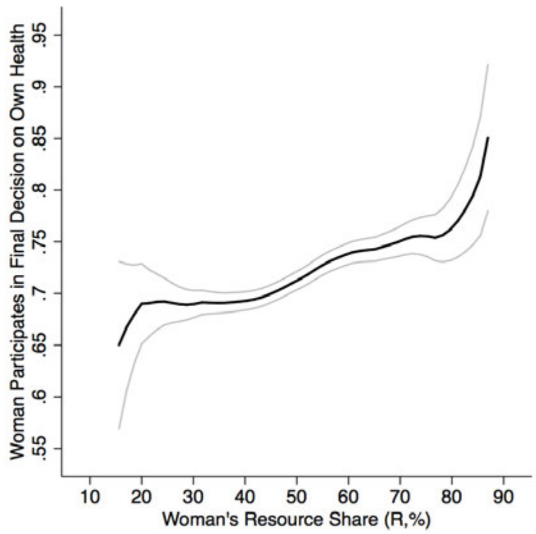
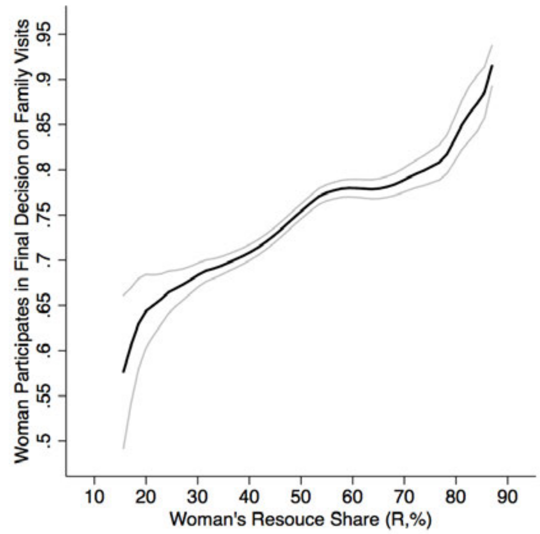


