Economic Development, the Nutrition Trap and Metabolic Disease

Nancy Luke, Kaivan Munshi, Anu Mary Oommen and Swapnil Singh*

January 20, 2022

Abstract

This research provides a single explanation for two seemingly unrelated facts that have recently been documented in developing countries: (i) the weak association between nutritional status, which we measure by BMI or weight conditional on height, and income, and (ii) the elevated risk of metabolic disease – diabetes, hypertension, cardiovascular disease – among normal weight individuals. Our model is based on a set point for BMI that is adapted to local food supply in the pre-modern economy, but which subsequently fails to adjust to rapid economic change. During the process of development, some individuals thus remain at their low-BMI set point, despite the increase in their income (food consumption), while others who have escaped the nutrition trap (but are not necessarily overweight) are at increased risk of metabolic disease. The model and the underlying biological mechanism, which are validated with micro-data from India, Indonesia and Ghana can jointly explain inter-regional (Asia-Africa) differences in nutritional status and the prevalence of diabetes.

*Luke: Pennsylvania State University, nkl10@psu.edu. Munshi: Yale University, kaivan.munshi@yale.edu. Oomen: Christian Medical College, anuoommen@cmcvellore.ac.in. Singh: Bank of Lithuania and Kaunas University of Technology, ssingh@lb.lt. We are grateful to Jere Behrman, Anne Ferguson-Smith, Nita Forouhi, Seema Jayachandran, K.M. Venkat Narayan, Nigel Unwin and numerous seminar participants for their constructive comments. Johannes Maywald, Krithika Raghupathi, and Astha Vohra provided outstanding research assistance. Research support from the National Institutes of Health through grant R01-HD046940, Cambridge-INET, the Keynes Fund and the Newton Trust at the University of Cambridge, and the Agence Nationale de la Recherche (ANR) under the EUR Project ANR-17-EURE-0010 is gratefully acknowledged. We are responsible for any errors that may remain. The views expressed here are those of the authors and do not necessarily reflect the position of the Bank of Lithuania.

1 Introduction

Two recently documented facts run counter to the conventional wisdom that economic development leads to better health: first, the absence of a clear link between nutritional status and income in developing countries (Deaton, 2007; Swaminathan et al., 2019) and second, the increased prevalence of metabolic disease – diabetes, hypertension, cardiovascular disease – among *normal* weight individuals with economic development (Narayan, 2017). Using nationally representative data from India, we corroborate these findings. There is an extremely weak association between BMI or weight conditional on height and income in the lower half of the income distribution, where individuals are disproportionately underweight. In parallel, there is a discontinuous increase in the risk of metabolic disease at a BMI level that is well within the normal range.[1] Our primary objective in this paper is to develop and test a model with three ingredients – adaptation, mismatch, and a set point – that can explain these seemingly unrelated observations.

Two models of early-life adaptation or developmental plasticity have been proposed by evolutionary developmental biologists: (a) The *developmental constraints* model in which developing organisms in severely resource limited environments make immediate tradeoffs to protect critical functions and improve survival in early life (Barker, 1995). (b) The *predictive* model in which maternal cues *in utero* predict the (normal) adult environment and the organism evolves accordingly in anticipation of future conditions (Gluckman and Hanson, 2006). While the developmental constraints model will apply to birth cohorts that face extreme or novel nutritional insults (Bateson et al., 2014), the predictive model will be relevant when conditions in past generations are an accurate predictor of (normal) conditions in the current generation (Burgess and Marshall, 2014). The pre-modern economy was characterized by wide short-term fluctuations in food supply, but had growth rates close to zero for centuries. The predictive model would have been especially relevant in such an environment, resulting in a population that was adapted physiologically to long-term (low) food supply, with the adaptation varying across space with fixed growing conditions.[2]

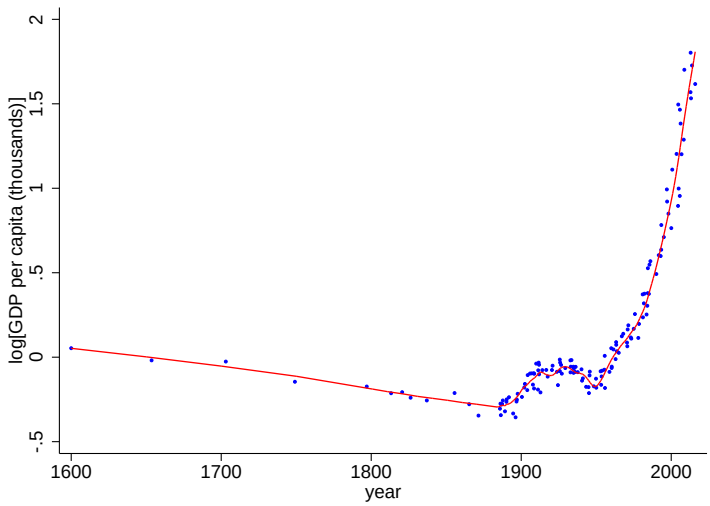
With economic development, there is a substantial increase in income. Figure 1, for example, plots GDP per capita (in logs) for India from 1600 to 2016. Income is stable (declining mildly) for the first 350 years, after which it starts to increase steeply. The standard assumption, which we verify empirically, is that there is a positive and continuous relationship between (food) consumption and (permanent) income. It has been hypothesized in the developmental origins of adult disease literature that the resulting mismatch between current and ancestral consumption (to which the population is adapted) has contributed to the high rates of metabolic disease in developing countries; e.g. Gluckman and Hanson (2004); Wells et al. (2016); Narayan (2017). Our model places additional restrictions on the relationship between metabolic disease and the mismatch, while simultaneously explaining the persistence of undernutrition in these countries by tying the initial adaptation to a set point.

Set point theory was originally motivated by the observation that a typical individual's bodyweight is

[1] There is an erroneous belief that the rapid increase in diabetes in countries like India is due to increased obesity; e.g. Diamond (2011). While obesity may well end up being the primary contributor to diabetes, once these countries have developed, just 4% of Indian adults in our representative sample are obese.

[2] Bateson et al. (2014) provide a point by point rebuttal of criticisms that have been levelled at the predictive model. As discussed below, heterogeneity in per capita food supply can be generated by a straightforward extension of the Malthusian model (see also Ashraf and Galor (2011)).

Figure 1: Evolution of Income in India



Source: Maddison Project Database (2018)
GDP per capita is measured in 2011 US dollars.

remarkably stable over time (Leibel, 2008). We assume that the set point for a given dynasty (family) is determined by food supply in the pre-modern economy and then persists across multiple generations at initial stages of economic development. Although it may be appropriate to characterize the set point with respect to weight for a given individual, we account for possible variation in height across generations by specifying the set point for members of a dynasty by their BMI; i.e. weight conditional on height. The additional advantage of this normalization is that BMI, rather than weight, is commonly associated with the risk of metabolic disease, the second outcome of interest in our analysis.

The set point for bodyweight is part of a homeostatic (stabilizing) system that maintains the body's energy balance against fluctuations in food intake by making metabolic and hormonal adjustments (Müller et al., 2010).[3] A property of all homeostatic systems is that they can only self-regulate within fixed bounds and will fail when these bounds are exceeded. To illustrate the functioning of such a system, consider the regulation of body temperature. The set point for normal human body temperature is approximately $37°C$. When the ambient temperature increases, the body responds by increasing heat loss through perspiration and other mechanisms. Past a particular level, however, the body is unable to defend the set point and body temperature starts to drift up with the ambient temperature. When the body temperature reaches $38.5°C$, hyperthermia sets in with adverse physiological consequences. Although the set point for body weight is individual-specific rather than universal (as with body temperature) the same general principles will apply. In our model, the set point for BMI at initial stages of economic development is determined by ancestral consumption or, equivalently, income in the pre-modern period. As long as current and pre-modern income remain sufficiently close to each other, the body will successfully defend its set point. BMI will be determined by pre-modern rather than current income and the risk of metabolic disease will be low.

[3] As discussed in Speakman et al. (2011), numerous studies indicate that when the energy balance is perturbed in either direction through a change in diet, the body returns to its original weight once the nutritional constraint is released. Furthermore, energy expenditures are modulated to resist the perturbation, indicating that the body is actively defending its set point.

Once the gap between current and pre-modern income crosses a threshold, however, the body will no longer be able to defend the set point and BMI will start to track current income. Escape from the set point is associated with imbalance in energy regulation, which will be accompanied by imbalance in underlying glucose regulation (given that glucose is the primary source of energy for cells in the human body). Failure of glucose homeostasis, in turn, manifests as diabetes and related metabolic disorders.

Based on the preceding discussion, the population will be partitioned into two distinct groups during the process of development: (i) Individuals who remain at their set point, despite the increase in their consumption, are partly responsible for the weak association between nutritional status, which we measure by BMI, and income. (ii) Individuals who have escaped the "nutrition trap," but are not necessarily overweight, are the primary contributors to the increased prevalence of metabolic disease that accompanies economic development. Note that this partition of the population is not permanent. The assumption in all models of developmental plasticity is that the initial adaptation is epigenetic; i.e. it involves changes in gene expression. This adaptation persists over the (animal or human) organism's lifetime and, in theory, can persist over multiple generations even after the conditions that gave rise to it have ceased to be relevant; e.g. Jablonka and Raz (2009). In contrast with traditional genetic alterations, however, epigenetic changes will only persist for a limited number of generations. This would explain why European populations, which were also under-nourished historically, no longer exhibit the traits we document in developing-country populations.[4]

If data on income, BMI, and metabolic disease were available for each dynasty over many generations, going back to the pre-modern period, then we could test the preceding argument directly. For a given dynasty, we would expect to observe a discrete increase in BMI in a particular generation in which the gap between current and ancestral income exceeded a threshold, with an accompanying increase in the risk of metabolic disease. In the absence of such multi-generational household-level data, we develop novel tests, based on current incomes, by characterizing the evolution of income in the population during the process of development. By making plausible assumptions about the distribution of (permanent) income shocks in each generation, our model generates the following implications at any point in time: (i) Although BMI is increasing in current household income at all levels, there is a discontinuous increase in the slope of this relationship at a particular income threshold. (ii) The risk of metabolic disease is constant below the same threshold and increasing in current income above the threshold. An appealing feature of the preceding implications is that they do not require knowledge of ancestral income (the set point) and can be tested with standard datasets that include information on income, BMI, and metabolic disease.

We use nationally representative household data from the India Human Development Survey (IHDS) for the core analysis because the Indian population is simultaneously characterized by high levels of under-nutrition and a high prevalence of metabolic disease. Our main result is that the BMI-income association (separately for children and adults) and the metabolic disease-income association (for adults alone) match the predictions of the model. The presence of a slope discontinuity, which we detect statistically using Hansen's (2017) threshold test is indicative of a set-point threshold. The weak association between BMI

[4]Cutler et al. (2006) note that there is a negative association between the risk of cardiovascular disease and income in advanced economies and Deaton (2007) notes that the weak association between height and income that he documents in developing countries is in sharp contrast with the corresponding associations in European populations.

and household income below the estimated threshold, which is located close to the median income level in the population, explains (in part) the persistence of undernutrition in this population. The steep increase in the probability of metabolic disease with income above the same threshold, which corresponds to a BMI that is at the lower end of the normal range, helps explain the second stylized fact.

Other mechanisms, based on childhood diarrhea (going back to Scrimshaw et al. (1968)), culturally determined dietary preferences (Atkin, 2013, 2016) and son preference (Jayachandran and Pande, 2017) have been proposed to explain the persistence of undernutrition in developing countries. Changes in diet and lifestyle with economic development are also believed to have contributed to the increased prevalence of metabolic disease (Narayan, 2017). The distinguishing feature of our model is that it predicts a discontinuous and inter-connected relationship between income and both BMI and metabolic disease, associated with the failure of an underlying homeostatic system, which is not implied by other potentially coexisting mechanisms.[5] The closest alternative to our model is the developmental constraints model in which adult outcomes are determined by early-life circumstances in each generation. Although both models generate the discontinuities described above, the adaptation to conditions in the distant past distinguishes our model from the alternative. An additional advantage of our model is that it is informative about the nutrition and health experiences of migrants from developing countries to advanced economies. Given the enormous income differential between origin and host country, most migrants to advanced economies will escape the nutrition trap in the first generation. This is consistent with Alacevich and Tarozzi's (2017) finding that the nutritional status of immigrants from South Asia (a historically poor region) to the U.K. converges to the level of the native population very swiftly. Nevertheless, their ancestral income should continue to determine the risk of metabolic disease, possibly for multiple generations, and South Asian immigrants residing in the U.K. and the U.S. are indeed many times more likely to have metabolic diseases, conditional on their BMI, than the native population (McKeigue et al., 1991; Oza-Frank and Narayan, 2010).[6]

The predictions of the model do not apply to India alone. To assess the external validity of the model, we test its predictions with data from the Indonesia Family Life Survey (IFLS) and the Ghana Socioeconomic Panel Survey (GSPS). While the pre-modern set point may be relevant in all developing countries, the fraction of the population that has escaped the set point will depend on a country's stage in the process of development. A cross-country comparison of current income and historical income indicates that the income-gap is substantially higher in Asia than in Africa; indeed, per capita income in Ghana is essentially unchanged from 1960 to 2010, whereas per capita incomes in India and Indonesia increased substantially. This suggests that a large fraction of the Indian and Indonesian populations will have crossed the threshold, whereas the Ghanaian population remains for the most part at its set point. In line with this hypothesis, the results with the IFLS match what we obtain with the IHDS; the association between both BMI and the risk of metabolic disease with income is characterized by a slope discontinuity (at the same income level). In contrast, there is a continuous relationship between BMI and household income with the Ghanaian data

[5] As discussed below, selective child mortality as emphasized by Deaton (2007) or poverty traps as in Dasgupta and Ray (1986) can generate a discontinuous BMI-income relationship, but cannot explain all the results that we obtain.

[6] Other studies, cited in Gujral et al. (2013), document similar patterns in countries such as Fiji, South Africa, and Singapore to which South Asians moved many generations ago as indentured workers and subsequently became relatively wealthy. Approximately 5-10% of type-2 diabetes risk can be attributed to genetic factors (Voight et al., 2010) and, hence, these population differences are unlikely to be genetic.

(information on metabolic disease is not available in the GSPS).

Although the basic paradigms of biological adaptation-persistence and the set point are well established, specific elements of our theoretical foundations remain to be verified: (i) Our assumption that the body will defend its pre-modern set point up to a threshold, although plausible, has not been previously examined. (ii) There is no direct evidence in humans, although there is in small mammals, that prolonged exposure to a low-nutrition environment can result in an adapted phenotype (body type) that persists for multiple generations after the initial environment has ceased to be relevant.[7] For the purpose of our analysis, what matters is the validity of the biological relationships that build on these assumptions and which serve as the starting point for our model: (a) BMI is determined by ancestral income, which is associated with the set point, below a threshold and by current income above the threshold. (b) The risk of metabolic disease is increasing in the difference between current and ancestral income, above but not below the same threshold. Our estimates of the model's structural parameters and the accompanying test of its internal validity allow us to verify not only that a threshold is present, but also the specific form that is imposed on the threshold function in the BMI-income relationship. In addition, we construct exogenous measures of ancestral (pre-modern) income and then verify the biological relationships directly. For this exercise, the location of the threshold is derived from the cross-sectional tests discussed above.

We construct measures of pre-modern income in two ways: First, we use FAO-GAEZ crop suitability data to construct consistent measures of per-household ancestral income at the district level for India and at the sub-regency (sub-district) level for Indonesia, using a method suggested by Galor and Özak (2016). These measures are merged with the IHDS and IFLS datasets that we use to test the cross-sectional implications of the model for India and Indonesia, respectively. Second, we use the agricultural revenue tax that was collected by the British colonial government in 1871, based on its independent assessment of local agricultural productivity, to construct a measure of per-household ancestral income at the village level. This measure, which is available for villages in the modern Indian state of Tamil Nadu, is merged with data from the South India Community Health Study (SICHS) which provides information on income, BMI and metabolic disease for a representative sample of households in rural Vellore district. The consistent finding, obtained independently with IHDS, IFLS, and SICHS data is that pre-modern income determines BMI below the threshold, whereas current income determines BMI above the threshold. Moreover, the difference between current and pre-modern income determines the risk of metabolic disease, above but not below the threshold, once again as specified in the model.[8]

Having tested the implications of the model and validated the biological relationships that it is built upon, we move from micro-data to cross-regional comparisons. Deaton (2007) observes that adult nutritional status, measured by height, is lower in South Asia than what would be predicted by GDP per capita, whereas

[7] In an experiment that was designed to mimic the development process, rats subjected to caloric restrictions over 50 generations continued to display altered epigenetic signatures and to have an elevated susceptibility to metabolic disease two generations after normal nutrition was restored (Hardikar et al., 2015). Although these findings are consistent with long-term adaptation and inter-generational persistence, results from animal studies may not completely capture the experience in human populations (see Luke et al. (2021) for a discussion on specific findings from this experiment).

[8] An exogenous shock to current income generated by economic development will increase the risk of metabolic disease on average, as documented by Sekhri and Shastry (2019). However, our model and empirical analysis indicate that this effect will only be observed for households whose current income has crossed a threshold, which is determined by their ancestral income.

the opposite is true for Africa. We document the same cross-regional patterns, using BMI instead of height to measure nutritional status. In addition, it has been observed that there is an unusually high prevalence of diabetes and related metabolic disorders among South Asians, despite the fact that they have low BMI on average (Narayan, 2017). We show that these seemingly unrelated findings can be interpreted through the lens of our model once we take account of the cross-regional income dynamics; i.e. that current income is higher in Asia (not just South Asia) but historical income, which determines the set point, was higher in Africa.

While the model is informative about a variety of health outcomes, at the micro and the macro level, it is important, particularly from a policy perspective, to go further and quantify the effect of the set point on undernutrition and the prevalence of metabolic disease. A comparison of counter-factual nutritional status and actual nutritional status with IHDS data, based on the estimated model, indicates that the fraction of underweight children aged 5-19 would decline by 10% and the fraction of underweight adults by 24% in the absence of a set point. To quantify the contribution of the set point to metabolic disease, we exploit an additional implication of the model, which is that the risk of disease will not be associated with BMI below a threshold, but will be increasing in BMI above the threshold. Estimates with IHDS data locate the BMI threshold at the lower end of the normal range; just under 22 for the country as a whole and below 21 for South India. We might expect to observe a similar co-existence of undernutrition and metabolic disease at relatively low BMI's, due to the set point mechanism, in other countries in the coming decades as they develop. Policy implications of these findings are discussed in the concluding section.

2 The Model

2.1 Population and Income

The population consists of a large number of infinitely lived dynasties (families). Each dynasty consists of a single individual in each generation, who is replaced by a single descendant in the generation that follows. There is a fixed return on wealth in each generation; i.e. an income flow, which is consumed, so that the stock is passed on (without depletion) to the next generation. We will thus use the terms (permanent) income and wealth interchangeably in the discussion that follows. Denote the logarithm of the dynasty's initial income, in period 0, by y_0. We can think of the initial period as spanning multiple generations in the pre-modern era during which epigenetic adaptation takes place, while subsequent periods, each spanning a single generation, describe the process of development. Income is the same in each generation during the pre-modern era, but subsequently evolves. We normalize so that the initial income distribution is bounded below at zero. Since we have no prior information on the distribution of income in the pre-modern economy, we place no other restrictions on this distribution for the moment. Permanent income in the modern economy is well approximated by the log-normal distribution (Battistin et al., 2009). We thus assume that each dynasty receives a permanent, additive and independent income shock u_τ in each subsequent period τ, where $u_\tau \sim N(\mu, \sigma^2)$. Solving recursively, log-income of a dynasty in period t is

$$y_t = y_0 + U_t, \tag{1}$$

where $U_t = \sum_{\tau=1}^{t} u_\tau \sim N(t\mu, t\sigma^2)$.[9] For ease of exposition, we will denote $t\mu$ by μ_t and $t\sigma^2$ by σ_t^2.

2.2 Biological Relationships

We now characterize the biological relationships between (i) BMI and income, and (ii) the risk of metabolic disease and income, during the process of economic development. This characterization is based on the verbal description from the preceding section.

There is a positive and continuous relationship between (food) consumption and income in all time periods.[10] Focussing first on the initial period, it follows that height and weight are increasing continuously in income (consumption). In addition, we assume that BMI (weight conditional on height) is increasing in income, which is consistent with the (continuing) positive association between these variables that we document below in developing-country populations. The relationship between BMI in the initial period, z_0, which determines the dynasty's BMI set point, and its initial income, y_0 is thus specified as follows:[11]

$$z_0 = a + by_0. \tag{2}$$

In subsequent periods, each descendant's body will defend her dynasty's set point in the face of fluctuations in consumption that arise due to the permanent income shocks. However, as noted, the body can only respond up to a point to deviations in income from the initial level, y_0, that determined the set point. There is thus a threshold α, such that BMI in period t,

$$z_t = \begin{cases} a + by_0 & \text{if} \quad U_t \leqslant \alpha \\ a + by_t & \text{if} \quad U_t > \alpha \end{cases} \tag{3}$$

Equation (3) imposes the restriction that the (linear) relationship between BMI and income is the same, below and above the threshold; what changes is the relevant measure of income, from y_0 to y_t. Later in the analysis, we will validate the structure we have imposed in equation (3) by separately estimating the b parameter, below and above the (estimated) threshold.

Notice that we do not specify a lower threshold for the set point; the implicit assumption is that dynasties do not regress with regard to nutritional status during a period of rapid economic growth. Given historically low levels of food supply in developing countries, the metabolism would have adapted to defend the set point especially vigorously against downward fluctuations in consumption.[12] Although mean income is increasing

[9] We do not include a dynasty-specific identifier when deriving and characterizing the income equation to simplify notation.

[10] The implicit assumption is that individuals do not alter their behavior to account for the effect of the set point on their nutritional status and the risk of metabolic disease during the process of development. This seems reasonable, given that the effect of the set point on these outcomes is the subject of our inquiry and, thus, is not known to the general population. This assumption also allows us to specify the biological relationships with respect to income rather than (more proximate) consumption.

[11] Although our model is concerned with inter-generational income dynamics, recall that intra-generational income volatility motivates the epigenetic adaptation to pre-modern income, y_0. It is possible that the set point adjusts in more volatile pre-modern environments to buffer individuals against more negative transitory income shocks, but we do not obtain empirical support for this response below.

[12] This is consistent with the conventional view that the regulation of bodyweight is more responsive to weight loss than to weight gain (Müller et al., 2010). For example, despite repeated weight cycling in response to seasonal fluctuations in food supply, minimal bodyweight in a sample of rural Gambian women remained extremely stable (within 1.5 kg.) over a period of

in our model, the distribution of income shocks is unbounded and, hence, a small number of dynasties could, nevertheless, face a sequence of very negative shocks that the body could not defend. However, all societies have consumption-smoothing mechanisms in place to insure against precisely such negative outcomes and these mechanisms improve with economic development. We thus assume that dynasties always successfully defend the set point z_0 in the face of negative income shocks, either biologically or by taking advantage of social safety nets to augment their consumption.[13]

As long as income remains within the threshold associated with the dynasty's set point, metabolic and hormonal adjustments ensure that the increases in consumption that accompany the increases in income due to economic development do not translate into increases in BMI. Once income crosses the threshold, however, the body can no longer defend the set point and BMI starts to track current income.[14] As discussed in the preceding section, this simultaneously increases the risk of metabolic disease. As in the developmental origins of adult disease literature, this risk is specified to be increasing in the mismatch between current income, y_t, and initial income, y_0. The additional feature of our model is that the income-gap only determines the risk of metabolic disease when it exceeds a threshold (and the individual escapes the nutrition trap). The relationship between the probability of metabolic disease, $P(D_t)$, and income can thus be characterized as follows:[15]

$$P(D_t) = \begin{cases} \gamma_1 & \text{if} \quad U_t \leqslant \alpha \\ \gamma_1 + \gamma_2(y_t - y_0) & \text{if} \quad U_t > \alpha \end{cases} \tag{4}$$

2.3 Cross-Sectional BMI-Income Relationship

Figure 2 describes the evolution of BMI across multiple generations (time periods) for a single dynasty, based on the biological relationship specified above. For expositional convenience, we assume that the dynasty only receives positive income shocks. Starting from an initial income, y_0, the dynasty's income thus increases monotonically across generations. However, it's members' BMI will remain at the dynasty's set point, $z_0 = a + by_0$, until y_t exceeds $y_0 + \alpha$. At that point in time, there will be a discrete increase in BMI, after which BMI will track y_t. If data over many generations, going back to the pre-modern period, were available for each dynasty, then these implications could be tested directly. In the absence of such multi-generational data, we proceed to derive the cross-sectional association between BMI and income, as implied by equation (3), when a dynasty-specific set point for BMI is present.