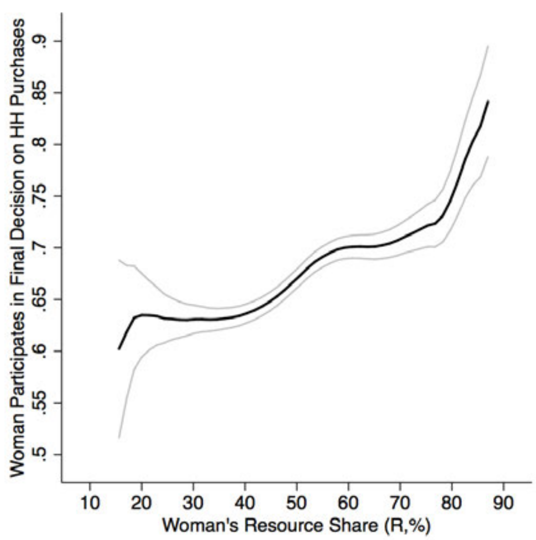
Figure A4: Illustrative Example



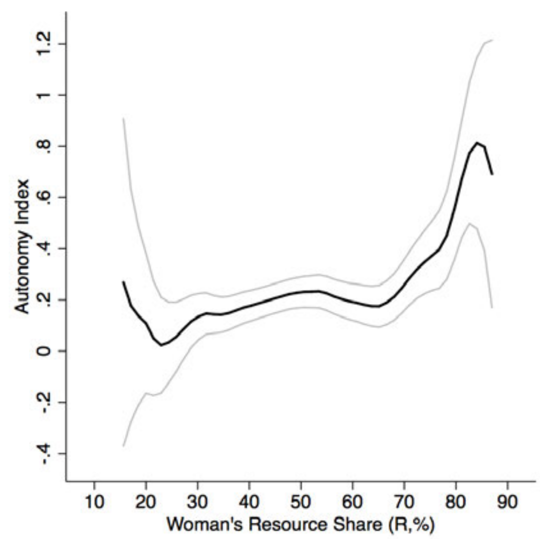
(A) Woman's Health



(B) Visits to Family and Relatives



(C) Household Purchases



(D) Autonomy Index

Figure A5: Structurally Recovered Bargaining Power and Household Decision Making

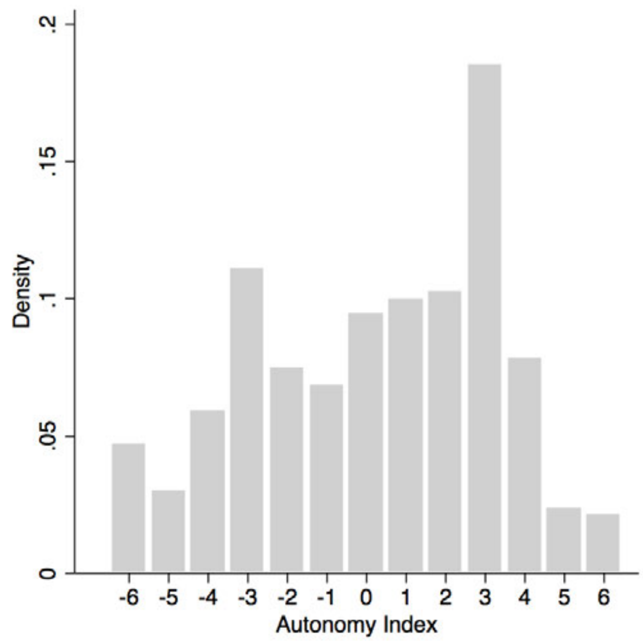


Figure A6: Autonomy Index (NFHS Sample)

Note: This index is based on women's response to 3 questions about decision making and 3 questions about mobility. We give each answer a value of one if the response is that wife controls that decision (or if she can go alone to places), a value of minus one if the response is that the husband controls that decision (or if she cannot go alone to places), and a value of zero for any other response. We then define our estimated index R to equal the sum of the responses across all of the questions.

Table A1: $|\text{Bias}_{\text{B-LATE}}|$

Panel A: $p_1^a = 0.9, p_0^b = 0.9$					
$p_1^b \downarrow \quad \| \quad p_0^a \rightarrow$	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: $p_1^a = 0.8, p_0^b = 0.8$					
$p_1^b \downarrow \quad \| \quad p_0^a \rightarrow$	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: $p_1^a = 0.7, p_0^b = 0.7$					
$p_1^b \downarrow \quad \| \quad p_0^a \rightarrow$	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

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

Table A2: $|\text{Bias}_{\text{MR-LATE}}|$

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 A3: OLS vs. 2SLS vs. MR-LATE when ϵ is Bounded at Known κ

	Panel A: N = 1,000								
	$\kappa = 0$			$\kappa = 5$			$\kappa = 10$		
	OLS	2SLS	MR-LATE	OLS	2SLS	MR-LATE	OLS	2SLS	MR-LATE
T	1.016	0.992	0.992	0.827	1.030	1.015	0.656	1.132	1.030
sd	0.048	0.320	0.320	0.057	0.342	0.286	0.065	0.393	0.309
Bias	0.016	-0.008	-0.008	-0.173	0.030	0.015	-0.344	0.132	0.030
MSE	0.003	0.103	0.103	0.033	0.118	0.082	0.123	0.172	0.096
	Panel B: N = 10,000								
	$\kappa = 0$			$\kappa = 5$			$\kappa = 10$		
	OLS	2SLS	MR-LATE	OLS	2SLS	MR-LATE	OLS	2SLS	MR-LATE
T	1.015	0.999	0.999	0.827	1.032	1.016	0.658	1.122	1.028
sd	0.014	0.097	0.097	0.018	0.105	0.085	0.020	0.119	0.094
Bias	0.015	-0.001	-0.001	-0.173	0.032	0.016	-0.342	0.122	0.028
MSE	0.000	0.009	0.009	0.030	0.012	0.007	0.118	0.029	0.010
	Panel C: N = 500,000								
	$\kappa = 0$			$\kappa = 5$			$\kappa = 10$		
	OLS	2SLS	MR-LATE	OLS	2SLS	MR-LATE	OLS	2SLS	MR-LATE
T	1.015	0.999	0.999	0.826	1.028	1.016	0.657	1.117	1.026
sd	0.002	0.014	0.014	0.002	0.015	0.012	0.003	0.017	0.013
Bias	0.015	-0.001	-0.001	-0.174	0.028	0.016	-0.343	0.117	0.026
MSE	0.000	0.000	0.000	0.030	0.001	0.000	0.118	0.014	0.001

Notes: In each Panel, the true value is set to 1.000. Results are based on 1,000 simulations for varying number of observations N. We simulate various measurement errors and the availability of a strong and exogenous instrument. When $\kappa = 0$ means that there is no measurement error. Whereas, measurement error with $\kappa = 0.05$ (0.10) means that we estimate R that are on average +/- 5 (10) percent of the true value. This means that roughly 10 (20) percent of sample is wrongly observed to either treatment or control.

Table A4: OLS vs. 2SLS vs. MR-LATE when ϵ is Unbounded

	Panel A: N = 1,000					
			MR-LATE			
	OLS	2SLS	$\kappa = 0$	$\kappa = 0.01$	$\kappa = 0.05$	$\kappa = 0.1$
D	0.780	1.063	1.063	1.042	1.021	1.031
sd	0.020	0.109	0.109	0.105	0.091	0.088
Bias	-0.220	0.063	0.063	0.042	0.021	0.031
MSE	0.049	0.016	0.016	0.013	0.009	0.009
	Panel B: N = 10,000					
			MR-LATE			
	OLS	2SLS	$\kappa = 0$	$\kappa = 0.01$	$\kappa = 0.05$	$\kappa = 0.1$
D	0.780	1.063	1.063	1.042	1.021	1.031
sd	0.020	0.109	0.109	0.105	0.091	0.088
Bias	-0.220	0.063	0.063	0.042	0.021	0.031
MSE	0.049	0.016	0.016	0.013	0.009	0.009
	Panel C: N = 500,000					
			MR-LATE			
	OLS	2SLS	$\kappa = 0$	$\kappa = 0.01$	$\kappa = 0.05$	$\kappa = 0.1$
D	0.780	1.063	1.063	1.042	1.021	1.031
sd	0.020	0.109	0.109	0.105	0.091	0.088
Bias	-0.220	0.063	0.063	0.042	0.021	0.031
MSE	0.049	0.016	0.016	0.013	0.009	0.009
	Panel D: N = 1,000,000					
			MR-LATE			
	OLS	2SLS	$\kappa = 0$	$\kappa = 0.01$	$\kappa = 0.05$	$\kappa = 0.1$
D	0.780	1.063	1.063	1.042	1.021	1.031
sd	0.020	0.109	0.109	0.105	0.091	0.088
Bias	-0.220	0.063	0.063	0.042	0.021	0.031
MSE	0.049	0.016	0.016	0.013	0.009	0.009

Notes: In each Panel, the true value is set to 1.000. Results are based on 1,000 simulations for varying number of observations N. We simulate a large measurement error and the availability of a strong and exogenous instrument.

Table A5: 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	7,480	0.91	0.77	0.74				
Man's Assignable Clothing Budget Share	7,480	0.92	0.70	0.87				
Total Expenditure (Rupees)	7,480	5968.30	4179.26	5064.75				
Number of Children	7,480	2.16	2.00	0.93	23,699	2.20	2.00	0.95
𝕀(1 child)	7,480	0.27	0.00	0.44	23,699	0.26	0.00	0.44
𝕀(2 children)	7,480	0.41	0.00	0.49	23,699	0.39	0.00	0.49
𝕀(3 children)	7,480	0.23	0.00	0.42	23,699	0.24	0.00	0.42
𝕀(4 children)	7,480	0.10	0.00	0.30	23,699	0.11	0.00	0.31
Fraction of Female Children	7,480	0.45	0.50	0.37	23,699	0.47	0.50	0.37
Woman's Age	7,477	30.96	30.00	6.46	23,697	30.22	30.00	7.32
Gender Age Gap (Man - Woman)	7,473	5.20	5.00	4.27	23,662	5.80	5.00	5.01
Children's Avg. Age	7,480	6.87	7.00	3.55	23,699	6.46	6.33	3.58
𝕀(Hindu, Buddhist, Jain, Sikh)	7,480	0.80	1.00	0.40	23,695	0.77	1.00	0.42
𝕀(Sch. Caste, Sch. Tribe, Oth. Back. Caste)	7,480	0.64	1.00	0.48	22,808	0.68	1.00	0.47
𝕀(Own Land)	7,447	0.69	1.00	0.46	23,694	0.35	0.00	0.48
𝕀(Woman Completed High School)	7,480	0.14	0.00	0.35	23,699	0.09	0.00	0.28
𝕀(Man Completed High School)	7,480	0.22	0.00	0.42	23,699	0.13	0.00	0.34
𝕀(Rural)	7,480	0.48	0.00	0.50	23,699	0.54	1.00	0.50
𝕀(North)	7,480	0.30	0.00	0.46	23,699	0.30	0.00	0.46
𝕀(East)	7,480	0.21	0.00	0.40	23,699	0.16	0.00	0.37
𝕀(North-East)	7,480	0.14	0.00	0.35	23,699	0.20	0.00	0.40
𝕀(South)	7,480	0.22	0.00	0.41	23,699	0.21	0.00	0.41
𝕀(West)	7,480	0.12	0.00	0.33	23,699	0.13	0.00	0.34