Recall that we normalize so that the initial income distribution is bounded below at zero. We also do not specify a lower threshold for the set point. It follows that all individuals with $y_t \leq \alpha$ must be at their

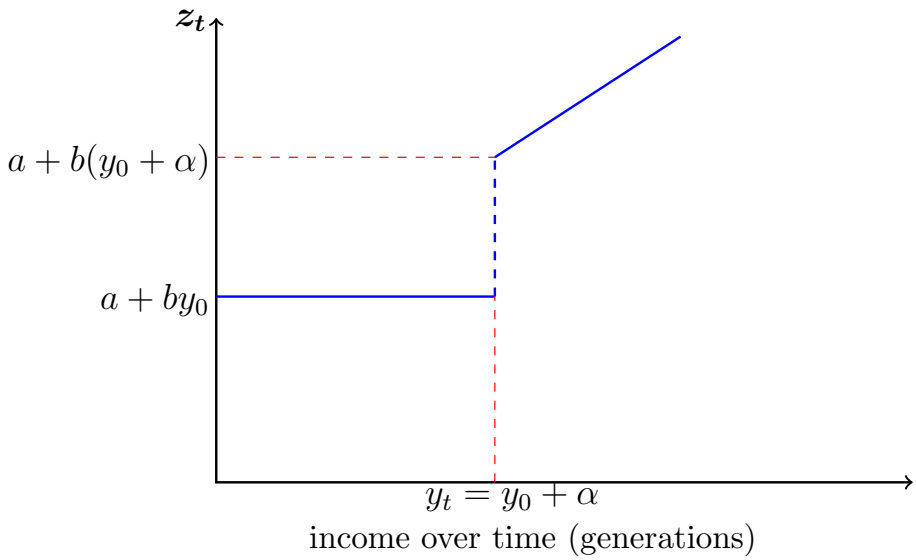
10 years (Prentice et al., 1992).

[13] Given that income shocks are positive on average and their distribution is symmetric, such redistribution is feasible. We are effectively ignoring catastrophic common shocks, such as famines, that can shift set points in an entire birth cohort. Such events have always been rare and are less relevant in the modern economy.

[14] This assumption is consistent with set point models that have recently been proposed for populations in advanced economies: (i) Nutritional status responds flexibly to food intake between genetically determined lower and upper set points. The upper set point (which becomes relevant in the modern economy) lies in the obese-overweight range for some individuals (Speakman et al., 2011). (ii) The inherited pre-modern low-BMI set point is replaced by "settling-points" which the body does not defend (Müller et al., 2010).

[15] $\gamma_1 > 0$, $\gamma_2 > 0$ in equation (4). The implicit assumption, which is consistent with recent evidence on diabetes reversal, is that the risk of metabolic disease can change in both directions over time as the individual's BMI shifts on either side of the threshold.

Figure 2: BMI - Income Relationship (within dynasty over generations)



set point; some of these individuals will belong to dynasties that had initial incomes below α and which subsequently increased their income by relatively little, whereas others will belong to dynasties whose income has drifted down over time. Mean BMI at any given level of income $y_t \leq \alpha$ can then be characterized by the following expression:

$$\mathbb{E}(z_t|y_t; y_t \leqslant \alpha) = \int_{-\infty}^{y_t} [a + b(y_t - U_t)] P(U_t \mid y_t) \ \mathrm{d}\, U_t$$

Let $f(\cdot)$ denote the density of the y_0 distribution. Applying Bayes' rule:

$$P(U_t \mid y_t) = \frac{P(U_t)P(y_t \mid U_t)}{\int_{-\infty}^{y_t} P(\widetilde{U}_t)P(y_t \mid \widetilde{U}_t)} = \frac{\phi(U_t; \mu_t, \sigma_t^2) f(y_t - U_t)}{\int_{-\infty}^{y_t} \phi(\widetilde{U}_t; \mu_t, \sigma_t^2) f(y_t - \widetilde{U}_t) \ \mathrm{d}\, \widetilde{U}_t}$$

In the absence of any prior knowledge about the distribution of pre-modern income, we make the simplifying assumption (validated below) that $f(\cdot)$ is approximately constant. It follows that

$$\mathbb{E}(z_t|y_t; y_t \leqslant \alpha) = \int_{-\infty}^{y_t} [a + b(y_t - U_t)] \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \ \mathrm{d}\, U_t = a + b\left(y_t - e^L(y_t)\right) \tag{5}$$

where $e^L(y_t) = \frac{1}{\Phi(y_t; \mu_t, \sigma_t^2)} \int_{-\infty}^{y_t} U_t \phi(U_t; \mu_t, \sigma_t^2) \ \mathrm{d}\, U_t$.

The assumption that initial income is (approximately) uniformly distributed allows us to derive an expression for $P(U_t \mid y_t)$ that is independent of y_0 in equation (5). Given that the uniform distribution has bounded support, the lower range of integration should then extend to $y_t - \overline{y}_0$, where $\overline{y}_0$ is the right support of the initial income distribution. The advantage of extending the range to $-\infty$ is that we can solve the model analytically, with simulations reported in Appendix A indicating that this approximation has no discernable effect on predicted BMI (and the risk of metabolic disease) except in the right tail of the y_t distribution. Consistent with the simulation results, we will see in Section 3.3 that predicted BMI, derived

from a structural model that incorporates the analytical approximation and the distributional assumptions we have imposed on the income generating process, matches actual BMI very closely.

For individuals with $y_t > \alpha$, some will have crossed their set point threshold, while others (who started with a higher initial income) will remain at their set point. The expression for mean BMI at income level $y_t > \alpha$ thus includes both types of individuals:

$$\begin{aligned}\mathbb{E}(z_t|y_t; y_t > \alpha) &= \int_{-\infty}^{\alpha} [a + b(y_t - U_t)] \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t + \int_{\alpha}^{y_t} [a + by_t] \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ &= a + b\left(y_t - e^H(y_t)\right), \quad \text{where } e^H(y_t) = \frac{1}{\Phi(y_t; \mu_t, \sigma_t^2)} \int_{-\infty}^{\alpha} U_t \phi(U_t; \mu_t, \sigma_t^2) \, \mathrm{d}\, U_t\end{aligned} \tag{6}$$

As shown in Appendix A, closed-form expressions for $e^L(y_t)$ and $e^H(y_t)$ can be derived using the properties of the normal and standard normal distributions:

$$e^L(y_t) = \mu_t - \sigma_t \frac{\phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)}{\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)} = \mu_t - \sigma_t \Lambda\left(\frac{y_t - \mu_t}{\sigma_t}\right) \tag{7}$$

$$e^H(y_t) = \frac{\mu_t \Phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right) - \sigma_t \phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right)}{\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)} \tag{8}$$

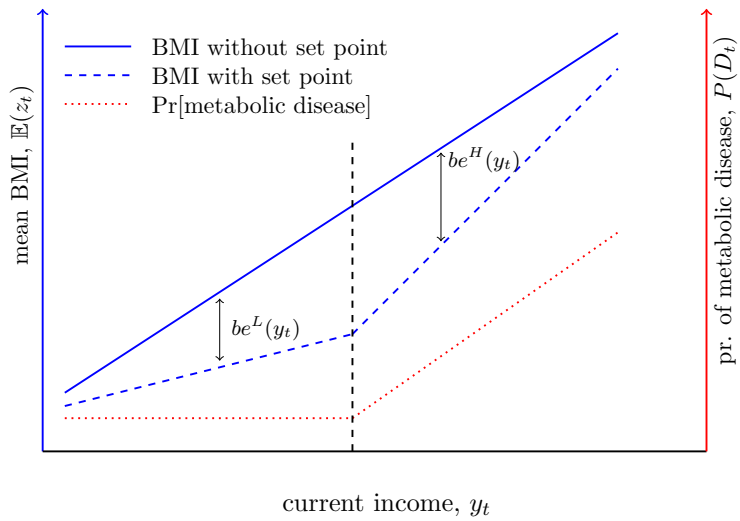
where $\Lambda(\bullet)$ is the inverse Mills ratio with the property that its derivative, $\frac{\mathrm{d}\,\Lambda(\bullet)}{\mathrm{d}(\bullet)}$, is negative, increasing and bounded on the interval $(-1, 0)$. Given the properties of the inverse Mills ratio, and noting that $e^H(y_t)$ is decreasing in y_t, we obtain the following result (the proof is in Appendix A):

Proposition 1 *(i) The slope of the BMI-income relationship is positive but less than b for $y_t \leqslant \alpha$ and greater than b for $y_t > \alpha$. (ii) There is a discontinuous change in the slope of the BMI-income relationship at $y_t = \alpha$. (iii) There is no level discontinuity in the BMI-income relationship at $y_t = \alpha$.*

The relationship between BMI and income implied by Proposition 1 is described graphically in Figure 3. Each dynasty transitions discretely to a higher BMI level, at a particular point in time, in Figure 2. This level-shift is smoothed out, and translates into a slope change at a particular income level, when we derive the corresponding cross-sectional BMI-income association across dynasties, at any point in time.

We complete the analysis of the BMI-income relationship by considering alternative specifications of the set point. Although an epigenetically determined set point may be heritable, it will ultimately cease to be relevant once a changed economic environment has been in place for a sufficient number of generations. Our model thus describes the relationship between nutritional status and income over a finite number of generations during the initial phase of economic development. During this phase, we assume that the set point, z_0, determined in period 0, is fixed. However, an alternative specification would allow the set point to adjust gradually across generations until it is no longer relevant. For example, the set point could be

Figure 3: Cross-Sectional Relationships



specified as a weighted average of y_0 and y_t, with the weight on y_t increasing over time. Alternatively, the set point could be determined by initial conditions (income) in each generation, as in the developmental constraints model discussed above. Since income does not vary within periods in our setup, the set point in period t with this alternative specification will then be parental income, y_{t-1}. As shown in Appendix A, the alternative specifications generate the same qualitative predictions as Proposition 1. What distinguishes the benchmark specification in equation (3) from the alternatives, as verified empirically, is that BMI below the estimated current-income threshold is determined exclusively by y_0. The test of internal validity, reported in Section 3.3, will provide additional and independent support for our preferred specification and, more generally, the structure we have imposed on the model. This stringent test is based on the observation, from equations (5) and (6), that once $e^L(y_t)$ and $e^H(y_t)$, respectively, are subtracted from y_t, the slope of the BMI-adjusted income relationship will be the same (equal to b) below and above the estimated threshold.

2.4 Cross-Sectional Disease-Income Relationship

Taking as given the biological relationship between the probability of metabolic disease, $P(D_t)$, and income, as specified in equation (4) for a single dynasty, the corresponding association in the cross-section across dynasties can be derived as follows:

Proposition 2 *(i) There is no relationship between $P(D_t)$ and y_t for $y_t \leqslant \alpha$, and a positive relationship for $y_t > \alpha$. (ii) There is a discontinuous change in the slope of the $P(D_t) - y_t$ relationship at $y_t = \alpha$. (iii) There is no level discontinuity in the $P(D_t) - y_t$ relationship at $y_t = \alpha$.*

The proof in Appendix A follows the same steps as the proof of Proposition 1. The $P(D_t) - y_t$ relationship specified by Proposition 2 is described graphically in Figure 3. This relationship is qualitatively the same as the $\mathbb{E}(z_t) - y_t$ relationship, except that the slope is zero below the threshold. This is because the risk of metabolic disease is constant (and the same) for all individuals who remain at their set point and because all individuals below the income threshold are at their set point. Above the threshold, in contrast, the risk

of metabolic disease is increasing in income. This is due to (i) the greater fraction of individuals who have escaped their set point, and (ii) the increased risk for those who have escaped. Note that the model predicts that the $\mathbb{E}(z_t) - y_t$ and $P(D_t) - y_t$ relationships will exhibit a slope discontinuity at the same income level: $y_t = \alpha$.[16]

Proposition 1 indicates that BMI is increasing with income at all levels, more steeply above a threshold, while Proposition 2 indicates that the risk of metabolic disease is only increasing in income above the same threshold. Bringing the two implications together, it follows that there will be no association between the risk of metabolic disease and BMI up to a BMI threshold (which corresponds to the underlying income threshold) and a positive association thereafter. Although our analysis focuses on the BMI-income and metabolic disease-income associations, we will examine this additional implication of the model, which is especially relevant for policy, in the concluding section of the paper.

3 Testing the Model

3.1 Cross-Sectional Analysis

The core data set that we use to test the model is the India Human Development Survey (IHDS). This nationally representative household survey, which was conducted in 2004-2005 and 2011-2012, includes detailed information on household income, nutritional status for children and adults residing in the household at the time of the survey, and the self-reported prevalence of different diseases among adult members of the household. The survey includes, in addition, information on household composition, food consumption expenditure in the last month, morbidity among the children in the last month, and detailed geographic locators, which will be used to supplement the analysis.[17]

Although a dynasty consists of a single individual in each generation in our model, multiple individuals will reside in a household in practice. Income is thus measured at the household level, as the average over the 2004 and 2012 rounds.[18] This smooths out noise in the round-specific income measures and given that the rounds were conducted nearly a decade apart, provides a more accurate estimate of the household's permanent income. Nutritional status for adults is measured by BMI, as in the model, in each survey round. To be consistent with the adults and in line with the model, nutritional status for children (aged 5-19) who have passed the formative early-life period is measured by the BMI-for-age (BFA) in each survey round. The BFA is constructed as a z-score, based on child growth standards provided by the WHO.[19] Metabolic

[16] Although we normalize so that the initial income distribution is bounded below at zero, it can more generally be bounded below at some income level $\underline{y}_0$, in which case the threshold would be located at $y_t = \underline{y}_0 + \alpha$. This would change the interpretation of the threshold location, but otherwise leave the analysis unchanged.

[17] The Demographic Health Survey (DHS), which is used by Deaton (2007) and Jayachandran and Pande (2017) also contains many of these variables. However, the DHS is not suitable for our purposes because it only collects indicators of asset ownership, which must then be converted into a crude wealth statistic using principal component analysis. The tests of the model, particularly the statistical tests to locate a slope-change at an income threshold, cannot be implemented without fine-grained income data.

[18] Household income, measured in thousands of Rupees per month, includes farm income, non-farm business income, wage income, remittances, and government transfers. To make incomes in the two rounds comparable, we adjust 2004-2005 incomes to 2011-2012 prices. For rural areas, the correction is based on the Consumer Price Index (CPI) for agricultural wage labor and for urban areas it is based on the CPI for industrial workers.

[19] The growth standard for children aged 5-19 is based on the 2007 WHO Reference, which is a reconstruction of the 1977

disease is constructed as an individual-specific binary variable that indicates whether the household head and his spouse report having been diagnosed with diabetes, hypertension, or cardiovascular disease in each survey round.[20]

Proposition 1 derives the cross-sectional association between BMI and income when a dynasty-specific set point is present: although the relationship is positive at all income levels, there will be a discontinuous shift to a steeper slope at a particular income threshold. Proposition 2 derives the corresponding association between the risk of metabolic disease and income: while a slope-change at the same income threshold is predicted, the difference is that variation in income is not expected to affect the risk of disease below the threshold. We test these predictions by nonparametrically estimating the BMI-income and metabolic disease-income relationships. Although our analysis focuses on the relationship with income, other individual and household characteristics could independently determine nutritional status and the risk of metabolic disease. All of the estimating equations in our analysis thus include the following standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district and survey-round.[21] The effect of gender bias on nutritional status, as documented by Jayachandran and Pande (2017), is captured by the gender and birth order dummies. Geographical variation in food tastes, as emphasized by Atkin (2013, 2016), or in the disease environment, as documented by Dandona et al. (2017), is captured by the district dummies and the rural-urban dummy. The covariates listed above are partialled out using the Robinson (1988) procedure prior to the nonparametric estimation reported in Figures 4a and 4b.[22]

The vertical lines in Figures 4a and 4b mark the point where we locate an income threshold, based on the statistical test described below. The shaded area around each line marks the 95% confidence interval for the threshold location, based on the same test. It is evident from both figures, and with all three outcomes, that the association with income is relatively weak below the estimated threshold, and much stronger above the threshold. The threshold location with children's BFA in Figure 4a matches precisely with the threshold location for the probability of adult metabolic disease as the outcome in Figure 4b. The estimated threshold location with adult BMI as the outcome is slightly lower than the other threshold locations. Such minor differences are to be expected, given that BMI is directly measured, whereas metabolic disease (although diagnosed) is self reported. Nevertheless, this discrepancy is not observed in the robustness tests that follow and in the subsequent analyses with South Indian and Indonesian data.

The threshold locations and confidence intervals in Figures 4a and 4b are estimated using a procedure developed by Hansen (2017). This procedure involves sequential estimation of the following piecewise linear equation:

$$z_i = \beta_0 + \beta_1 y_i + \beta_2 (y_i - \tau) \times \mathbb{I}(y_i - \tau > 0) + x_i \lambda + \epsilon_i, \tag{9}$$

where z_i is an outcome of interest; e.g. BMI, y_i is household $i's$ income, τ is the location of the income

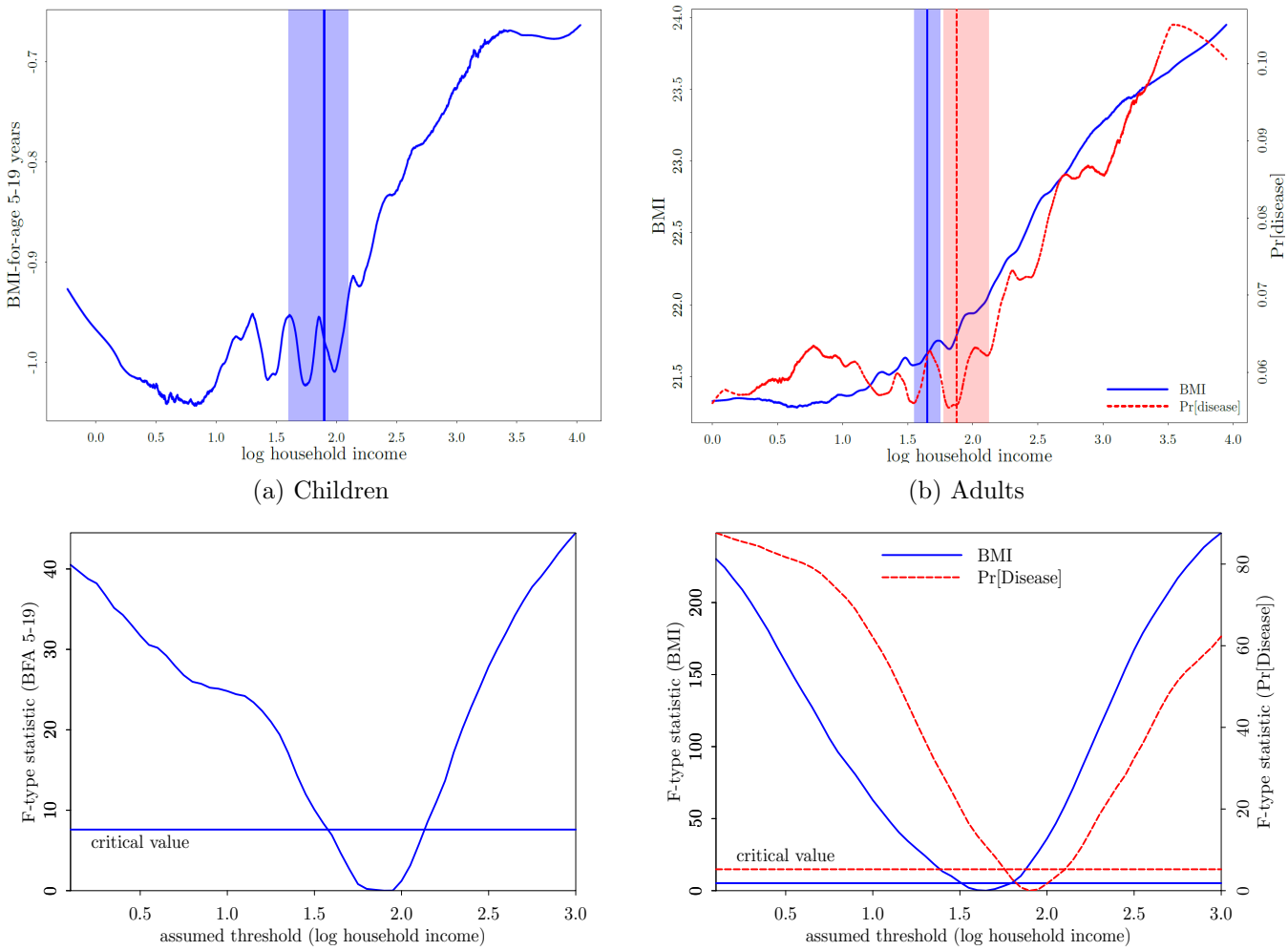
National Center for Health Statistics (NCHS) growth standard. Following the recommendation of the WHO, BMI-for-age z-scores outside the (-5,5) interval are dropped from the analysis.

[20]The model implicitly assumes that metabolic disease is reversible, consistent with recent experimental evidence (Taylor, 2013) and we allow for this possibility in the empirical analysis.

[21]Age is measured in years and the birth order is top coded at 3.

[22]Observations in the top and bottom 1% of the outcome distribution are excluded from the estimation sample in all of our analyses. This ensures that the estimation results are not driven by extreme outliers.

Figure 4: Nutritional Status and Metabolic Disease with respect to Household Income



(a) Children (b) Adults

(c) Children (d) Adults

Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district, and survey-round are partialled out prior to nonparametric estimation. The same set of covariates are included in the estimating equation at each assumed threshold for the threshold test.
The vertical lines mark the estimated threshold location and the shaded areas demarcate the corresponding confidence intervals. Cluster bootstrapped 5% critical values are used to bound the threshold location.

threshold (which must be estimated), $\mathbb{I}(\cdot)$ is an indicator function, β_1, β_2 are slope parameters, and x_i is a vector of additional covariates (the same covariates that are partialled out prior to nonparametric estimation). This equation is estimated at different assumed income thresholds (values of τ), starting at a very low income level and then covering the entire income range in small increments. An F-type statistic is computed at each assumed threshold, based on a comparison of the sum of squared residuals at that assumed threshold and the minimized value across all assumed thresholds. This statistic will have a minimum value of zero by construction, and the assumed income threshold corresponding to that value is thus our best estimate of the true threshold. If there is indeed a slope-change, then the F-type statistic will increase steeply as the assumed threshold moves away (on either side) from the income level at which it is minimized.

Table 1: Piecewise Linear Equation Estimates - nutritional status and metabolic disease

Dependent variable:	BFA 5-19	adult BMI	metabolic disease
	(1)	(2)	(3)
Baseline slope (β_1)	0.027	0.239**	0.002
	(0.028)	(0.057)	(0.002)
Slope change (β_2)	0.188**	0.940**	0.028**
	(0.037)	(0.066)	(0.003)
Threshold location (τ)	1.90	1.65	1.90
	[1.60, 2.10]	[1.55, 1.75]	[1.80, 2.05]
Threshold test $p-$value	0.000	0.000	0.000
Mean of dependent variable	-0.932	22.002	0.074
N	48,986	76,949	148,928

Source: India Human Development Survey (IHDS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district, and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
** significant at 5%, based on cluster bootstrapped confidence intervals.

Figures 4c and 4d plot the F-type statistic across the range of assumed thresholds for children's BFA and the adult outcomes, respectively. Bootstrapped, outcome-specific 5% critical values for the F-type statistic are also reported in the figures, allowing us to locate the threshold with the requisite degree of statistical confidence. The F-type statistic increases steeply as the assumed threshold moves away from the income level at which it is minimized, which implies, in turn, that the location of the threshold can be bounded with a relatively high degree of statistical precision. The estimated threshold locations in Figures 4a and 4b correspond to the income levels (assumed thresholds) at which the F-type statistic is minimized. The estimated confidence intervals are determined by the points of intersection between the F-type statistic and the critical value lines.