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. 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.

Table A6: 2005-2006 NFHS Individual Data

	Women ($N = 19,738$)			Men ($N = 10,785$)			Children ($N = 15,038$)		
	Mean	Median	St. Dev.	Mean	Median	St. Dev.	Mean	Median	St. Dev.
Body Mass Index (BMI)	21.00	20.24	3.97	21.27	20.73	3.54			
𝕀(BMI≤18.5)	0.29	0.00	0.46	0.24	0.00	0.42			
𝕀(Anemic)	0.15	0.00	0.36	0.09	0.00	0.28			
Weight for Age (z-score)							-1.75	-1.83	1.19
Height for Age (z-score)							-1.60	-1.62	1.55
𝕀(Cough in last 2 weeks)							0.18	0.00	0.38
𝕀(Fever in last 2 weeks)							0.15	0.00	0.36
𝕀(Diarrhea in last 2 weeks)							0.09	0.00	0.28
𝕀(Any Vaccination)							0.91	1.00	0.28
𝕀(HSA Exposed)	0.18	0.00	0.39	0.28	0.00	0.45	0.18	0.00	0.38
𝕀(1 child)	0.24	0.00	0.43	0.24	0.00	0.42	0.20	0.00	0.40
𝕀(2 children)	0.40	0.00	0.49	0.41	0.00	0.49	0.39	0.00	0.49
𝕀(3 children)	0.24	0.00	0.43	0.24	0.00	0.43	0.27	0.00	0.45
𝕀(4 children)	0.11	0.00	0.32	0.11	0.00	0.31	0.14	0.00	0.34
Fraction of Female Children	0.47	0.50	0.36	0.47	0.50	0.36	0.50	0.50	0.36
Women's Age	29.58	29.00	5.86	29.44	29.00	5.71	26.31	26.00	4.45
Gender Age Gap (Men - Women)	5.84	5.00	4.28	5.65	5.00	3.68	5.58	5.00	4.22
Children's Avg. Age	6.39	6.25	3.53	6.37	6.00	3.58	3.57	3.33	2.05
𝕀(Hindu, Buddhist, Jain, Sikh)	0.79	1.00	0.41	0.78	1.00	0.41	0.74	1.00	0.44
𝕀(Sch. Caste, Sch. Tribe, Oth. Back. Caste)	0.68	1.00	0.47	0.72	1.00	0.45	0.71	1.00	0.45
𝕀(Own Land)	0.36	0.00	0.48	0.34	0.00	0.47	0.34	0.00	0.47
𝕀(Woman Completed High School)	0.08	0.00	0.27	0.09	0.00	0.28	0.07	0.00	0.25
𝕀(Man Completed High School)	0.13	0.00	0.34	0.13	0.00	0.34	0.11	0.00	0.31
𝕀(Rural)	0.54	1.00	0.50	0.52	1.00	0.50	0.56	1.00	0.50
𝕀(North)	0.30	0.00	0.46	0.25	0.00	0.43	0.33	0.00	0.47
𝕀(East)	0.16	0.00	0.37	0.10	0.00	0.30	0.16	0.00	0.37
𝕀(North-East)	0.20	0.00	0.40	0.19	0.00	0.39	0.23	0.00	0.42
𝕀(South)	0.21	0.00	0.41	0.32	0.00	0.47	0.17	0.00	0.38
𝕀(West)	0.13	0.00	0.33	0.14	0.00	0.34	0.11	0.00	0.32
𝕀(Child is Female)							0.48	0.00	0.50
Child's Age							2.18	2.00	1.39

Notes: 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). 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.