The same (wild) bootstrap procedure, clustered at the level of the Primary Sampling Unit, that is used to compute the critical values and, hence, the 95% confidence interval for the threshold location in Figures 4c and 4d can also be used to compute standard errors for the slope coefficients, β_1 and β_2, in a piecewise linear equation estimated at the threshold we have located.[23] Moreover, a similar bootstrap procedure can be used to test our statistical model with a slope change at an income threshold, as described in equation (9), against the null hypothesis that there is a linear relationship between household income and each of the outcomes. These results are reported in Table 1. We can easily reject the null that the relationship is linear, without a discontinuity at a threshold, with each outcome. The reported point estimates of the baseline

[23]Following Hansen (2017) and Roodman et al. (2019), a coefficient's significance at the 5% level is determined by cluster bootstrapped 95% confidence intervals. For ease of exposition we report cluster bootstrapped standard errors for each coefficient.

slope coefficient (β_1) and the slope-change coefficient (β_2) are obtained at our best estimate of the true threshold, τ, for each outcome. As predicted by our model with a set point, the slope increases to the right of the threshold with each outcome (the slope-change coefficient is positive and significant). Proposition 1 indicates, in addition, that the slope to the left of the threshold should be positive with nutritional status as the outcome. This result is obtained for adults (Column 2) but not children (Column 1), perhaps because sample sizes are smaller for the children or because the income relationship strengthens over the life-course. In line with Proposition 2, there is no relationship between the probability of metabolic disease and household income below the threshold in Column 3, in contrast with the strong positive relationship above the threshold.[24] The estimated threshold location ranges from 1.65 to 1.9 for the three outcomes, with some amount of overlap in the confidence intervals between any pair of outcomes (the only exception is adult disease and BMI). The median income in our nationally representative sample of households is 1.8. This implies that the lower half of the income distribution in India remains stuck at the pre-modern BMI set point, whereas the upper half is at risk of metabolic disease.

We complete this section by verifying the robustness of this evidence in a number of ways: (i) We include average education among adult women and adult men in the household, as well as household composition, measured by the number of children, the number of teens, the number of adults, and the number of adults engaged in physical labor as additional covariates in the estimating equation in Appendix Figure B1 and Table B1.[25] While these variables could independently determine feeding practices, health seeking behavior, and other decisions that determine nutritional status and health outcomes, we see that the results are robust to their inclusion. (ii) We show in Appendix Figure B2 and Table B2 that the results are robust to restricting outcomes to the 2011-2012 survey round. (iii) We show in Appendix Figure B3 and Table B3 that the results for adults, with both BMI and metabolic disease as outcomes, are robust to separating men and women. (iv) Although we focus on a set point for BMI, given its link to metabolic disease and the attention placed on body weight in the recent literature, ancestral income would also have determined stature. Some past studies; e.g. Seckler (1984) have argued that there is a set point for stature that, like the set point for body weight, was part of a homeostatic system regulating the body's energy balance in the pre-modern economy. If that were the case, then Proposition 1 should hold, with BMI replaced by height. Appendix Figure B4a and Table B4, Columns 1-2, verify that this is true, using height-for-age for children aged 5-19 and height for adults to measure nutritional status. As with BMI, we cannot reject the hypothesis that the threshold for children and adults is located at the same income level. These findings can explain (in part) the weal association between height and income that has been documented in developing countries; e.g. Deaton (2007). One consequence of this weak association, as observed in Appendix Figure B4a, is that the range for height in the data is very narrow. This implies that our core results, using BMI as the measure of nutritional status, would be largely unchanged if weight (without conditioning for height) were used as the outcome instead. We verify in Appendix Figure B4b and Table B4, Columns 3-4 that this is

[24] The number of observations in Column 3 is substantially greater than in Column 2 for two reasons: (i) BMI, based on height and weight, can only be measured for individuals who were physically present at the time of the survey interview. (ii) BMI data were only collected for a small number of adult men in the 2004-2005 round.

[25] Household income and both average education and household composition are closely related, which is why we exclude these variables from the estimating equation in the benchmark specification.

indeed the case.[26] (v) We show in Appendix Figure B5 and Table B5 that the implications of the model are obtained with individual metabolic diseases, although a slope discontinuity cannot be located (statistically) with cardiovascular disease. It is particularly striking that the risk of hypertension and diabetes track very closely with income and that the precisely estimated threshold location is the same for both disorders.

3.2 Alternative Explanations

The additional covariates that we include in the estimating equations are meant to account for independent determinants of nutritional status and metabolic disease in India. However, such controls may not be complete and thus the discontinuity in the BMI-income and metabolic disease-income relationships plays an important role in ruling out alternative explanations. In particular, we do not expect other factors to generate such discontinuities and we verify that this is indeed the case with two important proximate determinants of nutritional status in developing countries: nutrient intake and children's illness, particularly diarrhoeal disease.[27]

Nonparametric estimates of the nutrient intake-household income relationship are reported in Figure 5a and corresponding estimates of the children's illness-household income relationship are reported in Figure 5b, using IHDS data. Nutrient intake is measured by the consumption of calories and fat (in grams) at the household level. Children's illness is measured by whether the child is reported to have had diarrhea and cough in the past month. The standard set of covariates, plus household composition and the number of adults engaged in physical labor are partialled out prior to estimation using Robinson's procedure. The additional covariates are meant to condition for energy expenditures, since energy (nutrient) intake net of these expenditures determines nutritional status.[28] We see that there is a positive and continuous relationship between the intake of calories and fat and household income in Figure 5a. In contrast, there is a negative and continuous relationship between the incidence of both diarrhea and cough with household income in Figure 5b. In neither figure do we observe a discontinuous slope-change at any income level. Indeed, Hansen's test cannot place bounds on the threshold location and, hence, fails to locate a slope-change at any assumed threshold in Figures 5c and 5d.[29] The same result (not reported) is obtained with other measures of nutrient intake – sugar consumption – and children's illness – the incidence of fever.[30]

[26] The results, with height and weight taken together, also rule out the possibility that the nonlinear and discontinuous association between BMI and income that we obtain is mechanically generated by positive, continuous, and distinct associations between income and both height and weight.

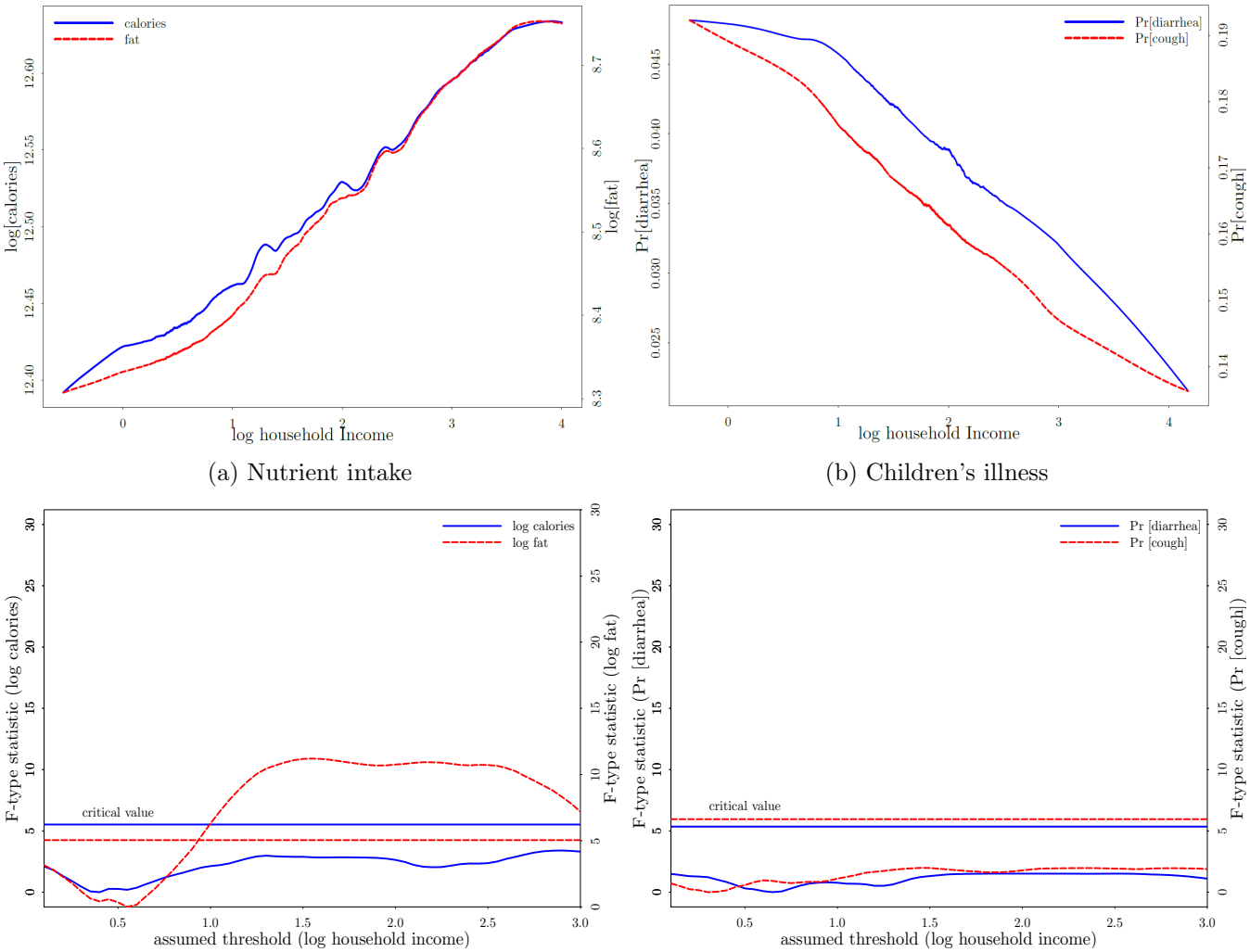
[27] Our model assumes a positive and continuous relationship between nutrient intake (consumption) and income. It is the biologically determined set point that breaks the smooth relationship between nutritional status and consumption and, by extension, income.

[28] Ng and Popkin (2012) decompose total energy expenditures into types of activity: work, active leisure, travel, and domestic tasks. The work category accounted for over 80% of the total energy expenditure in 2000 and 2005 in India.

[29] We include children aged 0-5 in addition to the 5-19 year olds when estimating the children's illness-income relationship because infectious diseases are concentrated in early childhood. This addition to the sample increases the likelihood of detecting a threshold. The corresponding tests (not reported) that separate the children into 0-5 year olds and 5-19 year olds also fail to detect a threshold. With regard to the nutrient intake-income relationship, there has been some controversy in the literature about the strength of this relationship; see, for example, Behrman and Deolalikar (1987) and Subramanian and Deaton (1996). However, none of these previous analyses suggest that there will be a discontinuity in this relationship.

[30] Appendix Figure B6 examines the relationship between household income and expenditures on nine food categories: wheat, rice, cereals and derivative products, meat and eggs, milk and derivative products, pulses, vegetables, sugar and derivative products, and oil. Although a positive association is observed with each category, a slope discontinuity cannot be detected with any category.

Figure 5: Nutrient Intake and Children's Illness with respect to Household Income



(a) Nutrient intake (b) Children's illness

(c) Nutrient intake (d) Children's illness

Source: India Human Development Survey (IHDS).
For the nutrient intake figures, the following covariates are partialled out prior to nonparametric estimation and included in the estimating equation at each assumed threshold: reported local price of rice, wheat, cereals and their derivative products, pulses, meat, sugar, oil, eggs, milk and its derivative products, vegetables and dummies for the number of children, adults, and teens in the household, dummies for the number of adults engaged in physical labor, caste group, rural area, district, and survey-round. For the children's illness figures, the standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order, caste group, rural area, district and survey-round are partialled out prior to nonparametric estimation and included in the estimating equation at each assumed threshold.
Cluster bootstrapped 5% critical values are used to bound the threshold location.

Although we do not observe a discontinuous association between income and a number of nutrient intake measures, it is possible (albeit unlikely) that an unobserved component of nutrient intake, or physical activity, is changing discontinuously at the income level at which we observe the discontinuous increase in BMI. However, this alternative explanation for the change in BMI would also need to generate a discontinuous increase in the risk of metabolic disease at the same income level. We will see below that BMI starts to increase from a base level that is at the lower end of the normal range (around 21). There is no obvious reason why this increase in BMI, starting from such a low base level, would be associated with an increase

in the risk of metabolic disease. In our model, BMI and the risk of metabolic disease are not directly related. The reason why these variables change discontinuously at the same income level is because they are simultaneously (and independently) impacted by the failure of an underlying homeostatic system, which is specific to developing country populations.

Could variation in child survival with income, which Deaton (2007) proposes as a possible explanation for the weak nutritional status-income association across countries, explain our results? As discussed in Appendix C, selective child mortality could generate the observed discontinuous relationship berween BMI and income, but it would be driven by households at the lower end of the BMI distribution at each income level. Quantile regressions reported in the Appendix do not detect such variation in the BMI-income relationship, ruling out this alternative mechanism.

Finally, we consider poverty traps as an explanation for our results. When poverty traps are generated by credit constraints and non-convexities, as in Galor and Zeira (1993) and Banerjee and Newman (1993), households with sufficiently low initial income y_0 will remain permanently at that level. This will change the distribution of current income, y_t, but without a set point, there will be no discontinuity in the cross-sectional BMI-income ($z_t - y_t$) relationship. Poverty trap models generated by undernutrition; e.g. Dasgupta and Ray (1986) could potentially generate such a discontinuity because of the feedback from z_t to y_t below a current income threshold. However, there is no role for y_0, conditional on y_t, below the threshold in this model. In contrast, as assumed in our model and verified below, z_t is determined exclusively by y_0 below the threshold. Moreover, no poverty trap model has implications for the risk of metabolic disease, and even if it did, it would not predict a discontinuous increase in this risk and in nutritional status at an income level close to the median in the population as we observe.[31]

3.3 Internal Validity

Our assumption that the body defends its inherited (pre-modern) set point up to a threshold has not been previously verified in developing country populations. Moreover, the model places additional structure on the threshold function in equation (3) by specifying that there is a linear relationship, with slope b, between BMI, z_t, and income, both below and above the threshold, with the relevant income measure switching from y_0 to y_t. The analysis that follows empirically validates the threshold assumption, the specific structure we have imposed on the threshold function in the BMI-income relationship, the distributional assumptions underlying the income generating process, and the analytical approximation that is used to derive closed-form expressions for e^L, e^H.[32]

Given our modeling assumptions, equation (3) implies the following cross-sectional $z_t - y_t$ relationships,

[31] We assume, as do Subramanian and Deaton (1996), that nutrient intake is determined by income. However, there is an older literature in development economics (see Strauss and Thomas (1998) for a survey) that is concerned with the effect of nutrient intake on productivity and, by extension, income. The general view in this literature is that the income-nutrient intake relationship will be strongest at low levels of nutrition and income. In the context of our analysis, reverse causation from nutrition to income would then bias our estimates of the nutritional status-income relationship upwards, particularly at low income levels, possibly leading to a false rejection of our model. Based on the nonparametric estimates for India in Figure 4 and the corresponding estimates for Indonesia that follow, this concern does not appear to be relevant in practice.

[32] The threshold test provides support for an underlying discontinuity in the nutritional status-income relationship, but while we can easily reject the null hypothesis that this relationship is linear, it is more difficult to formally rule out the possibility that it is highly nonlinear (without a discontinuity).

as specified in equations (5) and (6):

$$\mathbb{E}(z_t|y_t; y_t \leqslant \alpha) = a + b(y_t - e^L(y_t))$$

$$\mathbb{E}(z_t|y_t; y_t > \alpha) = a + b(y_t - e^H(y_t)).$$

Expressions for the adjustment terms, $e^L(y_t)$, $e^H(y_t)$, as functions of y_t and the parameters α, $\mu_t \equiv t\mu$, and $\sigma_t^2 \equiv t\sigma^2$ are derived in equations (7) and (8). If the parameter values can be independently obtained, then the appropriate adjustment term can be computed for each y_t. Once the adjustment term is included in the estimating equation, the structural slope parameter, b, can be independently estimated, below and above the income threshold. If the structure we have imposed on the model is empirically valid, the estimated b parameter will be statistically indistinguishable below and above the threshold.

The value of the α parameter can be obtained directly from the estimated location of the threshold in the cross-sectional tests. To determine the value of t, recall from Figure 1 that economic development in India commenced in the middle of the twentieth century. If each generation spans 30 years, then the grandparents of current working-age adults would have been the first generation to experience development; i.e. we are now in generation $t = 3$ of the model. To estimate the parameters of the distribution of income shocks, μ and σ^2, we require data on the income distribution over multiple time periods or generations. The distribution of pre-tax national income is available from the World Inequality Database from 1951 onwards for India (Chancel and Piketty, 2017). Assuming that each generation spans 30 years, as above, we use the (real) income distribution in 1951, 1981, and 2011 and, in particular, the change in these distributions, to estimate the μ and σ parameters.[33]

Table 2 reports coefficient estimates from a piecewise linear equation, using IHDS all-India data, with child (aged 5-19) BMI-for-age in Columns 1-2 and adult BMI in Columns 3-4 as outcomes. In addition to household income, the standard covariates are included in each estimating equation. The slope-change in the estimating equation is imposed at the income level where the threshold was previously located, separately for each outcome. Columns 1 and 3 report benchmark estimates without including the $e^L(y_t)$, $e^H(y_t)$ adjustment terms. This specification is essentially the same as what we estimated earlier in Table 1, except that we now report the slopes below and above the threshold (rather than the slope-change). Columns 2 and 4 report estimates with the adjustment terms included in the estimating equation. The slope coefficients can now be interpreted as the structural, b, parameter in the model. Although we can easily reject the null hypothesis that the slopes below and above the threshold are equal in Columns 1 and 3, without the adjustment, we cannot reject the null once the adjustment terms are included. Indeed, the point estimates of the slope coefficient are now remarkably similar, below and above the threshold. A comparison of the point estimates indicates, in addition, that the slope without the adjustment term is less than (greater than) b, below (above) the threshold, as implied by Proposition 1.

[33]The World Inequality Database provides the 99 fractiles of the income distribution; p_0p_1, ..., $p_{98}p_{99}$, where p_xp_y refers to the average income between percentiles x and y, in each of the three years. We set the number of dynasties in the economy to be equal to 10,000. We draw 10,000 times from the 1951 income distribution, with each fractile being equally represented, to generate the initial income distribution. For a given value of μ and σ^2 this allows us to simulate the income distribution in 1981 and 2011. Our best estimate of the parameters of the income-shock distribution is the value of μ and σ^2 for which the simulated income distribution in 1981 and 2011 matches most closely with the actual distribution.

Table 2: Piecewise Linear Equation Estimates - with and without adjustment terms

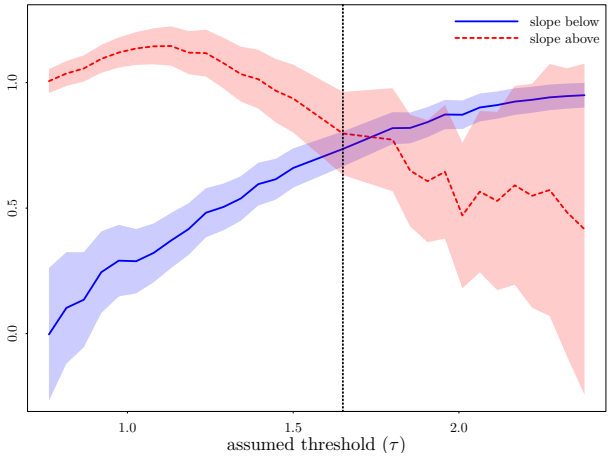
Dep. variable:	BFA 5-19		adult BMI	
Specification:	without adjustment (1)	with adjustment (2)	without adjustment (3)	with adjustment (4)
Slope below threshold (β_L)	0.030	0.118***	0.223***	0.735***
	(0.019)	(0.015)	(0.048)	(0.035)
Slope above threshold (β_H)	0.201***	0.101***	1.140***	0.797***
	(0.020)	(0.059)	(0.035)	(0.084)
$F-$statistic ($\beta_L = \beta_H$)	37.69	0.08	234.45	0.45
	[0.000]	[0.374]	[0.000]	[0.502]
Imposed threshold	1.90	1.90	1.65	1.65
N	48,986	48,986	76,949	76,949

Source: India Human Development Survey (IHDS)
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district, and survey-round are included in the estimating equation.
Least Squares standard errors are reported in parentheses and $p-$values associated with F-statistic are in square brackets.
$*$ significant at 10%, $**$ at 5% and $***$ at 1%

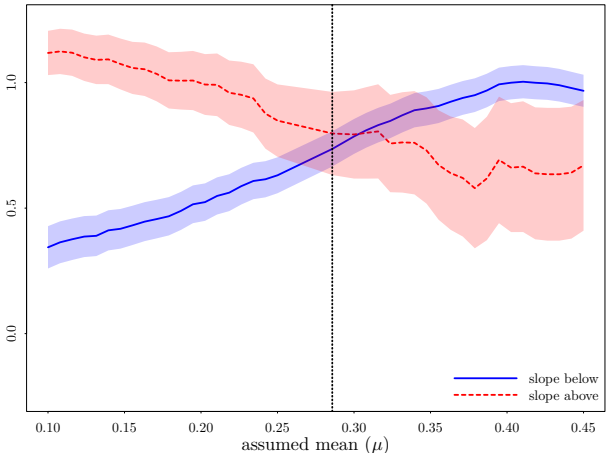
Figure 6 examines the sensitivity of the results in Table 2, with adult BMI as the outcome, to different values of the model's parameters. The panels in the figure report the estimated slope coefficient (with 95% confidence interval) below and above the threshold, for a range of values for each parameter: α, μ, σ, t. We see that the slope coefficients below (above) the threshold are increasing (decreasing) in the assumed parameter value for α, μ, t and are largely insensitive to the value of σ. The slope coefficients for α, μ, t coincide just around the values that we assign to these parameters in Table 2 (marked by the vertical lines in Figure 6). This indicates that all three parameter values need to line up precisely to equalize the slope coefficients in Table 2, which is especially striking given that these values are derived independently from different sources: the value of α is based on the income threshold location estimated with IHDS data, the value of μ is derived from the World Inequality Database, and t is based on the changes in per capita income over many centuries reported in Figure 1.

One benefit of the structural estimation is that it allows us to validate our modeling assumptions. An additional benefit is that it allows us to quantify the consequences of the nutrition trap. If the set point is irrelevant, there will be a linear relationship between household income and nutritional status: $\mathbb{E}(z_t) = a + by_t$. The estimated b parameter can thus be used to predict what nutritional status would have been in the absence of the nutrition trap. Figure 7a reports actual BMI-for-age, predicted BMI-for-age (based on the model), and the counter-factual BMI-for-age (in the absence of the nutrition trap) for children aged 5-19. Figure 7b reports the corresponding relationships with adult BMI as the outcome. The standard

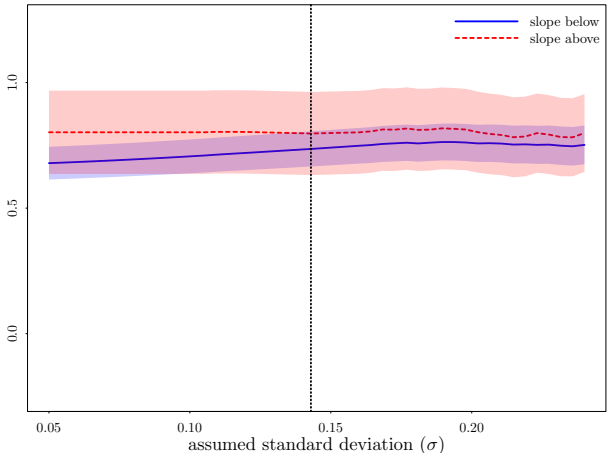
Figure 6: Sensitivity of Slope Coefficients with respect to Parameter Values



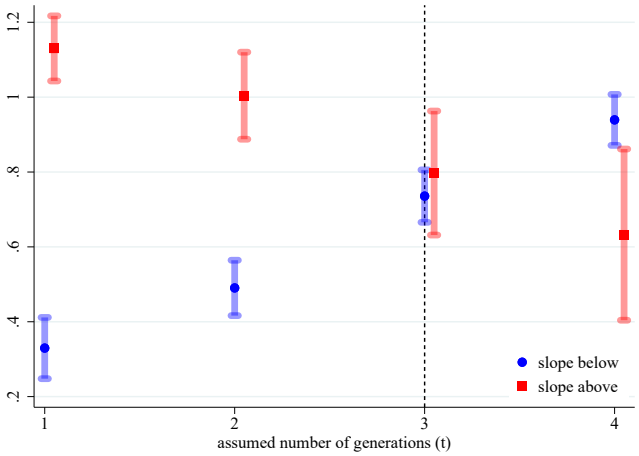
(a) Threshold



(b) Mean of the income shock



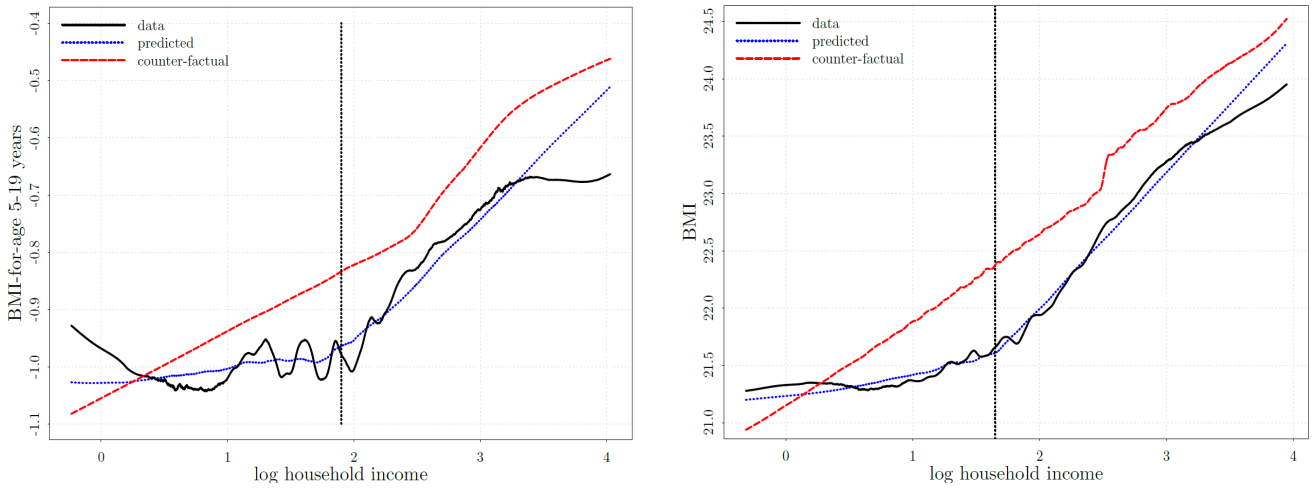
(c) Standard deviation of the income shock



(d) Number of generations

Notes: This figure plots the sensitivity of estimated slope coefficients, below and above the threshold, with respect to the value of each parameter of the model: (i) threshold location, (ii) mean of the income shock, (iii) standard deviation of the income shock, and (iv) the number of generations. The vertical line in each panel marks the parameter value that we use for estimation in Table 2.

Figure 7: Predicted Nutritional Status and Counter-factual Simulations



(a) BMI-for-age 5-19 years.

(b) BMI.

*Source:*India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district, and survey-round are partialled out prior to nonparametric estimation.

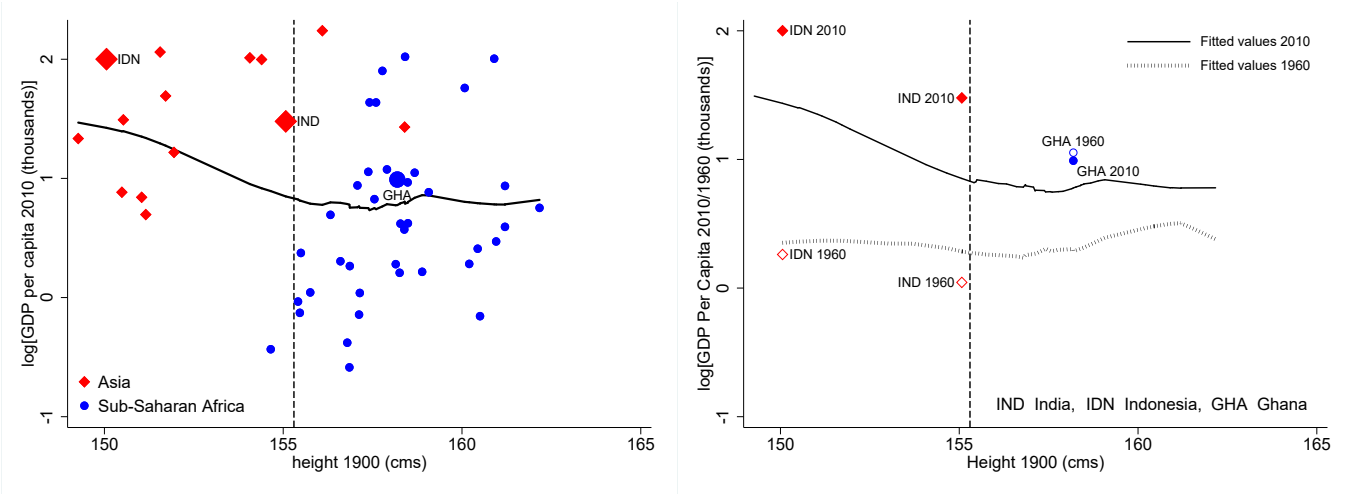
set of covariates are partialled out, and the dotted vertical line in each figure marks the location of the income threshold. Despite the model's parsimonious structure, and the simplifying assumptions we need to make to estimate its parameters, we see that the model fits the data very well. In our data, 22% of children are underweight (with a z-score below -2) and 20% of adults are underweight (with a BMI below 18.5). Based on the parameter estimates, the fraction of underweight children would decline by 10% and the fraction of underweight adults would decline by 24% if the set point were absent. The observed dampening of the nutritional status-current income relationship below the threshold, which we attribute to a predetermined set point, has important consequences for child and adult nutritional status in India.

3.4 External Validity

The presence of a set point is not unique to India. The next step in the analysis is thus to assess the applicability of the model to other developing countries. To test the cross-sectional implications of the model, the following data are required: (i) Household income, preferably at multiple points in time. (ii) Nutritional status (BMI). (iii) Indicators of metabolic disease. (iv) Household composition and detailed geographical indicators. The additional requirement is that a large sample is needed to locate a slope-change with precision. A search of publicly available data sets from other countries recovered two data sets that are suitable to test our model: the Indonesia Family Life Survey (IFLS) and the Ghana Socioeconomic Panel Survey (GSPS), although the GSPS does not contain information on metabolic disease.[34] We thus proceed to test the model with these two data sets, just as we did with the IHDS for India.

[34]Other well known data sets that we considered, but were determined to be unsuitable, include the Demographic Health Survey (DHS), the Living Standards Measurement Study (LSMS), Young Lives, and the China Health and Nutrition Survey (CHNS). As noted, the major limitation of the DHS for the purpose of our analysis is that it does not collect direct measures of income.

Figure 8: Current and Historical Income Across Countries



(a) GDP 2010.

(b) GDP 1960 and 2010.

Source: NCD-RisC and Penn World Table 9.0
Historical income is measured by height in the 1900 birth cohort.

While a set point may be present in other countries, the fraction of the population that has escaped its pre-modern set point will depend on a country's stage in the process of development. In the initial phase, when current income is relatively close to pre-modern income, most of the population remains in the nutrition trap. In the intermediate phase, as observed for India, a substantial fraction of the population continues to remain in the nutrition trap, but now a large number of individuals have also crossed the income threshold. This stage of development is characterized by the co-existence of low nutritional status, conditional on current income, in one segment of the population and a high prevalence of metabolic disease in a different segment of the population. At later stages of development, most of the population will have escaped the nutrition trap. Given that epigenetic inheritance will cease after a few generations, the pre-modern set point will also be irrelevant by that point in time.

At what stage in the development process are Indonesia and Ghana or, equivalently, how does current income in those countries compare with historical (pre-modern) income? It is standard practice to use adult height as a proxy for income, and the standard of living, in historical research.[35] Recall from the model that nutritional status, measured by height and BMI, is increasing continuously in contemporaneous income in the pre-modern economy (period 0). This relationship only weakens in subsequent periods (generations) with economic development on account of the persistent set point. We thus use historical adult height to measure historical income. Figure 8a plots the relationship between per capita GDP in 2010 and adult height for individuals born in 1900, which is available for a number of developing countries including India, Indonesia, and Ghana.[36] The first point to take away from the figure is that there has been a reversal of

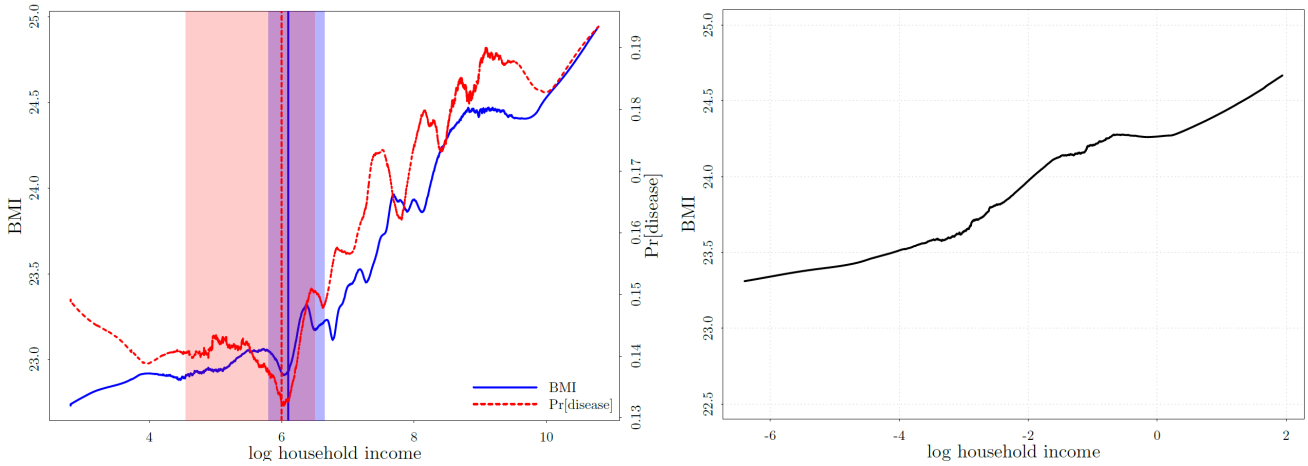
[35] As noted by Deaton (2007), genes are important determinants of individual height (and nutritional status more generally) but cannot explain variation across populations.

[36] We include all countries in South and South East Asia and Sub-Saharan Africa that satisfy the following requirement: their GDP per capita must be less than $10,000, which roughly corresponds to the upper bound for lower-middle income countries set by the World Bank. The same criterion is applied in the cross-regional analysis below.

fortunes over the past century, reflected by the negative relationship between current income and our proxy for historical income. The second point to take away from the figure is that the gap between current income and historical income is greater in Asia than in Africa. This is also true for the specific countries that we care about; the gap is greater for India and Indonesia than for Ghana (they have lower historical income but higher current income).

Figure 8b plots the relationship between per capita GDP, in 2010 and 1960, and adult height for the 1900 birth cohort. Notice that there is no apparent relationship between 1960 income and 1900 height across countries, in contrast with the negative relationship that is observed with 2010 income. Given the change in the slope over time, we expect that the sign would have been reversed – turning positive – if cross-country income data were available a few decades prior to 1960. Even if we restrict attention to the 1960-2010 period, it is evident that the gap between current and historical income is greater in Asian than in African countries (on either side of the vertical line). Focusing on individual countries, income in India and Indonesia increases substantially between 1960 and 2010, whereas it is unchanged in Ghana. Based on the preceding discussion, we expect that a substantial fraction of the Indonesian population will have crossed the income threshold, just as we observed for India. In contrast, we expect the population in Ghana to have remained (for the most part) at its pre-modern set point.

Figure 9: Nutritional Status and Metabolic Disease with respect to Income (Indonesia and Ghana)



(a) Nonparametric relationships (Indonesia). (b) Nonparametric relationship (Ghana)

Source: Indonesia Family Life Survey (IFLS), Ghana Socioeconomic Panel Survey (GSPS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, ethnicity (Indonesia) or tribe (Ghana), rural area, regency (indonesia) or district (Ghana), and survey-round are included in the estimating equation.
The vertical line marks the threshold location and the shaded region demarcates the cluster bootstrapped confidence interval.

Figure 9a nonparametrically estimates the relationships between adult BMI, the probability of metabolic disease, and household income using Indonesia Family Life Survey (IFLS) data. The same set of covariates that were included in the estimating equation with Indian data are included here as well, except that the district is replaced by the regency and caste is replaced by ethnicity. These covariates are partialled out, using Robinson's procedure, prior to nonparametric estimation. The IFLS has been conducted in five waves. To be consistent with the analysis using IHDS data in 2005 and 2011, the outcomes with IFLS

data are measured in the last two (2007 and 2014) waves. However, household income is averaged over all available waves to span as wide a time-window as possible and to smooth out transitory income shocks. The vertical lines in the figure mark the income levels at which Hansen's test locates thresholds for each outcome in Appendix Figure B7a and the shaded areas demarcate the corresponding confidence intervals. The estimated threshold locations are extremely close to each other, with an almost complete overlap in the confidence intervals. Moreover, as documented formally in Appendix Table B6, there is a weak association between household income and each outcome below the estimated threshold and a positive and significant slope-change above the threshold. As described above, the gap between current and historical income is even greater in Indonesia than in India. We would thus expect a larger fraction of the population to have escaped the nutrition trap in Indonesia and, based on our estimates of the threshold location, it appears that three-quarters of the Indonesian population has indeed crossed the threshold.

Figure 9b reports the nonparametric relationship between adult BMI and household income, using data from the Ghana Socioeconomic Panel Survey (GSPS). As noted, the GSPS does not collect data on metabolic disease. However, the full set of covariates that were used in the Indian and Indonesian analyses are available, with tribal affiliation replacing caste category and ethnicity, respectively. These covariates are partialled out prior to nonparametric estimation, as usual. The GSPS was conducted in three waves; 2009-2010, 2013, and 2017. The outcomes are measured in the 2009-2010 and 2013 waves, which correspond most closely to the IHDS waves, while household income is averaged over all three waves. In contrast with the discontinuous relationships that we estimated with Indian and Indonesian data, nutritional status is increasing smoothly with income in Figure 9b. Formal statistical support for this observation is provided in Appendix Figure B7b, where the Hansen test is unable to detect an income threshold. As reported in Appendix Table B6, there is a positive and statistically significant association between household income and adult BMI in Ghana. Where the Ghana data differ from the Indian and Indonesian data is that there is no slope change. Our interpretation of this finding, which is in line with the observation that current and historical incomes are relatively close in Africa is that the bulk of the Ghanaian population remains at its pre-modern set point. This will be shown to imply that the risk of metabolic disease conditional on BMI is relatively low in Africa, as verified with cross-country data in Section 5.

4 The Mechanism

Two biological relationships serve as the starting point for our model: (a) BMI is determined by ancestral income below a threshold and by current income above the threshold. (b) The risk of metabolic disease is constant below the threshold and increasing in the difference between current and ancestral income above the threshold. We next proceed to validate these relationships by constructing exogenous measures of ancestral income. The threshold location for this exercise is derived from the cross-sectional tests of the model; recall that households below the current income threshold remain at their set point.[37] As with the cross-sectional tests, we focus on India in the analysis that follows, but verify that the results hold up with Indonesian data

[37] Based on the model, some households above the current income threshold will also be at their set point. In a rapidly growing economy, however, most households above the threshold will have escaped the nutrition trap.

(with which a threshold can also be located).

An appealing feature of the cross-sectional tests of the model is that they do not require knowledge of the set point, y_0. This allowed us to include rural populations and urban populations (which include a large share of relatively recent migrants) in the analysis. When testing the biological relationships, however, we will need to link current income, y_t, to pre-modern ancestral income, y_0, and hence the analysis that follows is restricted to rural households. The implicit assumption is that rural households would have remained in their place of residence for many generations. While this may be true of the Indian sample, given that permanent migration by entire households in that country is especially low (Munshi and Rosenzweig, 2016), Indonesia has a long history of government sponsored internal migration (Bazzi et al., 2016). Despite this caveat, we obtain comparable results with Indian and Indonesian data.

4.1 District-Level Evidence

Measures of ancestral income going back many generations are unavailable at the family (dynasty) level. Our first measure of y_0 is constructed at the district level and is based on historical food supply. Agriculture was the dominant activity in the pre-modern economy and aggregate wealth would thus have been determined by crop productivity. The Food and Agriculture Organization Global Agro-Ecological Zones (FAO-GAEZ) project provides estimates of potential crop yields for 42 crops in 5 arc-minute by 5 arc-minute grid cells across the world. These grid cells can be matched to any administrative unit, such as a district, for which spatial shape files are available. Galor and Özak (2016) convert the potential yields to caloric production and then average across crops to construct a Caloric Suitability Index (CSI) which they document is a good indicator of the historical level of economic development or, equivalently, aggregate wealth across countries. We use the same index to measure pre-modern wealth at the district level, except that the baseline specification restricts attention to two staple crops – wheat and rice – that dominated historical agricultural production (and continue to account for a large share of agricultural production) in India.

The potential yields in the FAO-GAEZ database account for soil characteristics, temperature, moisture, and other growing conditions. They are provided for different levels of technology and for different levels of irrigation. We use low technology-rainfed agriculture to construct the CSI, as do Galor and Özak (2016), to match as closely as possible with pre-modern output. Aggregate wealth in the pre-modern economy would have determined the population (density) that could be supported in a local area (Diamond, 1997). If the CSI is a good measure of pre-modern aggregate wealth, then it should be closely related to historical population density. Appendix Figure B8a verifies this hypothesis at the district level by estimating a positive association between population density in 1951, when the Indian economy was just starting to develop, and CSI.[38]

While Appendix Figure B8a provides empirical support for our measure of historical aggregate wealth, it

[38]Recall from Figure 1 that the Indian economy only started to develop, after centuries of stagnation, in the middle of the twentieth century. The 1951 population densities, which are obtained from the first post-Independence census covering the entire country, will thus proxy for aggregate wealth prior to the onset of economic development. State fixed effects are partialled out in Appendix Figures B8a and B8b and are included in the estimating equations that follow to account for independent state-level determinants of historical population density, current household income, and the outcomes of interest (BMI and metabolic disease). For the analysis with Indonesian data, state fixed effects are replaced by regency fixed effects.

also indicates that the positive (and nonlinear) relationship between population and CSI must be accounted for when constructing measures of ancestral per household income. We do this by specifying that ancestral per household income is a flexible function of the CSI, $f(CSI)$. We then estimate the following equation:

$$y_t = f(CSI) + \epsilon_t, \tag{10}$$

where y_t is current household income, which is obtained as in the cross-sectional tests from the India Human Development Survey (IHDS), and CSI is based on the household's district of residence (the IHDS does not provide location identifiers below the district level). Equation (10) can be compared with the income equation (1) in the model:

$$y_t = y_0 + U_t.$$

Predicted income in equation (10) corresponds to ancestral income, y_0, and the residual in that estimating equation corresponds to the income mismatch, $U_t \equiv y_t - y_0$.[39] The objective when specifying the $f(CSI)$ function is to capture that part of the variation in current income that is explained by historical conditions and, by extension, ancestral per household income. Our preferred measure of y_0 will thus be predicted household income based on the most flexible nonparametric specification of the $f(CSI)$ function. Note that our measures of y_0, U_t are derived assuming, as in the model, that these variables are separable in the income equation. If places with higher y_0 started to develop earlier, for example, then they would have higher U_t, and our measures of y_0, U_t would be cross-contaminated. Such failure of the separability assumption would, however, lead to false rejection of the model and not the other way around.

Note also that spatial heterogeneity in y_0 is not inconsistent with the Malthusian model. Ashraf and Galor (2011) show theoretically that steady-state pre-modern per capita income would have varied with the predisposition towards having children and the cost of child rearing and there is recent empirical evidence in support of this argument (Dalgaard et al., 2021). Moreover, Diamond (1997) argues that greater agricultural productivity in the pre-modern period was associated with higher population densities and with more complex (vertically stratified) societies. Fertility would have varied by social class in such societies, with the elites consuming above subsistence. Once social stratification is incorporated in the Malthusian model, average per capita food consumption (income) will vary with agricultural productivity in a way that is theoretically ambiguous. In our data we document a nonmonotonic relationship (reasonably approximated by a quadratic function) between predicted household income and CSI in Appendix Figure B8b.

Table 3 reports the relationship between adult BMI, z_t, and both ancestral (predicted) income, y_0, and current income, y_t, below and above the estimated threshold. y_0 and y_t are normalized, by dividing by their respective standard deviations, to allow the magnitude of the income coefficients to be comparable. The standard set of covariates, with state fixed effects instead of district fixed effects since y_0 is measured at the district level, and with the exception of the rural-urban dummy since this is now a rural sample, are included in the estimating equations. The limitation of the district-level analysis is that ancestral income

[39] The residual, ϵ_t, is mean-zero by construction, whereas U_t has positive mean μ_t. Our estimates of y_0 and U_t are thus only identified up to a constant, but this has no bearing on the analysis that follows. Appendix Figure B9a uses binned scatter plots to (separately) describe the relationships between household income, y_t, and our measures of y_0 and U_t. These relationships are linear, matching the structure of the income equation (1) in the model.

Table 3: Nutritional Status - Income Relationship (below and above the threshold)

Dependent variable:	BMI			
Country:	India		Indonesia	
Sample:	Below	Above	Below	Above
Ancestral income	0.899***	0.165	1.059***	0.464
	(0.243)	(0.283)	(0.254)	(0.337)
Current income	0.185***	0.852***	-0.048	0.591***
	(0.040)	(0.047)	(0.119)	(0.064)
Threshold location	1.65	1.65	6.1	6.1
Dep. var. mean	20.482	21.851	22.317	23.021
N	27,164	20,296	3,182	10,610

Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group (India) or ethnicity (Indonesia), state (India) or regency (Indonesia) and survey-round are included in the estimating equation.The rural-urban dummy is excluded, since the sample is restricted to rural households.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

is constructed at an aggregate level and is based on a variable, CSI, that is not directly derived from pre-modern income. Nevertheless, as observed in Columns 1-2 with IHDS data, ancestral income has a positive and significant effect on adult BMI below the threshold (where households are at their set point) but not above it. Although the current income coefficient is also significant below the threshold, it is substantially smaller than the ancestral income coefficient and, moreover, is four times larger above the threshold.

Table 3, Columns 3-4, reports the adult BMI-income relationship with Indonesian (IFLS) data to verify its external validity. The analysis proceeds in exactly the same way as above, except that we restrict attention to a single staple crop – rice – which is by far the dominant crop in Indonesia. As noted, Indonesia has a long history of internal migration. While this could potentially weaken the relationship between our measure of ancestral income, which is based on the current place of residence, and nutritional status, the compensating advantage of the IFLS data is that they provide the sub-regency (sub-district) in which the household resides. The CSI can thus be constructed at a more disaggregate level than is possible with IHDS data. We see that the results in Columns 3-4 match closely with the biological relationships specified in the model. Ancestral income has a positive and significant effect on adult BMI below but not above the estimated threshold, whereas the converse is true for current income.

Current income in equation (10) can be decomposed into two orthogonal components: ancestral income, y_0, which is measured by predicted income and the income mismatch, $U_t \equiv y_t - y_0$, which is measured by the residual in that equation. While the standard assumption in models of developmental plasticity is that the risk of metabolic disease is increasing in the mismatch, the additional assumption in our model is that this relationship should be observed above but not below the threshold. The specification of the disease-income relationship in our model also assumes that there is no association between the risk of metabolic disease and ancestral income, below or above the threshold. To verify these assumptions, we estimate the

Table 4: Metabolic Disease - Income Relationship

Dependent variable:	Pr(metabolic disease)			
Country:	India		Indonesia	
Income component:	income mismatch (1)	ancestral income (2)	income mismatch (3)	ancestral income (4)
Income component	0.001	0.012*	-0.004	-0.011
	(0.002)	(0.006)	(0.011)	(0.019)
Income component × $\mathbf{1\{}$current income $> \boldsymbol{\tau}\mathbf{\}}$	0.018***	-0.002	0.032**	0.001
	(0.004)	(0.002)	(0.011)	(0.008)
Joint signficance				
$F-$statistic [$p-$value]	14.983	1.889	13.811	0.170
	[0.000]	[0.153]	[0.000]	[0.844]
Threshold location (τ)	1.90	1.90	6.00	6.00
Dep. var. mean	0.054	0.054	0.162	0.162
N	90,879	90,879	11,001	11,001

Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group (India) or ethnicity (Indonesia), state (India) or regency (Indonesia) and survey-round are included in the estimating equation.The rural-urban dummy is excluded, since the sample is restricted to rural households.
$F-$statistic measures the joint significance of the uninteracted and interacted income component coefficients.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

relationship between the probability of metabolic disease and (separately) each income component. The estimating equation includes the relevant income component as well as its interaction with a binary variable that indicates whether current income, y_t, exceeds the estimated threshold.