Table A7: Bounds κ^a and κ^b

	Women		Men		Children	
	κ^a	κ^b	κ^a	κ^b	κ^a	κ^b
$\mathscr{K}=0$	50.00	50.00	50.00	50.00	50.00	50.00
$\mathscr{K}=1$	50.20	49.80	50.19	49.83	50.18	49.82
$\mathscr{K}=5$	51.04	48.95	51.05	49.04	50.92	49.09
$\mathscr{K}=10$	52.22	47.99	52.30	48.12	51.98	48.16
$\mathscr{K}=20$	54.73	45.94	54.99	46.10	54.26	46.18

Note: NFHS data.

Table A8: First Stage Estimates

	Women		Men		Children	
	T^a	T^b	T^a	T^b	T^a	T^b
	Panel A: $\mathscr{K}=0$					
$\mathbb{I}$(HSA)	0.0781	-0.0781	0.0706	-0.0706	0.0778	-0.0778
	(0.0121)	(0.0164)	(0.0137)	(0.0161)	(0.0085)	(0.0258)
First Stage F-stat.	58.9640	58.9640	37.2133	37.2133	13.8997	13.8997
	Panel B: $\mathscr{K}=1$					
$\mathbb{I}$(HSA)	0.0790	-0.0828	0.0692	-0.0763	0.0695	-0.0816
	(0.0117	(0.0122)	(0.0113)	(0.0136)	(0.0292)	(0.0217)
First Stage F-stat.	61.4723	65.7576	36.4079	42.6378	10.4811	15.3693
	Panel C: $\mathscr{K}=5$					
$\mathbb{I}$(HSA)	0.0719	-0.0806	0.0692	-0.0732	0.0705	-0.0892
	(0.0096)	(0.0117)	(0.0117)	(0.0138)	(0.0187)	(0.0193)
First Stage F-stat.	52.4354	59.4068	38.6722	37.4487	14.3894	15.2712
	Panel D: $\mathscr{K}=10$					
$\mathbb{I}$(HSA)	0.0573	-0.0633	0.0496	-0.0598	0.0655	-0.0743
	(0.0109)	(0.0134)	(0.0119)	(0.0137)	(0.0108)	(0.0196)
First Stage F-stat.	34.7818	35.3245	21.1301	23.6691	13.5925	12.3206
	Panel E: $\mathscr{K}=20$					
$\mathbb{I}$(HSA)	0.0259	-0.0658	0.0156	-0.0602	0.0606	-0.0809
	(0.0072)	(0.0128)	(0.0075)	(0.0186)	(0.0193)	(0.0222)
First Stage F-stat.	7.0816	34.3959	2.1093	21.7688	15.7231	15.9062

Note: NFHS data. 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).

Table A9: 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
$T = \mathbb{I}(R \geq 50)$	0.0245	0.0303	0.0195	0.228
	(0.0147)	(0.0125)	(0.0160)	(0.0994)
Mean Dependent Variable	0.6642	0.7130	0.7400	65.8703

Note: NFHS data. The sample includes married women of age 15 to 49 in nuclear households with up to 4 children. 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).