Table 4 reports results with Indian (IHDS) data in Columns 1-2 and Indonesian (IFLS) data in Columns 3-4. As assumed, the (uninteracted) income mismatch coefficient, which reflects the association with the risk of metabolic disease below the threshold, is economically and statistically insignificant in Columns 1 and 3. In contrast, the interaction coefficient, reflecting the change in the association above the threshold, is positive and significant in both columns. Moreover, as assumed once again, the ancestral income coefficients in Columns 2 and 4 are insignificant, with one exception (the uninteracted coefficient with Indian data in Column 2). Summarizing the estimation results and in line with the specification of the metabolic disease -income relationship in the model, we observe that the uninteracted and interacted coefficients are jointly significant in Columns 1 and 3, which measure the association between metabolic disease and the income mismatch, but jointly insignificant in Columns 2 and 4, where we measure the corresponding association with ancestral income.

We complete the district-level analysis by verifying the robustness of the results: (i) To a less flexible quadratic specification of the $f(CSI)$ function in Appendix Table B7 and Appendix Table B9. (ii) To construction of the CSI with an expanded set of major crops; wheat, rice, barley, sorghum, rye and millet

for India and rice, sorghum, cassava and maize for Indonesia. Estimates with these additional crops are reported in Appendix Table B8 and Appendix Table B10. As a final robustness test, we include volatility in historical district-level food supply, measured by the standard deviation of temperature and rainfall over the 1900-2010 period, in the estimating equation with BMI as the outcome. This allows for the possibility that volatility, together with average food supply, determined the set point. However, we do not obtain empirical support for this hypothesis, as reported in Appendix Table B11.

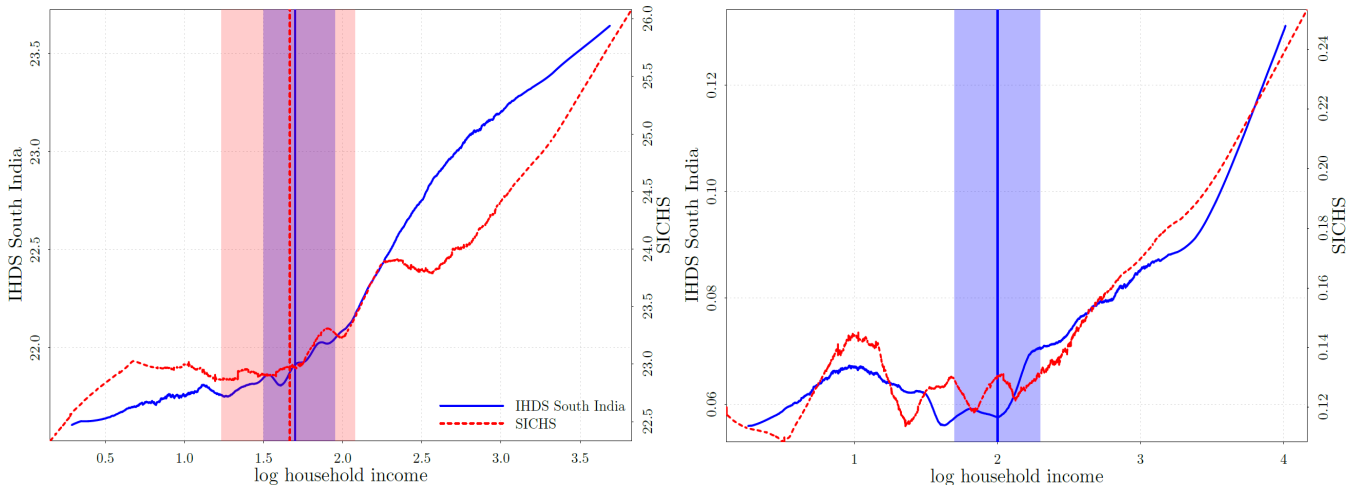
4.2 Village-Level Evidence

The district-level measures of ancestral income, y_0, allow us to validate both biological relationships specified in the model. The advantage of these measures is that they can be constructed, in a consistent fashion, using nationally representative data from India and Indonesia. However, as noted, the district and the sub-regency are aggregate spatial units and the CSI is not a direct measure of pre-modern income. We improve on both of these dimensions by using data from the South India Community Health Study (SICHS) for the analysis that follows. The SICHS covers a rural population of 1.1 million individuals residing in Vellore district in the South Indian state of Tamil Nadu. Two components of the SICHS are relevant for our analysis: a census of all 298,000 households residing in the study area, completed in 2014, and a detailed survey of 5,000 representative households, completed in 2016. The SICHS census collected each household's income in the preceding year. The SICHS survey collected information on the marriage of the household head and his spouse, and their parents, and in addition covers all variables included in the analysis using IHDS and IFLS data above. More importantly, the SICHS data are supplemented with historical records, obtained from the British Library in London, on the agricultural revenue tax per acre of cultivated land that was collected from each village in the Northern Tamil Nadu region (encompassing the study area) in 1871.[40] The revenue tax was based on a detailed assessment, made by the colonial government, of crop suitability, soil quality, precipitation, and other growing conditions. Like the CSI, this is a measure of potential agricultural productivity, but it is (i) defined at the village level, (ii) based explicitly on pre-modern growing conditions, and (iii) provides a direct measure of pre-modern income; i.e. the monetary value of agricultural output.

We begin the analysis with SICHS data by establishing that the cross-sectional relationships estimated above with nationally representative IHDS data are obtained in the study area as well. Figure 10a reports the association between adult BMI, for the household head and his spouse, and current household income. To smooth out transitory shocks, we take the average of the household income reported in the SICHS census and the SICHS survey as our measure of permanent household income. The standard set of covariates, excluding the district dummies and the rural-urban dummy since the rural sample is drawn from a single district, are partialled out prior to nonparametric estimation. The SICHS study area was purposefully selected to be representative of rural South India, defined as in Munshi and Rosenzweig (2016) by the states of Tamil Nadu, Andhra Pradesh, Karnataka, and Maharashtra, with respect to socioeconomic and

[40] There are 377 *panchayats* or village governments in the SICHS study area. These *panchayats* were historically single villages, which over time sometimes divided or added new habitations. The *panchayat* as a whole, which often consists of multiple modern villages, can thus be linked back to a single historical village. What we refer to as a "village" in the discussion that follows is thus a historical village or, equivalently, a modern *panchayat*.

Figure 10: Nutritional Status and Metabolic Disease with respect to Income (IHDS and SICHS)



(a) BMI (b) Metabolic disease

Source: India Human Development Survey (IHDS), South India Community Health Study (SICHS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, and (for IHDS) rural area, district and survey-round are included in the estimating equation.
The same set of covariates are included in the estimating equation at each assumed threshold for the threshold test.
The vertical lines mark the estimated threshold location and the shaded areas demarcate the corresponding 95% confidence intervals.

demographic characteristics.[41] As a basis for comparison, we thus report the corresponding nonparametric plot obtained with IHDS data, for the South Indian states, in Figure 10a. We go through the same steps as above to plot the relationship between the risk of metabolic disease and current income, with SICHS and IHDS South India data, in Figure 10b.

The estimated relationships, with SICHS and IHDS South India data, match very closely across the income distribution in both figures.[42] The vertical lines mark the spot where Hansen's test (shown in Appendix Figure B10) locates an income threshold, with the shaded area demarcating the associated 95% confidence interval. The threshold locations with adult BMI as the outcome are precisely estimated and almost identical with the two data sets. With the risk of metabolic disease as the outcome, in contrast, a threshold is precisely estimated with IHDS South India data but not SICHS data.[43] The tests that follow will thus be restricted to the adult BMI-income relationship.

One advantage of the SICHS analysis is that ancestral income can be measured at the village level.

[41] Borker et al. (2021) provide a detailed description of the study area, documenting that it is representative of rural Tamil Nadu and rural South India with respect to socioeconomic and demographic characteristics; e.g. age distribution, marriage patterns, literacy rates, labor force participation, child and adult sex ratios, and religious composition.

[42] BMI and the risk of metabolic disease are systematically higher with SICHS data relative to IHDS South India data (this can be observed by comparing the range of the Y-axes in Figure 10). In line with this finding, Alacevich and Tarozzi (2017) document that average heights for children under 5 are lower in the IHDS than in the Demographic Health Survey (DHS). They also document that heights and weight are measured with error in the IHDS, with heaping at particular focal points. Once we control for the level, however, the SICHS and the IHDS South India data track very closely with household income.

[43] This is because the sample size is much smaller with SICHS data and the threshold location is more difficult to estimate with the risk of metabolic disease as the outcome. For those outcomes for which thresholds can be located in Figure 10, the piecewise linear equation estimates at the estimated thresholds are reported in Appendix Table B12. In line with previous results, we cannot reject the hypothesis with South Indian (IHDS) data that the thresholds with BMI and metabolic disease as outcomes are located at the same income level.

However, this creates a new complication because ancestors can be drawn from multiple villages. Epigenetic inheritance was traditionally assumed to occur along the female line; i.e. via the mother, although recent evidence indicates that paternal traits can also be transmitted epigenetically (Jablonka and Raz, 2009). We allow for both possibilities, in which case ancestral incomes along the male and female line are relevant. Marriage in India is patrilocal, with women often leaving their natal (birth) village when they marry. In a patrilocal society, men do not move when they marry and, hence, ancestral income along the male line is determined by historical income in the individual's natal village. Ancestral income along the female line, in contrast, will be determined by historical income in the (possibly) many different villages from which the female ancestors were drawn.

To construct a single measure of ancestral income, we take advantage of the fact that families in rural India match assortatively on wealth (permanent income) in the marriage market, as documented with SICHS data by Borker et al. (2021). Although ancestral income, y_0, will not match perfectly on the male and female side in any marriage on account of the $U_t \equiv y_t - y_0$ term in the income equation, it will still be highly correlated for husbands and wives. We verify that this is the case, with SICHS data, for the household head and his spouse in Appendix Figure B11a and for their parents in Appendix Figure B11b, using the 1871 tax revenue in each individual's natal village to measure y_0.[44] The strong correlation in ancestral income that we document does not arise mechanically because couples are drawn from the same natal village. 80% of women in the SICHS study area leave their natal village when they marry, although almost all of them marry within the district, and we expect that similarly strong correlations in ancestral incomes would be observed if data from earlier generations were available. This implies that the 1871 tax revenue in any village from which ancestors were drawn could be used to construct y_0. To be consistent with our measure of current income, we use 1871 tax revenue in the current village of residence, both for the household head and his spouse, to construct their ancestral income.

The tax revenue per acre of cultivated land in 1871 measures historical wealth at the level of the village. As with the construction of the district level measure of ancestral per household income, we allow for an endogenous (village level) population response by specifying that per household ancestral income, y_0, is a flexible function, $g(R)$, of the 1871 tax revenue, R. The analysis thus proceeds in two steps: First, we estimate the relationship between current household income, y_t, and $g(R)$; the predicted income provides us with a measure of y_0, following the same argument as above. Second, we estimate the relationship between adult BMI and both y_0 and y_t, below and above the threshold located in Figure 10. As seen in Table 5, the ancestral income coefficient is positive and significant below, but not above, the threshold. In contrast, the current income coefficient is positive and significant above, but not below, the threshold. This result is robust to alternative (nonparametric and quadratic) specifications of the $g(R)$ function. Note that this result would not be obtained with the alternative developmental constraints model in which the set point is based on early-life conditions (income) in the current generation. We would expect our measure of current income to be a better proxy for those conditions than ancestral incomes measured in 1871, in which case current income would also determine BMI below the threshold.

[44] The household's ancestral income, y_0, is specified as a continuous function of the 1871 village-level tax revenue below. However, this has no bearing on the analysis of assortative matching.

Table 5: SICHS Nutritional Status - Income Relationship (below and above the threshold)

Dependent variable:	adult BMI			
$g(R)$ specification:	nonparametric		quadratic	
Sample:	below	above	below	above
	(1)	(2)	(3)	(4)
Ancestral income	0.334***	0.170	0.375***	0.026
	(0.124)	(0.150)	(0.128)	(0.123)
Current income	0.012	0.834***	0.048	0.834***
	(0.190)	(0.119)	(0.191)	(0.120)
Threshold location	1.69	1.69	1.69	1.69
Dependent var. mean	23.033	23.755	23.033	23.755
N	1810	3844	1810	3844

Source: South India Community Health Study (SICHS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender and caste group are included in the estimating equation. The rural-urban dummy and district dummies are excluded, since the rural sample is drawn from a single district.
Bootstrapped standard errors, clustered at the level of the village, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

We close this section by considering alternative explanations for the results in Table 5. The statistical challenge when testing the mechanism underlying the model is that the set point is determined by fixed pre-modern conditions. Even if these conditions are exogenously determined, they could still be associated with factors that independently determine the outcomes of interest. For example, village-level tax revenue in 1871, which we use to construct ancestral income, is also associated with pre-modern aggregate wealth and contemporaneous levels of economic development. These historical economic conditions could potentially determine nutritional status today through a variety of channels. A second alternative argument posits that ancestral income proxies for poorly measured current income at low income levels. However, any alternative mechanism must explain the additional restrictions imposed by our model: ancestral income should only be relevant below the estimated threshold and current income should only be relevant above the threshold. These sharp discontinuities cannot be explained by historical development or by measurement error, and they continue to hold in Appendix Table B13 even when ancestral income and current income are included separately in piecewise linear equations. The ancestral income coefficient continues to be positive and statistically significant below (but not above) the threshold, whereas the converse is true with current income.[45]

[45] We include ancestral income and current income, above and below the threshold, in Tables 3 and 5 to avoid the possibility that one variable simply proxies for the other. This is because ancestral income and current income are correlated by construction.

5 Cross-Regional Implications

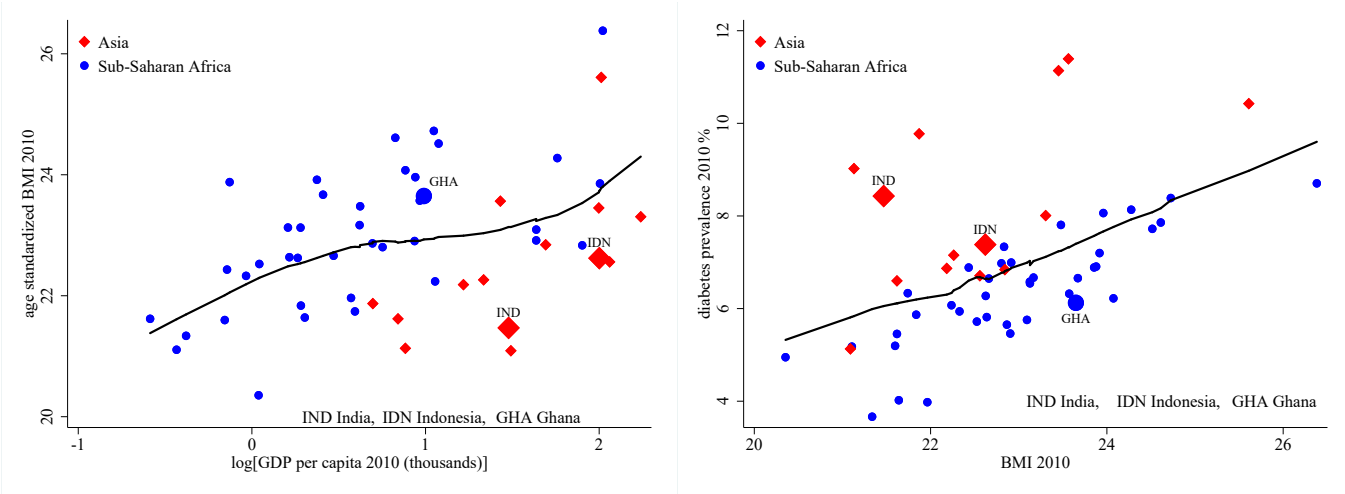
The nutrition-income puzzle that Deaton (2007) uncovered is that nutritional status, which he measures by height, is lower in South Asia than what would be predicted by GDP per capita, whereas the reverse is true for Africa. As shown below, our model, adapted to a cross-country setting with aggregate data can generate the same fact, but with BMI as the measure of nutritional status. We make the following assumptions in the aggregate model: (i) A fixed fraction of the population, π, remains within its set point in each country, j, in the current time period, t. This simplifying assumption is evidently at odds with the reported cross-country variation and we will discuss the consequences of relaxing it below. (ii) Each dynasty in country j has the same income, y_0^j, in the initial period, 0. (iii) Log income in the population in the current time period, y_t^j, is normally distributed; its mean, μ_t^j, varies across countries, but the variance, σ_t^2 is the same. Given these assumptions, and taking advantage of the properties of the normal distribution, equation (3) implies that average BMI in country j in the current period, z_t^j, can be expressed as a weighted average of initial income, y_0^j, and average current income, μ_t^j (the derivation is in Appendix A):

$$z_t^j = a + b\left[\pi y_0^j + (1-\pi)\left(\mu_t^j + \sigma_t \frac{\phi\left[\Phi^{-1}(\pi;0,1);0,1\right]}{1-\pi}\right)\right] \tag{11}$$

Taking expectations conditional on μ_t^j, $E(z_t^j|\mu_t^j)$ is increasing in $E(y_0^j|\mu_t^j)$. Looking back at Figure 8, if we drew a horizontal line through the figure at any level of current (average) income, it is evident that historical heights (which proxy for historical incomes) would be higher for African countries. $E(y_0^j|\mu_t^j)$ is higher in Africa, which implies that $E(z_t^j|\mu_t^j)$ is higher in Africa from equation (11). The qualifier to this argument is that a greater fraction of Asian populations have escaped their set point; i.e. π is not constant, but is in fact smaller for Asian countries. Given that $\mu_t^j > y_0^j$, this adjustment will weaken the preceding argument. Keeping this in mind, we plot average BMI against current GDP per capita in Figure 11a. Drawing a vertical line through the figure at any level of current income, BMI is indeed higher in African countries than in Asian (not just South Asian) countries.

Although other mechanisms have been proposed to explain the weak association between nutritional status and income in developing countries, an appealing feature of our mechanism, based on a biologically determined set point is that it also has implications for the emergence of metabolic diseases. The micro evidence, presented above, indicates that the risk of these diseases increases when (normal weight) individuals escape the nutrition trap. While we expect to observe this phenomenon in any developing economy, the prevalence of metabolic disease at a particular point in time will depend on the fraction of the population that has escaped the nutrition trap, together with the mismatch between current income and historical income for those who have escaped. We would expect these conditions to vary across populations, and the literature has indeed identified large differences in the prevalence of diabetes and related metabolic conditions. As with the nutrition literature, South Asians have received disproportionate attention. While diabetes was virtually nonexistent in South Asia until a few decades ago, rapid economic growth in India in particular has been accompanied by a substantial increase in the prevalence of the disease among normal weight adults (Narayan, 2017).

Figure 11: BMI - Current Income and Diabetes - BMI Relationship Across Countries



(a) BMI - current income

(b) Diabetes - BMI

Source: NCD-RisC and Penn World Table 9.0

Making the same assumptions as above, the aggregate version of the disease-income relationship specified in equation (4) can be expressed as:

$$D_t^j = \gamma_1 + \gamma_2(1-\pi)\left[\mu_t^j + \sigma_t \frac{\phi\left[\Phi^{-1}(\pi;0,1);0,1\right]}{1-\pi} - y_0^j\right], \tag{12}$$

where D_t^j is the fraction of the population in country j in the current period t that has contracted metabolic disease and $(1-\pi)$ is the fraction of the population that has escaped the nutrition trap and is at elevated risk of the disease. The term in square brackets in the preceding equation measures the average mismatch between current income and historical income (which determines the pre-modern set point) for individuals who have escaped the nutrition trap. As in equation (4), the risk of metabolic disease is increasing in this mismatch, whereas the risk is constant below the threshold.

Taking expectations conditional on average BMI, z_t^j, in equation (12), $E(D_t^j|z_t^j)$ is increasing in $E(\mu_t^j - y_0^j|z_t^j)$. Recall from Figure 8 that for any level of average current income, μ_t^j, average historical income, y_0^j, is higher in African countries than in Asian countries. We know from equation (11) that z_t^j is a weighted average of μ_t^j and y_0^j. Thus, if an African and Asian country have the same average BMI, then the Asian country must have higher μ_t^j and lower y_0^j. Based on this argument, $E(\mu_t^j - y_0^j|z_t^j)$ is higher in Asia than in Africa and, hence, $E(D_t^j|z_t^j)$ must be higher as well. This prediction is reinforced by the fact that a larger fraction of Asian populations have escaped the nutrition trap; i.e. $(1-\pi)$ is larger for Asian countries in equation (12). Figure 11b tests the preceding prediction by plotting diabetes prevalence against average BMI across countries. Diabetes prevalence is increasing with BMI across countries, both in Asia and in Africa. The important difference is that when we draw a vertical line through the figure at any BMI level, diabetes is higher in Asian countries than in African countries. Notice that while India is somewhat of an outlier in the figure, other Asian countries are even bigger outliers and not all of them are South Asian.

Although the diabetes literature has tended to focus on South Asians as a particularly vulnerable group, our analysis, as with the analysis of the BMI-income relationship, indicates that inter-regional differences in diabetes prevalence extend to the Asian continent as a whole.

6 Conclusion

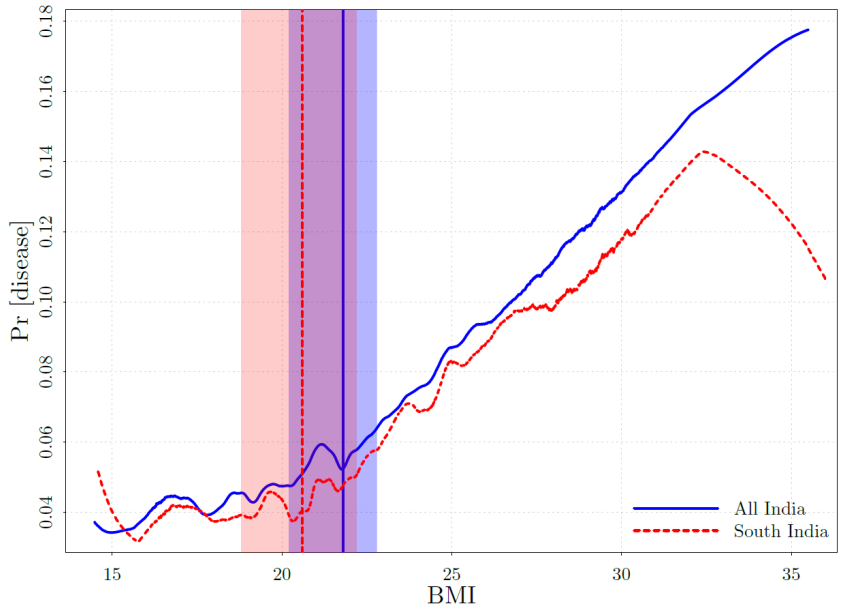
This paper proposes a single explanation for two seemingly unrelated observations in developing countries: (i) the weak association between BMI and income, lower in the income distribution where individuals are disproportionately underweight and (ii) the elevated risk of metabolic disease among normal weight individuals. Our explanation is based on a set point for BMI that is adapted to conditions in the pre-modern economy, but which fails to subsequently adjust to rapid economic change. This implies that during the process of development, the population will be divided into two distinct groups: Individuals in the first group remain at their low-BMI set point, despite the increase in their consumption. Individuals in the second group, who have escaped the nutrition trap, but are not necessarily overweight, are the primary contributors to the increased prevalence of metabolic disease.

To provide support for the preceding argument, we develop a model of nutrition and health that incorporates an individual-specific set point. A notable feature of the model is that it generates predictions for the cross-sectional BMI-income and metabolic disease-income relationships that can be tested without knowledge of the set point. These predictions are validated with micro-data from multiple countries; India, Indonesia, and Ghana. Additional tests verify the biological relationships that serve as the point of departure for the model. To complete the analysis, the model is adapted to aggregate data. This allows us to explain why BMI in Africa (Asia) is higher (lower) than what would be predicted by current GDP per capita, complementing Deaton's (2007) findings with height as the measure of nutritional status, while simultaneously explaining why there is a higher prevalence of diabetes, conditional on BMI, in Asian versus African countries.

Our structural estimates and accompanying counter-factual simulations for India, a country where both stylized facts have been well documented, indicate that the set point contributes substantially to undernutrition in children and adults. Nutrition supplementation in childhood would appear to be an obvious strategy to offset the set point, but it has been observed that nutrition programs are often ineffective, especially when targeted at older children (Schroeder et al., 1995; Victora et al., 2008). Our analysis provides an explanation for these findings; once the set point has been fixed, nutrition interventions will only be successful if they are intense enough to move individuals out of their set point and then remain permanently in place. Such interventions will also need to account for their potentially negative consequences for (future) chronic disease.

With or without nutrition interventions, an increasing fraction of the population will escape the nutrition trap in the coming decades with economic development, with an accompanying increase in the prevalence of metabolic disease. Screening will be an important component of public health programs targeting these conditions, and successful screening requires the at-risk population to be accurately identified. It has been recommended that the lower bound for the overweight range in Asian populations be reduced from 25 to

Figure 12: Metabolic Disease - BMI Relationship



Source: India Human Development Survey (IHDS)

23, to account for the fact that these populations are at elevated risk of metabolic disease at lower BMI (Deurenberg-Yap et al., 2002; Pan et al., 2004). However, this recommendation is not based on rigorous statistical analysis. Based on our model, there is no association between metabolic disease and BMI up to a BMI threshold, and a positive relationship thereafter. Figure 12 verifies this prediction with IHDS all-India and IHDS South India data, after partialling out the standard set of covariates. The vertical line marks the spot where the (precisely estimated) thresholds are located, which is at an extremely low BMI of 21.8 for all-India and 20.6 for South India. To put these findings in perspective, 11.2% of adult Indians have a BMI in the 21.8-23 range and 24.1% of South Indians have a BMI in the 20.6-23 range. While this suggests that the burden of metabolic disease in developing countries like India may be even greater than currently envisaged, the fact that many diabetics will thus have relatively low BMI's may also be an advantage. Recent evidence on diabetes reversal through a weight loss program (Taylor and Holman, 2015) indicates that there is an individual-specific BMI threshold, which is independent of initial BMI, below which diabetes is reversed. In a developing country population, we expect that this threshold will be associated with the pre-modern set point, which for many (lean) diabetics will not be far from their existing BMI.

References

Alacevich, C. and A. Tarozzi (2017): "Child Height and Intergenerational Transmission of Health: Evidence from Ethnic Indians in England," *Economics & Human Biology*, 25, 65–84.

Ashraf, Q. and O. Galor (2011): "Dynamics and Stagnation in the Malthusian Epoch," *American Economic Review*, 101, 2003–41.

Atkin, D. (2013): "Trade, Tastes, and Nutrition in India," *American Economic Review*, 103, 1629–63.

——— (2016): "The Caloric Costs of Culture: Evidence From Indian Migrants," *American Economic Review*, 106, 1144–81.

Banerjee, A. V. and A. F. Newman (1993): "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101, 274–298.

Barker, D. J. (1995): "Fetal Origins of Coronary Heart Disease." *BMJ: British Medical Journal*, 311, 171–174.

Bateson, P., P. Gluckman, and M. Hanson (2014): "The biology of developmental plasticity and the Predictive Adaptive Response hypothesis," *The Journal of Physiology*, 592, 2357–2368.

Battistin, E., R. Blundell, and A. Lewbel (2009): "Why is Consumption More Log Normal Than Income? Gibrat's Law Revisited," *Journal of Political Economy*, 117, 1140–1154.

Bazzi, S., A. Gaduh, A. D. Rothenberg, and M. Wong (2016): "Skill transferability, migration, and development: Evidence from population resettlement in Indonesia," *American Economic Review*, 106, 2658–98.

Behrman, J. R. and A. B. Deolalikar (1987): "Will developing country nutrition improve with income? A case study for rural South India," *Journal of political Economy*, 95, 492–507.

Borker, G., J. Eeckhout, N. Luke, S. Minz, K. Munshi, and S. Swaminathan (2021): "Wealth, Marriage, and Sex Selection," Tech. rep., Working Paper, Yale University.

Burgess, S. C. and D. J. Marshall (2014): "Adaptive parental effects: the importance of estimating environmental predictability and offspring fitness appropriately," *Oikos*, 123, 769–776.

Chancel, L. and T. Piketty (2017): "Indian Income Inequality, 1922-2014: From British Raj to Billionaire Raj?" *Working Paper*.

Cutler, D., A. Deaton, and A. Lleras-Muney (2006): "The Determinants of Mortality," *Journal of Economic Perspectives*, 20, 97–120.

Dalgaard, C.-J., J. B. Madsen, and H. Strulik (2021): "Physiological constraints and the transition to growth: implications for comparative development," *Journal of Economic Growth*, 26, 241–289.

Dandona, L., R. Dandona, G. A. Kumar, D. Shukla, V. K. Paul, K. Balakrishnan, D. Prabhakaran, N. Tandon, S. Salvi, A. Dash, and Others (2017): "Nations Within a Nation: Variations in Epidemiological Transition Across the States of India, 1990–2016 in the Global Burden of Disease Study," *The Lancet*, 390, 2437–2460.

Dasgupta, P. and D. Ray (1986): "Inequality as a Determinant of Malnutrition and Unemployment: Theory," *The Economic Journal*, 96, 1011–1034.

Deaton, A. (2007): "Height, Health, and Development," *Proceedings of the National Academy of Sciences*, 104, 13232–13237.

Diamond, J. (1997): *Guns, Germs and Steel: The Fates of Human Societies*, Vintage.

——— (2011): "Medicine: Diabetes in India," *Nature*, 469, 478–479.

Galor, O. and Ö. Özak (2016): "The Agricultural Origins of Time Preference," *American Economic Review*, 106, 3064–3103.

Galor, O. and J. Zeira (1993): "Income Distribution and Macroeconomics," *The Review of Economic Studies*, 60, 35–52.

Gluckman, P. D. and M. A. Hanson (2004): "Living With the Past: Evolution, Development, and Patterns of Disease," *Science*, 305, 1733–1736.

——— (2006): "The Conceptual Basis for the Developmental Origins of Health and Disease," in *Developmental Origins of Health and Disease*, ed. by P. Gluckman and M. Hanson, Cambridge University Press, 33 – 50.

Gujral, U. P., R. Pradeepa, M. B. Weber, K. V. Narayan, and V. Mohan (2013): "Type 2 Diabetes in South Asians: Similarities and Differences With White Caucasian and Other Populations," *Annals of the New York Academy of Sciences*, 1281, 51–63.

Hansen, B. E. (2017): "Regression Kink With an Unknown Threshold," *Journal of Business & Economic Statistics*, 35, 228–240.

Hardikar, A. A., S. N. Satoor, M. S. Karandikar, M. V. Joglekar, A. S. Puranik, W. Wong, S. Kumar, A. Limaye, D. S. Bhat, A. S. Januszewski, et al. (2015): "Multigenerational undernutrition increases susceptibility to obesity and diabetes that is not reversed after dietary recuperation," *Cell metabolism*, 22, 312–319.

Jablonka, E. and G. Raz (2009): "Transgenerational Epigenetic Inheritance: Prevalence, Mechanisms, and Implications for the Study of Heredity and Evolution," *The Quarterly Review of Biology*, 84, 131–176.

Jayachandran, S. and R. Pande (2017): "Why are Indian Children so Short? The Role of Birth Order and Son Preference," *American Economic Review*, 107, 2600–2629.

Leibel, R. (2008): "Molecular physiology of weight regulation in mice and humans," *International journal of obesity*, 32, S98–S108.

Luke, N., K. Munshi, A. Oommen, and S. Singh (2021): "Economic Development, the Nutrition Trap and Metabolic Disease," Tech. rep., National Bureau of Economic Research.

McKeigue, P., B. Shah, and M. Marmot (1991): "Relation of Central Obesity and Insulin Resistance With High Diabetes Prevalence and Cardiovascular Risk in South Asians," *The Lancet*, 337, 382–386.

Müller, M. J., A. Bosy Westphal, and S. B. Heymsfield (2010): "Is There Evidence for a Set Point That Regulates Human Body Weight?" *F1000 Medicine Reports*, 2.

Munshi, K. and M. Rosenzweig (2016): "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 106, 46–98.

Narayan, K. V. (2017): "Public Health Challenges for the 21st Century: Convergence of Demography, Economics, Environment and Biology: Nalanda Distinguished Lecture," *The National Medical Journal of India*, 30, 219.

Ng, S. W. and B. M. Popkin (2012): "Time Use and Physical Activity: A Shift Away From Movement Across The Globe," *Obesity Reviews*, 13, 659–680.

Oza-Frank, R. and K. V. Narayan (2010): "Overweight and Diabetes Prevalence Among US Immigrants," *American Journal of Public Health*, 100, 661–668.

Prentice, A. M., S. A. Jebb, G. R. Goldberg, W. A. Coward, P. Murgatroyd, S. Poppitt, and T. Cole (1992): "Effects of Weight Cycling on Body Composition," *The American Journal of Clinical Nutrition*, 56, 209–216.

Robinson, P. M. (1988): "Root-N-consistent Semiparametric Regression," *Econometrica*, 931–954.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019): "Fast and wild: Bootstrap inference in Stata using boottest," *The Stata Journal*, 19, 4–60.

Schroeder, D. G., R. Martorell, J. A. Rivera, M. T. Ruel, and J.-P. Habicht (1995): "Age differences in the impact of nutritional supplementation on growth," *The Journal of nutrition*, 125, 1051S–1059S.

Scrimshaw, N. S., C. E. Taylor, J. E. Gordon, W. H. Organization, et al. (1968): *Interactions of nutrition and infection*, World Health Organization.

Seckler, D. (1984): "The'small but healthy?'Hypothesis: a reply to critics," *Economic and Political Weekly*, 1886–1888.

Sekhri, S. and G. K. Shastry (2019): "The Curse of Plenty: Early Childhood Roots of the Rise in Chronic Disease," Tech. rep., Working Paper, University of Virginia.

Speakman, J. R., D. A. Levitsky, D. B. Allison, M. S. Bray, J. M. De Castro, D. J. Clegg, J. C. Clapham, A. G. Dulloo, L. Gruer, S. Haw, et al. (2011): "Set points, settling points and some alternative models: theoretical options to understand how genes and environments combine to regulate body adiposity," *Disease models & mechanisms*, 4, 733–745.

Strauss, J. and D. Thomas (1998): "Health, Nutrition, and Economic Development," *Journal of Economic Literature*, 36, 766–817.

Subramanian, S. and A. Deaton (1996): "The demand for food and calories," *Journal of Political Economy*, 104, 133–162.

Swaminathan, S., R. Hemalatha, A. Pandey, N. J. Kassebaum, A. Laxmaiah, T. Longvah, R. Lodha, S. Ramji, G. A. Kumar, A. Afshin, et al. (2019): "The burden of child and maternal malnutrition and trends in its indicators in the states of India: the Global Burden of Disease Study 1990–2017," *The Lancet Child & Adolescent Health*, 3, 855–870.

Taylor, R. (2013): "Type 2 Diabetes: Etiology and Reversibility," *Diabetes Care*, 36, 1047–1055.

Taylor, R. and R. R. Holman (2015): "Normal Weight Individuals Who Develop Type 2 Diabetes: The Personal Fat Threshold," *Clinical Science*, 128, 405–410.

Victora, C. G., L. Adair, C. Fall, P. C. Hallal, R. Martorell, L. Richter, H. S. Sachdev, Maternal, C. U. S. Group, et al. (2008): "Maternal and child undernutrition: consequences for adult health and human capital," *The Lancet*, 371, 340–357.

Voight, B. F., L. J. Scott, V. Steinthorsdottir, A. P. Morris, C. Dina, R. P. Welch, E. Zeggini, C. Huth, Y. S. Aulchenko, G. Thorleifsson, et al. (2010): "Twelve Type 2 Diabetes Susceptibility Loci Identified Through Large-Scale Association Analysis," *Nature Genetics*, 42, 579.

Wells, J. C., E. Pomeroy, S. R. Walimbe, B. M. Popkin, and C. S. Yajnik (2016): "The Elevated Susceptibility to Diabetes in India: An Evolutionary Perspective," *Frontiers in Public Health*, 4, 145.

A Mathematical Appendix

A.1 Proofs of Propositions

Proof of Proposition 1: We first derive closed-form expressions for $e^L(y_t)$, $e^H(y_t)$. The expression for $e^L(y_t)$ is given as

$$e^L(y_t) = \frac{1}{\Phi(y_t; \mu_t, \sigma_t^2)} \int_{-\infty}^{y_t} U_t \phi(U_t; \mu_t, \sigma_t^2) \, \mathrm{d}\, U_t$$

Focusing on the numerator we can write

$$\begin{aligned}\int_{-\infty}^{y_t} U_t \phi(U_t; \mu_t, \sigma_t^2) \, \mathrm{d}\, U_t &= \int_{-\infty}^{y_t} U_t \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left[-\frac{1}{2}\left(\frac{U_t - \mu_t}{\sigma_t}\right)^2\right] \mathrm{d}\, U_t \\ &= \int_{-\infty}^{\frac{y_t-\mu_t}{\sigma_t}} (\sigma_t x_t + \mu_t) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} x_t^2\right] \mathrm{d}\, x_t\end{aligned}$$

where the second equality comes from the substitution $x_t = \frac{U_t - \mu_t}{\sigma_t}$. The last equality can be written as

$$\mu_t \Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right) - \sigma_t \phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)$$

given that $\frac{\mathrm{d}\, \phi(x_t;0,1)}{\mathrm{d}\, x_t} = -x_t \phi(x_t; 0, 1)$. A similar transformation of $\Phi(y_t; \mu_t, \sigma_t^2)$ in the denominator of the $e^L(y_t)$ expression in (5) gives us the closed-form expression for $e^L(y_t)$ in equation (7). The corresponding expression for $e^H(y_t)$ in equation (8) is derived by replacing y_t with α in the upper limit for integration.

To establish that the slope of the BMI-income relationship is positive but less than b below the threshold, substitute the expression for $e^L(y_t)$ from equation (7) in equation (5) and differentiate with respect to y_t. Given the properties of the inverse Mill's ratio, BMI-income relationship for $y_t \leqslant \alpha$ is given as

$$\frac{\mathrm{d}\, \mathbb{E}(z_t | y_t; y_t \leqslant \alpha)}{\mathrm{d}\, y_t} = b\left[1 + \Lambda'\left(\frac{y_t - \mu_t}{\sigma_t}\right)\right] \in (0, b)$$

Further, to demonstrate that the slope of the BMI-income relationship above the threshold is greater than b, observe from the expression for $e^H(y_t)$ in equation (8), that the numerator is independent of y_t and the denominator is increasing in y_t. Hence, $\frac{\mathrm{d}\, e^H(y_t)}{\mathrm{d}\, y_t} < 0$, which implies $\frac{\mathrm{d}\, \mathbb{E}(z_t | y_t; y_t > \alpha)}{\mathrm{d}\, y_t} > b$ for $y_t > \alpha$.

Note, from equations (7) and (8), that $e^L(y_t) = e^H(y_t)$ at $y_t = \alpha$, and thus, from equations (5) and (6), there is no level discontinuity at the threshold. To prove that there is, nevertheless, a slope discontinuity at the threshold, $y_t = \alpha$, we need to show that

$$\lim_{y_t \uparrow \alpha} \frac{\mathrm{d}\, \mathbb{E}(z_t | y_t)}{\mathrm{d}\, y_t} \neq \lim_{y_t \downarrow \alpha} \frac{\mathrm{d}\, \mathbb{E}(z_t | y_t)}{\mathrm{d}\, y_t}$$

From equations (5) and (6), a necessary and sufficient condition for the preceding inequality to be satisfied is that $\frac{\mathrm{d}\, e^L(y_t)}{\mathrm{d}\, y_t} \neq \frac{\mathrm{d}\, e^H(y_t)}{\mathrm{d}\, y_t}$ at $y_t = \alpha$. Using equations (7) and (8), it can be established that this is indeed the case. For this result, first denote $v_t = \frac{y_t - \mu_t}{\sigma_t}$. From equation (7), $e^L(y_t) = \frac{\mathcal{L}(v_t)}{\Phi(v_t;0,1)}$, where

$\mathcal{L}(v_t) = \mu_t\Phi(v_t;0,1) - \sigma_t\phi(v_t;0,1)$. From equation (8), $e^H(y_t) = \frac{\mathcal{L}(\overline{v})}{\Phi(v_t;0,1)}$ where $\overline{v} = \frac{\alpha-\mu_t}{\sigma_t}$. Given that the denominator and the numerator (evaluated at $y_t = \alpha$) of the $e^L(y_t), e^H(y_t)$ expressions are the same, a necessary condition for $\frac{\mathrm{d}\, e^L(y_t)}{\mathrm{d}\, y_t} \neq \frac{\mathrm{d}\, e^H(y_t)}{\mathrm{d}\, y_t}$ is that $\frac{\mathrm{d}\,\mathcal{L}(v_t)}{\mathrm{d}\, y_t} \neq \frac{\mathrm{d}\,\mathcal{L}(\overline{v})}{\mathrm{d}\, y_t}$ at $y_t = \alpha$. $\frac{\mathrm{d}\,\mathcal{L}(\overline{v})}{\mathrm{d}\, y_t} = 0$. From the property of the standard normal distribution, $\phi'(v_t;0,1) = -v_t\phi(v_t;0,1)$, and, hence, $\left.\frac{\mathrm{d}\,\mathcal{L}(v_t)}{\mathrm{d}\, y_t}\right|_{y_t=\alpha} = \frac{\alpha}{\sigma_t}\phi(\overline{v};0,1) > 0$.

Proof of Proposition 2: From equation (4),

$$P(D_t|y_t; y_t \leqslant \alpha) = \int_{-\infty}^{y_t} \gamma_1 \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t = \gamma_1 \tag{A.1}$$

$$P(D_t|y_t; y_t > \alpha) = \int_{-\infty}^{\alpha} \gamma_1 \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t + \int_{\alpha}^{y_t} (\gamma_1 + \gamma_2 U_t)\frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t$$

$$= \gamma_1 + \gamma_2 \int_{\alpha}^{y_t} U_t \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t$$

Following the same steps that we used to derive the expression for $e^L(y_t)$ in (7), we can write

$$P(D_t|y_t; y_t > \alpha) = \gamma_1 + \gamma_2 \left[\mu_t - \sigma_t\Lambda\left(\frac{y_t-\mu_t}{\sigma_t}\right) - \frac{\mu_t\Phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right) - \sigma_t\phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right)}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)}\right] \tag{A.2}$$

From equation (A.1), $\frac{\mathrm{d}\,P(D_t|y_t;y_t\leqslant\alpha)}{\mathrm{d}\, y_t} = 0$, and from equation (A.2), $\frac{\mathrm{d}\,P(D_t|y_t;y_t>\alpha)}{\mathrm{d}\, y_t} > 0$ for $y_t > \alpha$ because $\Lambda'(\cdot) < 0$ and $\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)$ is increasing in y_t. This also establishes that there is a slope discontinuity at $y_t = \alpha$. Further, substituting $y_t = \alpha$ in equation (A.2) eliminates the term inside square brackets, implying that there is no level discontinuity at $y_t = \alpha$.

A.2 Placing an upper bound on y_0

BMI-income relationship: Assume that the period 0 income has both lower and upper bounds i.e. $y_0 \in [0, \overline{y}_0]$. Hence the range of U_t for any given value of y_t is $[y_t - \overline{y}_0, y_t]$. The mean BMI at any y_t, for $y_t \leqslant \alpha$, is given by

$$\begin{aligned}
\mathbb{E}(z_t|y_t; y_t \leqslant \alpha) &= \int_{y_t-\overline{y}_0}^{y_t} [a + b(y_t - U_t)]\frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t-\overline{y}_0;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t \\
&= a + by_t - b\int_{y_t-\overline{y}_0}^{y_t} U_t \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t-\overline{y}_0;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t \\
&= a + b(y_t - \overline{e}^L(y_t))
\end{aligned}$$

where $\overline{e}^L(y_t)$ corresponds to $e^L(y_t)$ in the model without an upper bound on y_0. Following the same steps as in the proof of Proposition 1 above:

$$\overline{e}^L(y_t) = \mu_t - \sigma_t \frac{\left[\phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right) - \phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right]}{\left[\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right) - \Phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right]} \tag{A.3}$$

For $y_t > \alpha$ there are two cases: (i) $y_t \in [\alpha, \overline{y}_0 + \alpha]$ and (ii) $y_t > \overline{y}_0 + \alpha$. In the first case, at each level of y_t, there are two types of individuals: those who remain at their set point and those who have crossed the threshold. The mean BMI at any y_t, for $y_t \in [\alpha, \overline{y}_0 + \alpha]$, is thus described by the following expression:

$$\begin{aligned}
\mathbb{E}\left(z_t|y_t; y_t \in [\alpha, \overline{y}_0 + \alpha]\right) &= \int_{y_t-\overline{y}_0}^{\alpha} [a + b(y_t - U_t)] \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\, U_t \\
&\quad + \int_{\alpha}^{y_t} [a + by_t] \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\, U_t \\
&= a + by_t - b \int_{y_t-\overline{y}_0}^{\alpha} U_t \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\, U_t \\
&= a + b(y_t - \overline{e}^H(y_t))
\end{aligned}$$

where $\overline{e}^H(y_t)$ corresponds to $e^H(y_t)$ in the model without an upper bound. As above, this expression can be simplified as

$$\overline{e}^H(y_t) = \frac{\mu_t\left[\Phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right) - \Phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right] - \sigma_t\left[\phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right) - \phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right]}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right) - \Phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)} \tag{A.4}$$

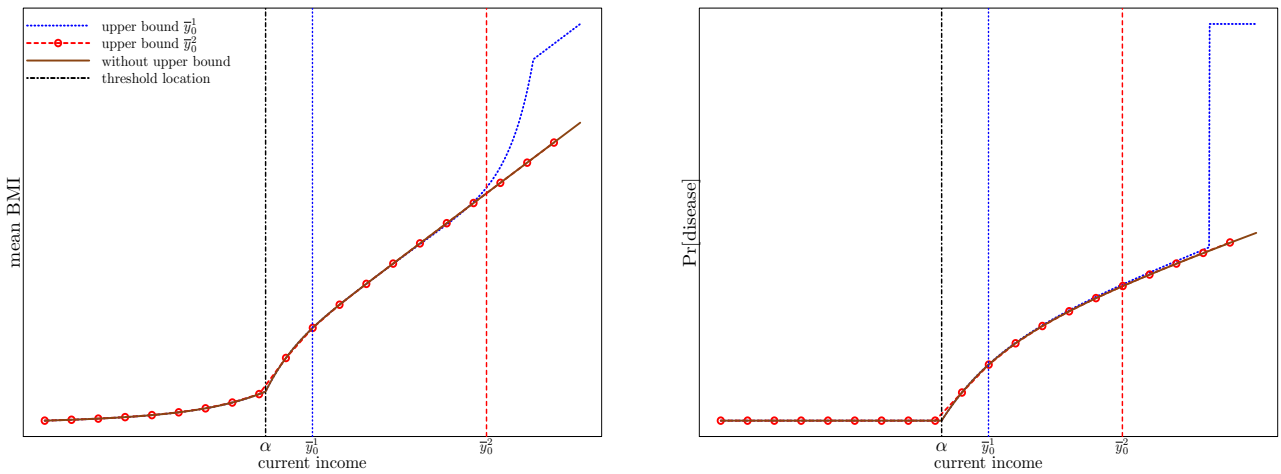
For $y_t > \overline{y}_0 + \alpha$, everyone has escaped the set point. Hence, the mean BMI at any y_t is

$$\begin{aligned}
\mathbb{E}(z_t|y_t; y_t > \ \overline{y}_0 + \alpha) &= \int_{\alpha}^{\infty} (a + by_t) \frac{\phi(U_t;\mu_t,\sigma_t^2)}{1 - \Phi(\alpha;\mu_t,\sigma_t^2)} \,\mathrm{d}\, U_t \\
&= a + by_t
\end{aligned}$$

Metabolic disease-income relationship: For $y_t \leqslant \alpha$,

$$\begin{aligned}
P(D_t|y_t; y_t \leqslant \alpha) &= \int_{y_t-\overline{y}_0}^{y_t} \gamma_1 \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\, U_t \\
&= \gamma_1
\end{aligned}$$

Figure A1: Simulated Cross-Sectional Relationships with upper bound on y_0



(a) BMI

(b) Metabolic disease

For $y_t \in [\alpha, \overline{y}_0 + \alpha]$,

$$\begin{aligned} P(D_t|y_t; y_t \in [\alpha, \overline{y}_0 + \alpha]) &= \int_{y_t - \overline{y}_0}^{\alpha} \gamma_1 \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2) - \Phi(y_t - \overline{y}_0; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t + \\ &\quad \int_{\alpha}^{y_t} (\gamma_1 + \gamma_2 U_t) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2) - \Phi(y_t - \overline{y}_0; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ &= \gamma_1 + \gamma_2 \int_{\alpha}^{y_t} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2) - \Phi(y_t - \overline{y}_0; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \end{aligned}$$

Solving the integral,

$$P(D_t|y_t; y_t \in [\overline{y}_0 + \alpha]) = \gamma_1 + \gamma_2 \frac{\mu_t \left[\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right) - \Phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right) \right] - \sigma_t \left[\phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right) - \phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right) \right]}{\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right) - \Phi\left(\frac{y_t - \overline{y}_0 - \mu_t}{\sigma_t}; 0, 1\right)} \tag{A.5}$$

For $y_t > \overline{y}_0 + \alpha$, as everyone has escaped their set point, we can write,

$$\begin{aligned} P(D_t|y_t; y_t > \overline{y}_0 + \alpha) &= \int_{\alpha}^{\infty} (\gamma_1 + \gamma_2 U_t) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{1 - \Phi(\alpha; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ &= \gamma_1 + \gamma_2 \left[\mu_t + \sigma_t \frac{\phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right)}{1 - \Phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right)} \right] \end{aligned} \tag{A.6}$$

which is independent of y_t.