Table A10: Self-reported Measures of Power and Health: Example 1

	𝕀(Woman Participates in Final Decisions on)								
	Own Health			Household Purchases			Visits to Family and Relatives		
	BMI	Pr(BMI≤18.5)	Pr(Anemic)	BMI	Pr(BMI≤18.5)	Pr(Anemic)	BMI	Pr(BMI≤18.5)	Pr(Anemic)
D	154.4886	-14.4061	-7.3242	77.0713	-10.5959	-8.0653	1.6255	-0.3637	0.6134
	(181.9458)	(6.4603)	(4.5583)	(161.4285)	(4.8256)	(3.7369)	(28.1084)	(0.6721)	(0.4610)
First stage F-statistics:									
T^a	0.1570	0.1570	0.1570	4.0934	4.0934	4.0934	1.1688	1.1688	1.1688
T^b	0.0125	0.0125	0.0125	0.0218	0.0218	0.0218	4.8848	4.8848	4.8848

Note: Estimates obtained with NFHS data and the MR-LATE estimator. The sample includes married women of age 15 to 49 in nuclear households with up to 4 children. 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). D is the true, unobserved treatment: 𝕀(Woman Participates in Final Decisions on [X]). $T^a = 1$ if the answer is "the wife," otherwise $T^a = 0$. Similarly, one can let $T^b = 1$ if the answer is "the husband," otherwise $T^b = 0$.

Table A11: Self-reported Measures of Power and Health: Example 2

	BMI	Pr(BMI≤18.5)	Pr(Anemic)
D ($k=0$)	18.1815	-1.7050	-1.0356
	(6.7474)	(0.5801)	(0.3344)
First stage F-statistics:			
T^a	3.9385	3.9385	3.9385
T^b	3.9385	3.9385	3.9385
D ($k=1$)	20.6218	-2.6167	-1.0582
	(20.4949)	(1.4365)	(0.4026)
First stage F-statistics:			
T^a	3.6756	3.6756	3.6756
T^b	0.5468	0.5468	0.5468
D ($k=2$)	20.3989	-2.1133	-0.8088
	(19.1140)	(0.9096)	(0.4782)
First stage F-statistics:			
T^a	1.6341	1.6341	1.6341
T^b	1.3387	1.3387	1.3387
D ($k=3$)	41.1914	-3.3276	-2.5234
	(86.8981)	(1.0959)	(1.0992)
First stage F-statistics:			
T^a	0.1286	0.1286	0.1286
T^b	2.1703	2.1703	2.1703

Note: Estimates obtained with NFHS data and the MR-LATE estimator. The sample includes married women of age 15 to 49 in nuclear households with up to 4 children. 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). D is the true, unobserved treatment $D = \mathbb{I}(R^* \geq e)$. T^a and T^b are based on the answers to six questions about women's participation in household decision and women's mobility. $T^a = \mathbb{I}(R \geq \kappa^a)$, with $\kappa^a = k$, and $T^b = \mathbb{I}\big(R < \kappa^b\big)$, with $\kappa^b = -k$.

Table A12: Woman's Health: Linear Model

	First Stage		Second Stage		
	Woman's Resource Share (R)		BMI	Pr(BMI≤18.5)	Pr(Anemic)
$\mathbb{I}$(HSA)	-0.0337	0.0702			
	(0.0443)	(0.0278)			
Woman's Resource Share (R, %)			2.962	-0.954	-0.608
			(3.048)	(0.408)	(0.318)
Observations	19,738	10,765	10,765	10,765	10,765

Note: NFHS data. 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). Column 1: the sample includes married women of age 15 to 49 in nuclear households with up to 4 children. Column 2 to 5: the sample includes married women of age 15 to 49 in nuclear households with up to 4 children with $R \in (40, 60)$.

Table A13: Children's Health: Linear Model

	First Stage	Second Stage					
	Woman's Res. Share (R)	W-for-age (z-score)	H-for-age (z-score)	Pr (Cough)	Pr (Fever)	Pr (Diarrhea)	Pr (Vacc.)
$\mathbb{I}$(HSA)	0.107						
	(0.0501)						
Woman's Resource Share (R, %)		1.481	2.037	-0.483	-0.445	-0.330	-0.0938
		(0.909)	(1.154)	(0.297)	(0.274)	(0.201)	(0.219)

Note: NFHS data. 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). The sample includes children 0 to 5 in nuclear households with up to 4 children.