Although analytical results can no longer be derived as in Propositions 1 and 2, expressions (7), (8), (A.3), (A.4), (A.5) and (A.6) can be used to simulate the relationship between current income and both BMI and the probability of metabolic disease. We use the actual income from the IHDS and the estimates of μ_t, σ_t from the structural estimation exercise for the simulation. The left panel in Figure A1 plots the

relationship between BMI and current income, with and without the upper bound on y_0. For the upper bound we choose two values of $\overline{y}_0$. The first value $\overline{y}_0^1$, marked by the blue dotted vertical line, is close to the threshold α whereas the second value $\overline{y}_0^2$, marked by the red dashed line, is further to the right. The simulated BMI-income and metabolic disease-income relationships track together, almost exactly, with the three specifications, except in the right tail of the income distribution where we observe a second discontinuity with $\overline{y}_0^1$. In our data, we do not observe a second discontinuity, at a high income level, with either BMI or the risk of metabolic disease as outcomes.

A.3 Alternative Specifications for the Set Point

A.3.1 Set point determined by ancestral and current income

Assume that a dynasty's set point is determined, each period, by the weighted average of ancestral income and current income. The relationship between BMI and income can now be written as

$$z_t = \begin{cases} a + b[r_t y_0 + (1 - r_t) y_t] & \text{if } \; y_t - [r_t y_0 + (1 - r_t) y_t] \leqslant \widetilde{\alpha} \\ a + b y_t & \text{if } \; y_t - [r_t y_0 + (1 - r_t) y_t] > \widetilde{\alpha} \end{cases} \tag{A.7}$$

where $r_1 = 1$ and $\lim_{t \to \infty} r_t = 0$. $y_t - [r_t y_0 + (1 - r_t) y_t] = r_t (y_t - y_0) = r_t U_t$. Hence, the threshold becomes time variant and is given by $\frac{\widetilde{\alpha}}{r_t}$. The mean BMI at $y_t \leqslant \frac{\widetilde{\alpha}}{r_t}$ can then be expressed as

$$\begin{aligned} \mathbb{E}\left[z_t | y_t; y_t \leqslant \frac{\widetilde{\alpha}}{r_t} \right] &= \int_{-\infty}^{y_t} (a + b[r_t y_0 + (1 - r_t) y_t]) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} dU_t \\ &= \int_{-\infty}^{y_t} (a + b[y_t - r_t U_t]) \frac{\phi(y_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} dU_t \\ &= a + b(y_t - r_t e^L(y_t)) \end{aligned}$$

where $e^L(y_t)$ is defined in (7). Similarly, for $y_t > \frac{\widetilde{\alpha}}{r_t}$, we can write

$$\begin{aligned} \mathbb{E}\left[z_t | y_t; y_t > \frac{\widetilde{\alpha}}{r_t} \right] &= \int_{-\infty}^{\frac{\widetilde{\alpha}}{r_t}} (a + b[y_t - r_t U_t]) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} dU_t + \int_{\frac{\widetilde{\alpha}}{r_t}}^{y_t} (a + b y_t) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} dU_t \\ &= a + b y_t - b r_t \int_{-\infty}^{\frac{\widetilde{\alpha}}{r_t}} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} dU_t \\ &= a + b(y_t - r_t \widetilde{e}^H(y_t)) \end{aligned}$$

where the expression for $\widetilde{e}^H(y_t)$ is the same as in equation (8) when α is replaced by $\frac{\widetilde{\alpha}}{r_t}$.

A.3.2 Set point determined by previous generation income

Assume that a dynasty's set point is determined, each period, by the previous generation's income. The relationship between nutritional status and income can be written as

$$z_t = \begin{cases} a + by_{t-1} & \text{if } \; y_t - y_{t-1} \leqslant \overline{\alpha} \\ a + by_t & \text{if } \; y_t - y_{t-1} > \overline{\alpha} \end{cases} \tag{A.8}$$

Assuming that $y_{t-1} \geqslant 0$, and using $u_t = y_t - y_{t-1}$ where $u_t \sim N(\mu, \sigma^2)$, we can write mean BMI at $y_t \leqslant \overline{\alpha}$ as

$$\begin{aligned} \mathbb{E}[z_t | y_t; y_t \leqslant \overline{\alpha}] &= \int_{-\infty}^{y_t} [a + by_{t-1}] \frac{\phi(u_t; \mu, \sigma^2)}{\Phi(y_t; \mu, \sigma^2)} du_t \\ &= a + by_t - b \int_{-\infty}^{y_t} u_t \frac{\phi(u_t; \mu, \sigma^2)}{\Phi(y_t; \mu, \sigma^2)} du_t \\ &= a + b(y_t - e^L(y_t; \mu, \sigma^2)) \end{aligned}$$

Similarly, mean BMI at $y_t > \overline{\alpha}$ is given as

$$\begin{aligned} \mathbb{E}[z_t | y_t; y_t > \overline{\alpha}] &= \int_{-\infty}^{\overline{\alpha}} [a + by_{t-1}] \frac{\phi(u_t; \mu, \sigma^2)}{\Phi(y_t; \mu, \sigma^2)} du_t + \int_{\overline{\alpha}}^{y_t} [a + by_t] \frac{\phi(u_t; \mu, \sigma^2)}{\Phi(y_t; \mu, \sigma^2)} du_t \\ &= a + by_t - b \int_{-\infty}^{\overline{\alpha}} u_t \frac{\phi(u_t; \mu, \sigma^2)}{\Phi(y_t; \mu, \sigma^2)} du_t \\ &= a + b(y_t - e^H(y_t; \mu, \sigma^2)) \end{aligned}$$

A.4 Aggregate BMI Equation Derivation

Let $\underline{y}_t^j$ denote the income threshold above which households escape their set point.

$$\pi = Pr[y_t^j \leqslant \underline{y}_t^j] = \Phi(\underline{y}_t^j; \mu_t^j, \sigma_t^2).$$

By the property of the normal distribution,

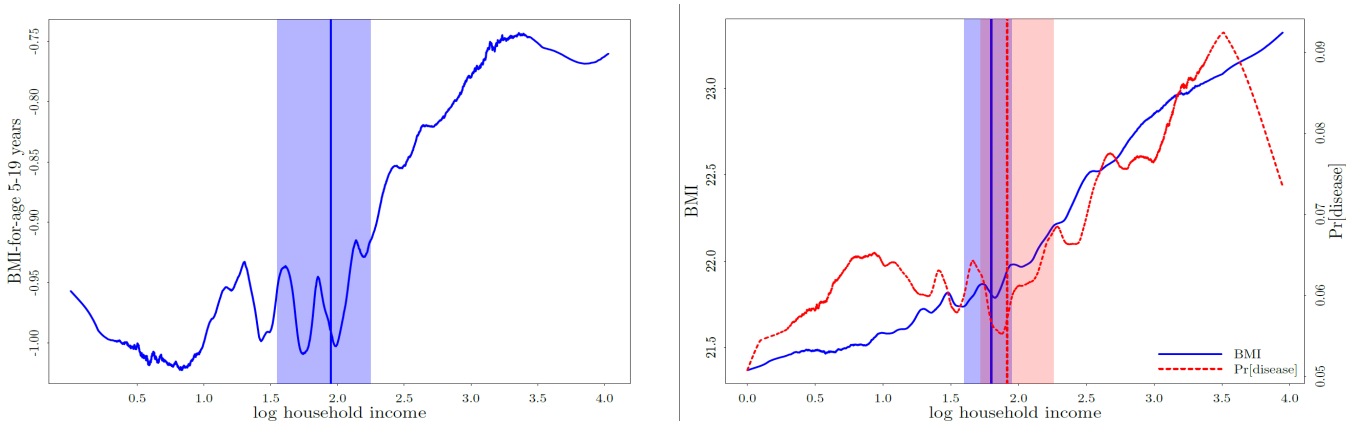
$$\underline{y}_t^j = \Phi^{-1}(\pi; \mu_t^j, \sigma_t^2) = \mu_t^j + \sigma_t \Phi^{-1}(\pi; 0, 1).$$

By the property of the normal distribution, once again, and substituting the expression for $\underline{y}_t^j$ derived above, average income above the threshold can be expressed as:

$$\mathbb{E}[y_t^j | \underline{y}_j^t < y_t^j < \infty] = \mu_t^j + \sigma_t \frac{\phi\left(\frac{\underline{y}_t^j - \mu_t^j}{\sigma_t}; 0, 1\right)}{1 - \Phi\left(\frac{\underline{y}_t^j - \mu_t^j}{\sigma_t}; 0, 1\right)} = \mu_t^j + \sigma_t \frac{\phi\left[\Phi^{-1}(\pi; 0, 1); 0, 1\right]}{1 - \pi}$$

B Appendix Figures and Tables

Figure B1: Nutritional Status and Metabolic Disease with respect to Household Income (adult education and household composition included as additional covariates)



(a) Children (b) Adults

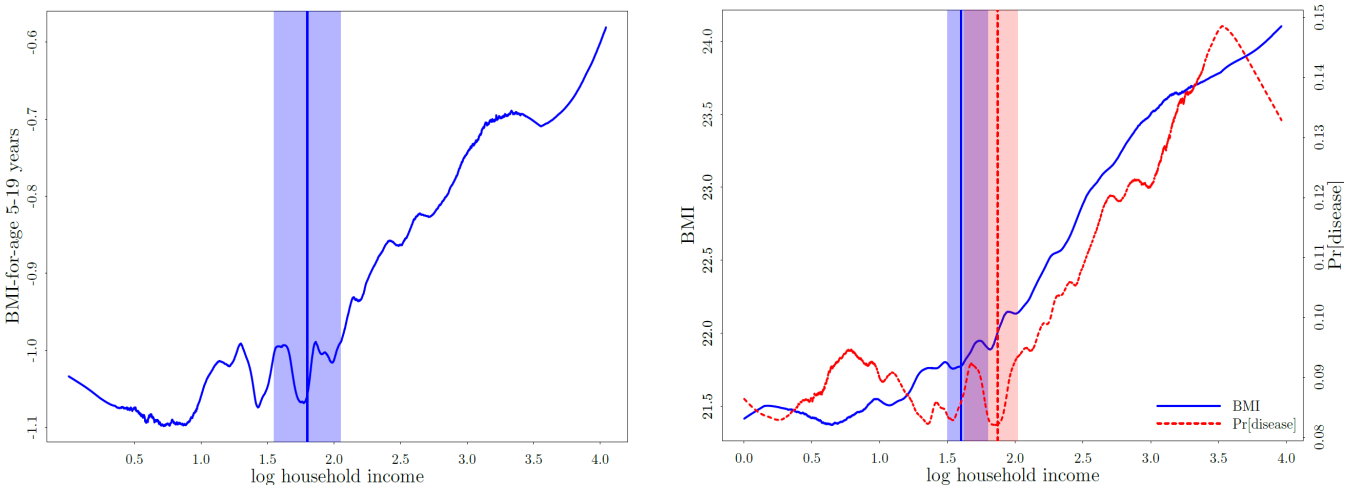
Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district, and survey-round are partialled out prior to nonparametric estimation.
Additional covariates include dummies for the number of adults, teens, and children in the household, dummies for the number of individuals engaged in manual labor, and dummies for the highest education of adult females and males.
The vertical lines mark the estimated threshold location and the shaded areas demarcate the corresponding confidence intervals.

Table B1: Piecewise Linear Equation Estimates (adult education and household composition included as additional covariates)

Dependent Variable:	BFA 5-19	adult BMI	metabolic disease
	(1)	(2)	(3)
Baseline slope (β_1)	0.020	0.281**	0.003
	(0.028)	(0.052)	(0.002)
Slope change (β_2)	0.148**	0.516**	0.014**
	(0.039)	(0.069)	(0.003)
Threshold location (τ)	1.95	1.80	1.95
	[1.55, 2.25]	[1.60, 1.95]	[1.75, 2.30]
Threshold test $p-$value	0.000	0.000	0.000
Mean of dependent variable	-0.932	22.002	0.074
N	48,986	76,949	148,928

Source: India Human Development Survey (IHDS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district, and survey-round are included in the estimating equation. Additional covariates include dummies for the number of adults, teens, and children in the household, dummies for the number of individuals engaged in manual labor, and dummies for the highest education of adult females and males.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses. Cluster bootstrapped 95% confidence bands for the threshold location are in brackets. ** significant at 5%, based on cluster bootstrapped confidence intervals.

Figure B2: Nutritional Status and Metabolic Disease with respect to Household Income (outcomes restricted to IHDS 2011-2012)



(a) Children (b) Adults

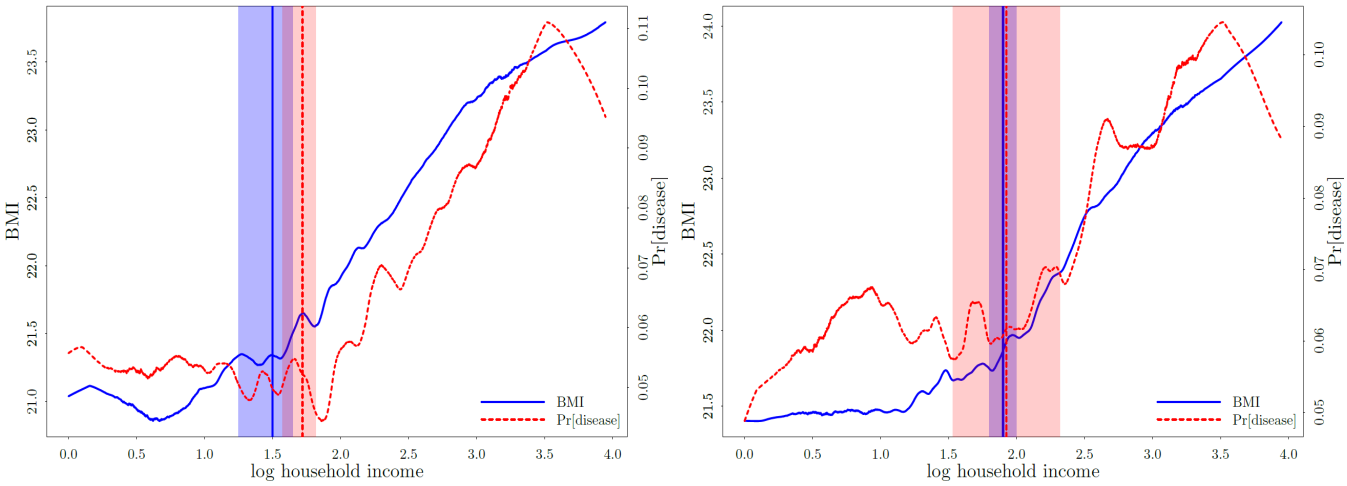
Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, and district are partialled out prior to nonparametric estimation.
The vertical lines mark the estimated threshold location and the shaded areas demarcate the corresponding confidence intervals.

Table B2: Piecewise Linear Equation Estimates (outcomes restricted to IHDS 2011-2012)

Dependent variable:	BFA 5-19	adult BMI	metabolic disease
	(1)	(2)	(3)
Baseline slope (β_1)	0.040	0.294**	-0.001
	(0.033)	(0.068)	(0.003)
Slope change (β_2)	0.172**	0.862**	0.036**
	(0.039)	(0.078)	(0.005)
Threshold location (τ)	1.80	1.60	1.90
	[1.55, 2.05]	[1.50, 1.80]	[1.65, 2.05]
Threshold test $p-$value	0.000	0.000	0.000
Mean of dependent variable	-0.965	22.189	0.098
N	35,810	53,006	74,166

Source: India Human Development Survey (IHDS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, and district are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
** significant at 5%, based on cluster bootstrapped confidence intervals.

Figure B3: Nutritional Status and Metabolic Disease with respect to Household Income (separately for men and women)



(a) Men (b) Women

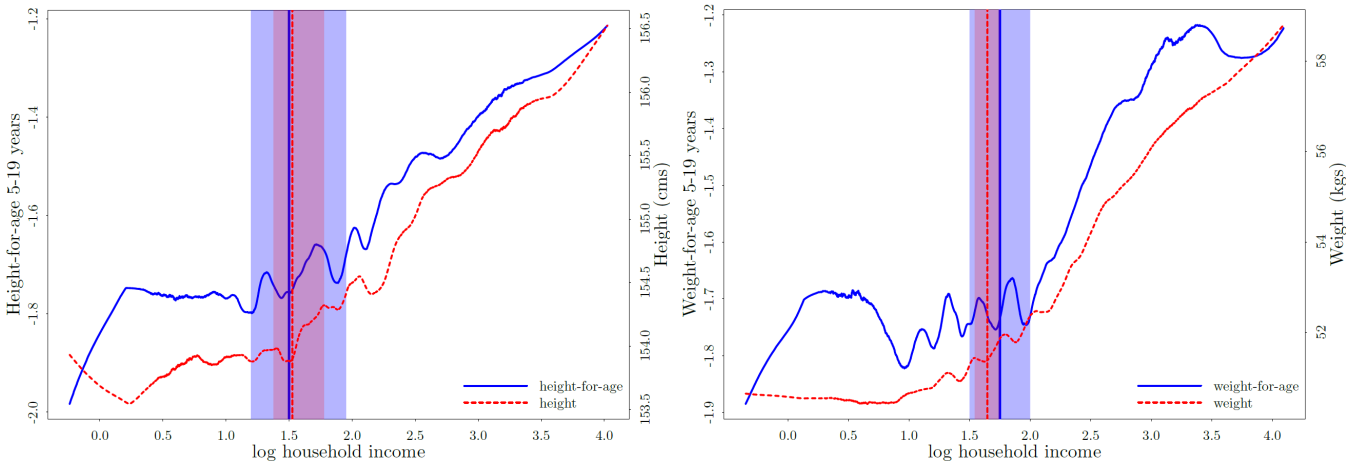
Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for caste group, rural area, district, and survey-round are partialled out prior to nonparametric estimation.
The vertical lines mark the estimated threshold location and the shaded areas demarcate the corresponding confidence intervals.

Table B3: Piecewise Linear Equation Estimates (separately for men and women)

Dependent variable:	adult BMI		metabolic disease	
Sample:	men	women	men	women
	(1)	(2)	(3)	(4)
Baseline slope (β_1)	0.342**	0.225**	−0.001	0.005
	(0.104)	(0.062)	(0.003)	(0.003)
Slope change (β_2)	0.877**	0.980**	0.038**	0.018**
	(0.112)	(0.079)	(0.004)	(0.005)
Threshold location (τ)	1.50	1.75	1.90	1.95
	[1.25, 1.65]	[1.60, 1.85]	[1.80, 2.00]	[1.55, 2.35]
Threshold test $p-$value	0.000	0.000	0.000	0.002
Mean of dependent variable	21.854	22.060	0.071	0.077
N	20,596	56,044	71,768	77,160

Source: India Human Development Survey (IHDS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for caste group, rural area, district, and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
** significant at 5%, based on cluster bootstrapped confidence intervals.

Figure B4: Nutritional Status with respect to Income (alternative nutritional status measures)



(a) Height (adults) and height-for-age (5-19 years)

(b) Weight (adults) and weight-for-age (5-19 years)

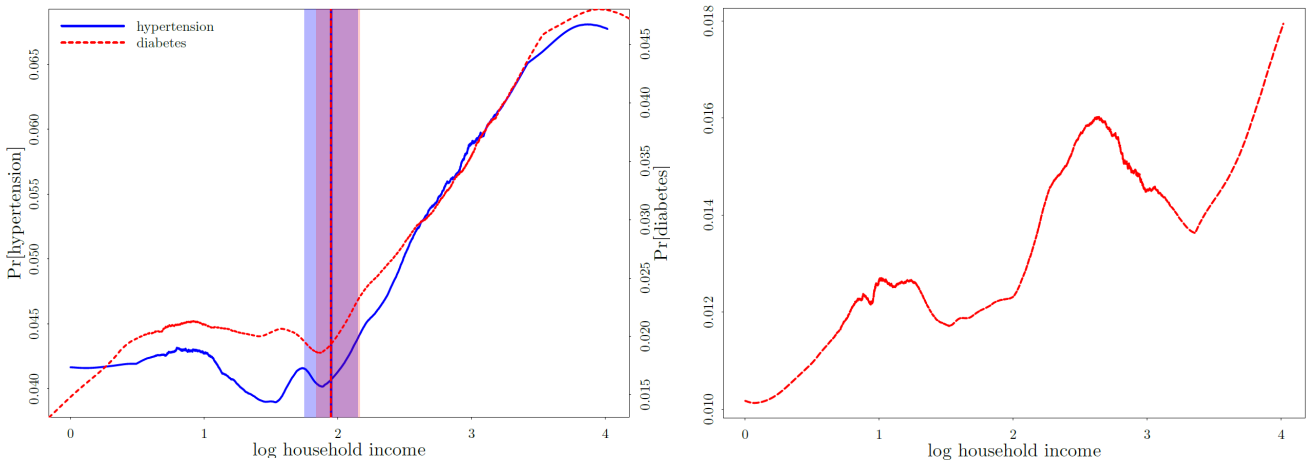
Source: India Human Development Survey (IHDS)
The following covariates are partialled out prior to the nonparametric estimation: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for children), caste group, rural area, district, and survey-round.
The vertical lines mark the estimated threshold locations and the shaded areas demarcate the corresponding cluster bootstrapped 95% confidence intervals.

Table B4: Piecewise Linear Equation Estimates (alternative nutritional status measures)

Alternative measure:	Height		Weight	
Dependent variable:	HFA 5-19	adult height	WFA 5-19	adult weight
	(1)	(2)	(3)	(4)
Baseline slope (β_1)	0.024	0.191	-0.004	0.656**
	(0.044)	(0.135)	(0.033)	(0.150)
Slope change (β_2)	0.206**	0.836**	0.331**	2.863**
	(0.045)	(0.144)	(0.041)	(0.174)
Threshold location (τ)	1.50	1.45	1.75	1.60
	[1.20, 1.95]	[1.30, 1.70]	[1.50,2.00]	[1.50,1.70]
Threshold test $p-$value	0.000	0.000	0.000	0.000
Mean of dependent variable	-1.649	154.483	-1.634	52.578
N	48,845	77000	23030	77143

Source: India Human Development Survey (IHDS)
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, birth order (for the children), caste group, rural area, district, and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
** significant at 5%, based on cluster bootstrapped confidence intervals.