Table A14: Engel Curves Estimation Results (NSS Sample)

	$R(X)$	$\widehat{\delta}_w(X)$	$\widehat{\delta}_m(X)$	$\widehat{\beta}(X)$
$\mathbb{1}$(2 children)	-0.0554	-0.2660	-0.2260	0.0260
	(0.0166)	(0.3370)	(0.3350)	(0.0409)
$\mathbb{1}$(3 children)	-0.0395	-0.5610	-0.3470	0.0455
	(0.0211)	(0.3880)	(0.3860)	(0.0471)
$\mathbb{1}$(4 children)	-0.0824	0.0203	0.1510	-0.0180
	(0.0270)	(0.5900)	(0.5920)	(0.0721)
Fraction of Female Children	-0.0146	-0.4530	-0.5150	0.0544
	(0.0178)	(0.3600)	(0.3620)	(0.0437)
Gender Age Gap (Man - Woman)	0.0618	1.6710	1.3630	-0.1110
	(0.1320)	(2.6590)	(2.7870)	(0.3390)
Woman's Age	-0.571	2.3910	1.6600	-0.1510
	(0.1200)	(3.1000)	(3.1200)	(0.3810)
Children's Avg. Age	-0.1570	-3.4250	-2.6360	0.2270
	(0.2410)	(5.3780)	(5.3550)	(0.6520)
$\mathbb{1}$(Hindu, Buddhist, Jain, Sikh)	0.0951	1.520	1.038	-0.141
	(0.0214)	(0.3430)	(0.3480)	(0.0418)
$\mathbb{1}$(Sch. Caste, Sch. Tribe, Oth. Back. Caste)	-0.0313	-0.0360	-0.0002	-0.0180
	(0.0158)	(0.3090)	(0.3130)	(0.0380)
$\mathbb{1}$(Own Land)	0.0060	-0.1300	-0.0490	0.0299
	(0.0167)	(0.3270)	(0.3250)	(0.0393)
$\mathbb{1}$(Woman Completed High School)	0.0610	-0.2490	-0.2380	0.0368
	(0.0259)	(0.4840)	(0.4860)	(0.0563)
$\mathbb{1}$(Man Completed High School)	0.0254	0.0243	-0.2200	0.0272
	(0.0207)	(0.4020)	(0.4060)	(0.0477)
$\mathbb{1}$(Rural)	-0.0115	1.545	1.740	-0.194
	(0.0155)	(0.3200)	(0.3200)	(0.0390)
$\mathbb{1}$(North)	-0.0588	0.1130	0.987	-0.0423
	(0.0270)	(0.5250)	(0.5230)	(0.0630)
$\mathbb{1}$(East)	0.128	0.2750	0.0353	-0.0094
	(0.0276)	(0.5440)	(0.5320)	(0.0656)
$\mathbb{1}$(North-East)	0.19	-1.505	-2.055	0.150
	(0.0321)	(0.5970)	(0.5810)	(0.0713)
$\mathbb{1}$(South)	-0.0631	1.521	1.216	-0.139
	(0.0263)	(0.5590)	(0.564)	(0.0681)
Constant	0.623	6.743	6.997	-0.718
	(0.0476)	(1.0060)	(1.0070)	(0.1240)

Note: NSS data. Robust standard errors in parenthesis. Age variables are are divided by 100 to ease computation.

Table A15: Difference-in-Difference Test for External Validity

	$\widehat{\lambda}_{TZ}$	Bootstrap St. Error
Women:		
Body Mass Index (BMI)	0.6920	1.0184
Pr(BMI≤18.5)	0.0070	0.0347
Pr(Anemic)	0.0190	0.0270
Men:		
Body Mass Index (BMI)	1.0120	1.3337
Pr(BMI≤18.5)	-0.0190	0.2031
Pr(Anemic)	-0.0790	0.0755
Children:		
Height-for-age (z-score)	1.4550	8.4582
Weight-for-age (z-score)	-0.0270	9.1306
Pr(Cough)	-0.5960	0.6306
Pr(Fever)	-0.1570	0.4526
Pr(Diarrhea)	-0.1310	0.2607
Pr(Any Vaccination)	-0.2210	0.3046

Note: NFHS data. Test based on Kowalski (2016) and Kowalski et al. (2016) and implemented in Stata with `mtebinary`. $\lambda_{TZ} = 0$ implies no treatment effect heterogeneity and global external validity (under the assumption that T is the correct treatment).