Figure B5: Metabolic diseases (separately) with respect to income



(a) Hypertension/Diabetes (b) Cardiovascular disease

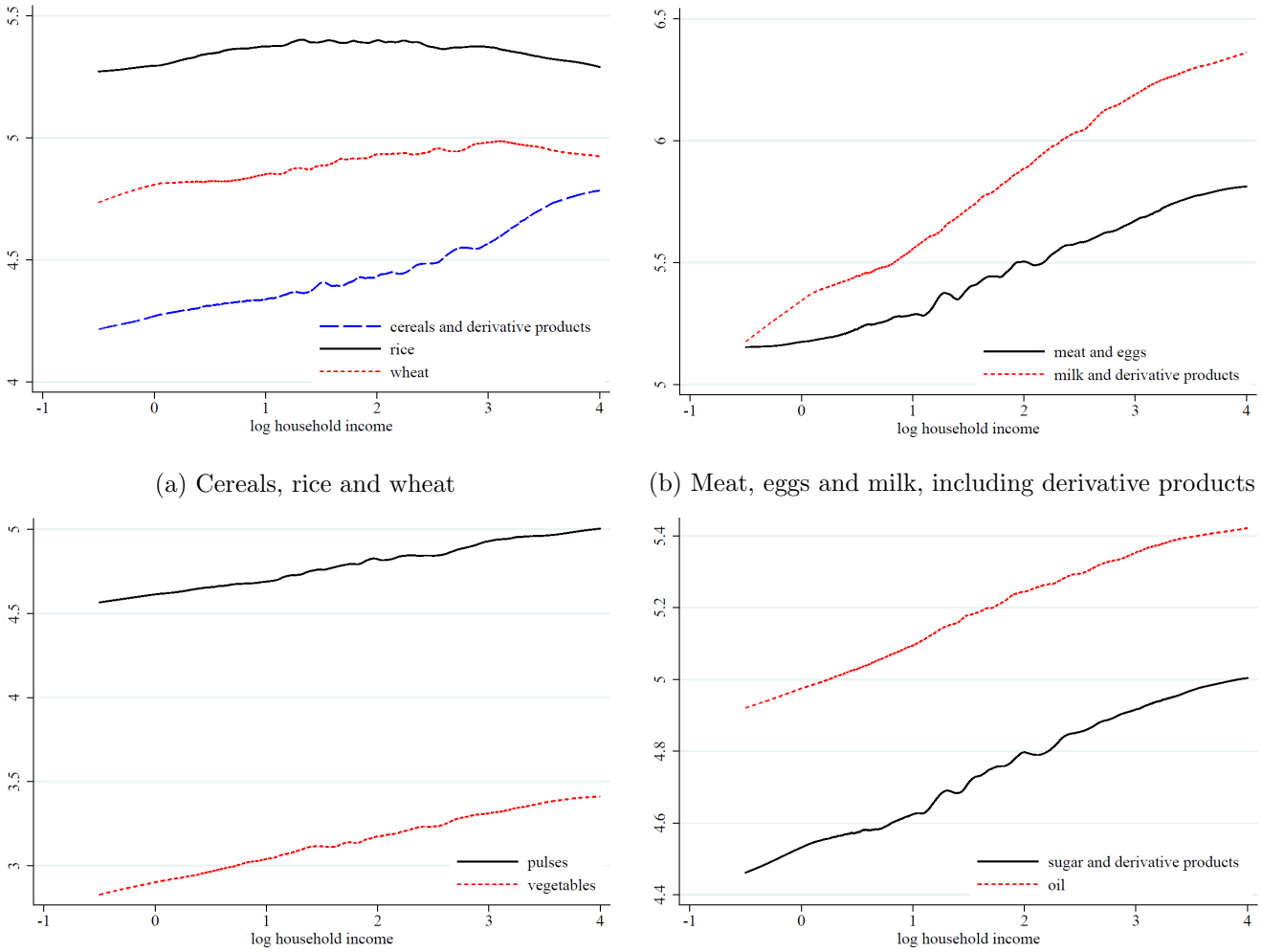
Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are included in the estimating equation.
The vertical lines mark the estimated threshold locations and the shaded areas demarcate the corresponding cluster bootstrapped 95% confidence intervals.

Table B5: Piecewise Linear Equation Estimates (hypertension and diabetes)

Dependent variable:	Hypertension (1)	Diabetes (2)
Baseline slope (β_1)	0.001	0.001
	(0.002)	(0.001)
Slope change (β_2)	0.018**	0.017**
	(0.003)	(0.002)
Threshold location (τ)	1.95	1.95
	[1.75, 2.15]	[1.85, 2.15]
Threshold test $p-$value	0.000	0.000
Mean of dependent variable	0.049	0.027
N	147,858	147,684

Source: India Human Development Survey (IHDS)
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
** significant at 5%, based on cluster bootstrapped confidence intervals.

Figure B6: Expenditure on different food categories with respect to household income



(a) Cereals, rice and wheat

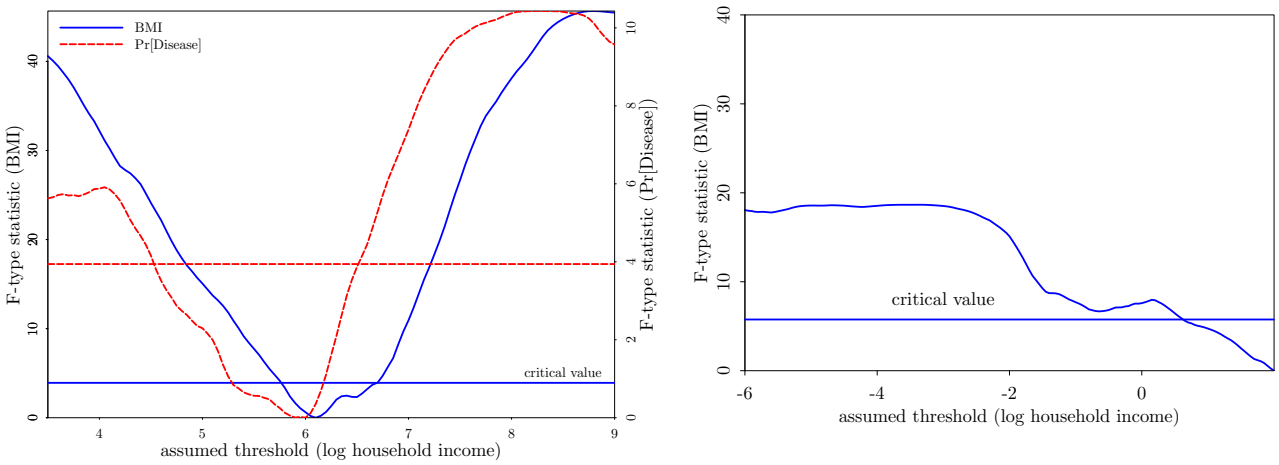
(b) Meat, eggs and milk, including derivative products

(c) Pulses and vegetables

(d) Oil and sugar and derivative products

Source: India Human Development Survey (IHDS).
This figure plots the nonparametric relationship between expenditures on different food categories and household income. Food expenditures are measured as the log of monthly expenditures in Rupees. The following covariates are partialled out prior to the nonparametric estimation: reported local price of rice, wheat, cereals and their derivative products, pulses, meat, sugar, oil, eggs, milk and its derivative products, vegetables and dummies for the number of children, adults, and teens in the household, occupation, caste group, rural area, district, and survey-round.

Figure B7: Nutritional Status and Metabolic Disease with respect to Income (Indonesia and Ghana)



(a) Threshold tests (Indonesia)

(b) Threshold test (Ghana)

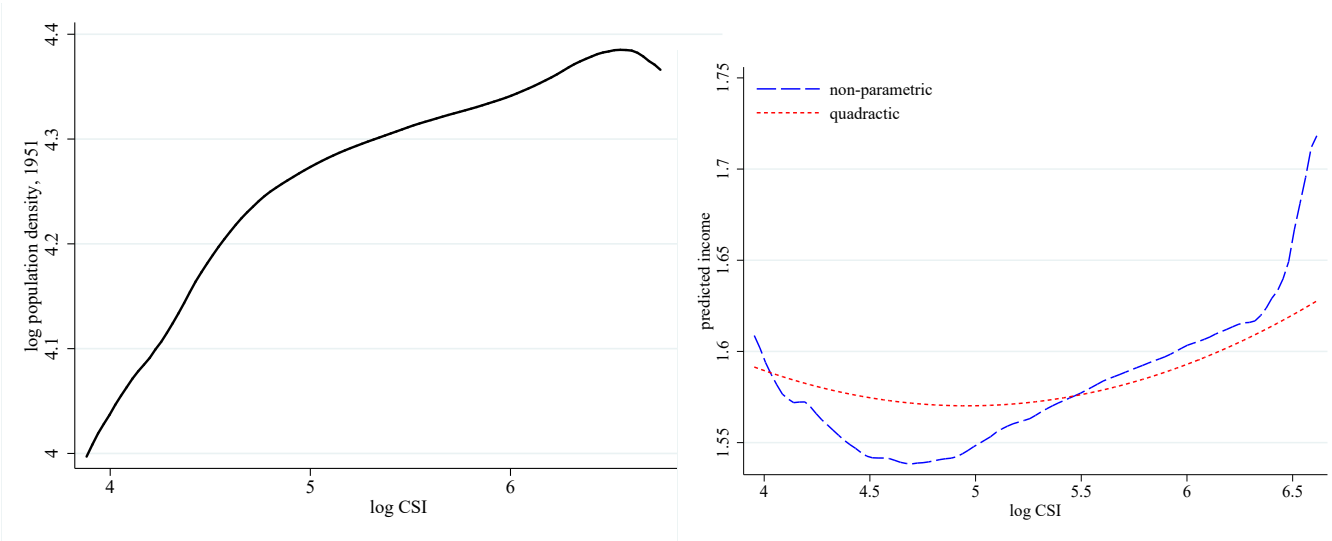
Source: Indonesia Family Life Survey (IFLS), Ghana Socioeconomic Panel Survey (GSPS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, ethnicity (Indonesia) or tribe (Ghana), rural area, regency (Indonesia) or district (Ghana), and survey-round are included in the estimating equation at each assumed threshold for the threshold test.
Indonesia: bootstrapped 5% critical values, clustered at the sub-regency level. Ghana: bootstrapped 5% critical values, clustered by enumeration area.

Table B6: Piecewise Linear Equation Estimates (Indonesia and Ghana)

Sample country:	Indonesia		Ghana
Dependent variable:	adult BMI	metabolic disease	adult BMI
	(1)	(2)	(3)
Slope below (β_L)	0.067	-0.001	0.165***
	(0.065)	(0.010)	(0.036)
Slope above (β_H)	0.398**	0.022**	–
	(0.069)	(0.011)	
Threshold location (τ)	6.10	6.00	–
	[5.80, 6.65]	[4.55, 6.50]	
Threshold test $p-value$	0.000	0.004	–
Dep. var. mean	23.532	0.181	23.934
N	30,812	24,788	11,372

Source: Indonesia Family Life Survey (IFLS), Ghana Socioeconomic Panel Survey (GSPS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
Logarithm of household income is the independent variable.
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, ethnicity (Indonesia) or tribe (Ghana), rural area, regency (Indonesia) or district (Ghana), and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the sub-regency level for Indonesia and by enumeration area for Ghana, are in parentheses.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
** significant at 5%, *** at 1%, based on cluster bootstrapped confidence intervals.

Figure B8: Relationship between Population Density, Predicted Household Income and Caloric Suitability Index (CSI)

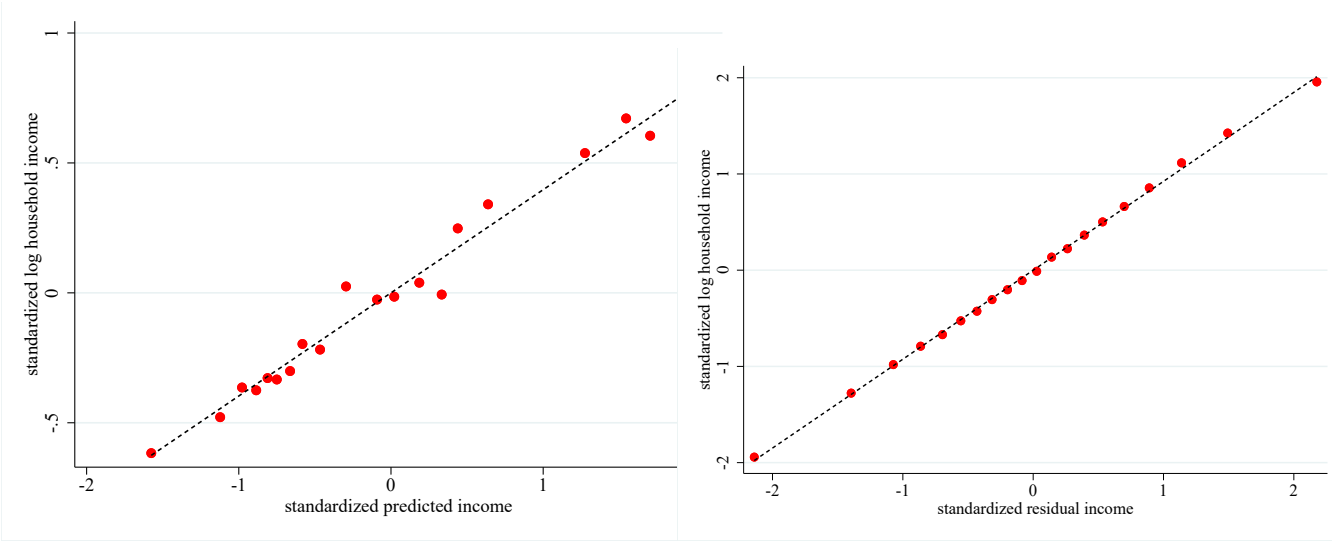


(a) Population density

(b) Predicted household income

Source: FAO-GAEZ dataset, 1951 population census, India Human Development Survey (IHDS)

Figure B9: Relationship between Household Income, Predicted Income, and Residual Income



(a) Predicted income

(b) Residual income

Source: FAO-GAEZ dataset, India Human Development Survey (IHDS)
This figure reports binned scatter plots describing the relationship between current household income, y_t, and (i) predicted income, which is our measure of y_0, and (ii) residual income, which is our measure of U_t. All variables are standardized.

Table B7: Nutritional Status - Income Relationship (below and above the threshold, quadratic $f(CSI)$ function)

Dependent variable:	BMI			
Country:	India		Indonesia	
Sample:	Below	Above	Below	Above
Ancestral income	0.530	-0.009	0.966***	0.339
	(0.243)	(0.202)	(0.354)	(0.475)
Current income	0.194***	0.854***	-0.048	0.601***
	(0.040)	(0.047)	(0.120)	(0.065)
Threshold location (τ)	1.65	1.65	6.1	6.1
Dep. var. mean	20.482	21.851	22.317	23.021
N	27,164	20,296	3,182	10,610

Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group (India) or ethnicity (Indonesia), state (India) or regency (Indonesia), and survey-round are included in the estimating equation.The rural-urban dummy is excluded, since the sample is restricted to rural households.
For India, staple crops are wheat and rice. For Indonesia, the staple crop is rice.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

Table B8: Nutritional Status - Income Relationship (below and above the threshold, additional crops)

Dependent variable:	BMI			
Country:	India		Indonesia	
Sample:	Below	Above	Below	Above
Ancestral income	0.538**	0.406	1.332***	0.552
	(0.193)	(0.210)	(0.281)	(0.381)
Current income	0.186***	0.848***	-0.051	0.589***
	(0.040)	(0.047)	(0.118)	(0.063)
Threshold location (τ)	1.65	1.65	6.1	6.1
Dep. var. mean	20.482	21.851	22.317	23.021
N	27,164	20,296	3,182	10,610

Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group (India) or ethnicity (Indonesia), state (India) or regency (Indonesia) and survey-round are included in the estimating equation.The rural-urban dummy is excluded, since the sample is restricted to rural households.
For India, staple crops are wheat, rice, sorghum, barley and millet. For Indonesia, staple crops are rice, sorghum, cassava and maize.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

Table B9: Metabolic Disease - Income Relationship (quadratic $f(CSI)$ function)

Dependent variable:	Metabolic disease			
Country:	India		Indonesia	
Income component:	Income mismatch	Ancestral income	Income mismatch	Ancestral income
	(1)	(2)	(3)	(4)
Income component	0.001	0.006	-0.003	0.016
	(0.002)	(0.006)	(0.011)	(0.037)
Income component ×				
$\mathbf{1\{}$income $> \boldsymbol{\tau}\mathbf{\}}$	0.018***	-0.002	0.030**	0.006
	(0.004)	(0.002)	(0.011)	(0.009)
Joint significance				
$F-$statistic [$p-$value]	16.070	0.734	12.699	0.341
	[0.000]	[0.481]	[0.000]	[0.711]
Threshold location (τ)	1.90	1.90	6.00	6.00
Dep. var. mean	0.054	0.054	0.162	0.162
N	90,879	90,879	11,001	11,001

Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group (India) or ethnicity (Indonesia), state (India) or regency (Indonesia) and survey-round are included in the estimating equation. The rural-urban dummy is excluded, since the sample is restricted to rural households.
For India, staple crops are wheat and rice. For Indonesia, staple crop is rice.
$F-$statistic measures the joint significance of the uninteracted and interacted income component coefficients.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

Table B10: Metabolic Disease - Income Relationship (additional crops)

Dependent variable:	Metabolic disease			
Country:	India		Indonesia	
Income component:	Income mismatch	Ancestral income	Income mismatch	Ancestral income
	(1)	(2)	(3)	(4)
Income component	0.001	0.005	-0.004	0.006
	(0.002)	(0.007)	(0.011)	(0.020)
Income component ×				
$\mathbf{1\{}$income $> \boldsymbol{\tau}\mathbf{\}}$	0.018***	-0.002	0.031**	0.004
	(0.003)	(0.002)	(0.011)	(0.009)
Joint significance				
$F-$statistic [$p-$value]	15.646	0.435	13.121	0.236
	[0.000]	[0.648]	[0.000]	[0.790]
Threshold location (τ)	1.90	1.90	6.00	6.00
Dep. var. mean	0.054	0.054	0.162	0.162
N	90,879	90,879	11,001	11,001

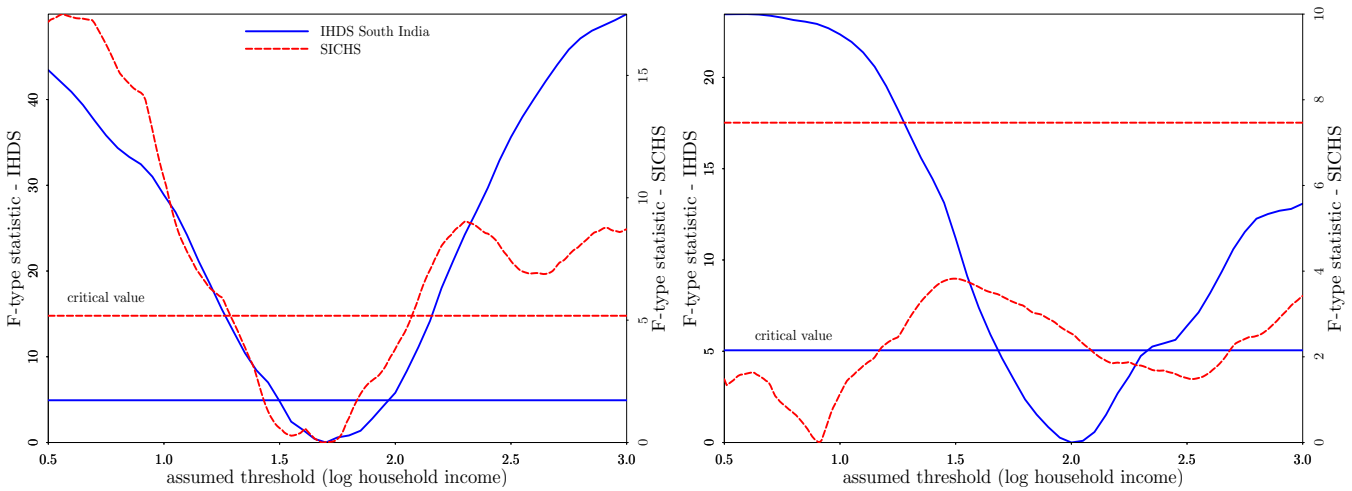
Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group (India) or ethnicity (Indonesia), state (India) or regency (Indonesia) and survey-round are included in the estimating equation.The rural-urban dummy is excluded, since the sample is restricted to rural households.
For India, staple crops are wheat, rice, sorghum, barley and millet. For Indonesia, staple crops are rice, sorghum, cassava and maize.
$F-$statistic measures the joint significance of the uninteracted and interacted income component coefficients.
. Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

Table B11: Nutritional Status - Income Relationship (below and above the threshold, conditional on climate volatility)

Dependent variable:	BMI			
Country:	India		Indonesia	
Sample:	Below	Above	Below	Above
Ancestral income	0.917***	0.166	1.045***	0.501
	(0.229)	(0.245)	(0.231)	(0.297)
Current income	0.185***	0.853***	-0.044	0.589***
	(0.040)	(0.047)	(0.119)	(0.063)
Temperature volatility	-2.740	-6.090**	1.298	-7.345
	(2.696)	(2.838)	(2.148)	(5.134)
Rainfall volatility	-0.003	0.015	-0.233	0.325
	(0.011)	(0.014)	(0.386)	(0.313)
Threshold location (τ)	1.65	1.65	6.1	6.1
Dep. var. mean	20.482	21.851	22.317	23.021
N	27,164	20,296	3,182	10,610

Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: rainfall and temperature volatility at the district level for India, and the sub-regency level for Indonesia, age (linear, quadratic, and cubic terms) and dummies for gender, caste group (India) or ethnicity (Indonesia), state (India) or regency (Indonesia) and survey-round are included in the estimating equation.The rural-urban dummy is excluded, since the sample is restricted to rural households.
For India, staple crops are wheat and rice. For Indonesia, the staple crop is rice.
Rainfall and temperature volatility is calculated as standard deviations over the period 1901-2010.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

Figure B10: Threshold Tests - Nutritional Status and Metabolic Disease (IHDS and SICHS)



(a) adult BMI (b) metabolic disease

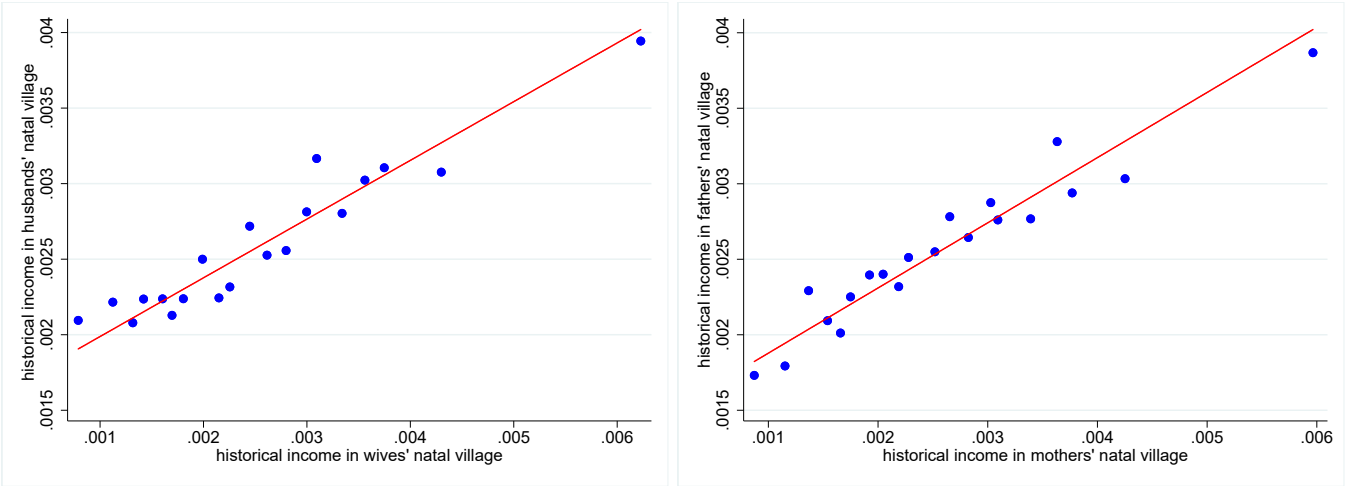
Source: India Human Development Survey (IHDS), South India Community Health Study (SICHS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, and (for IHDS) rural area, district and survey-round are included in the estimating equation at each assumed threshold for the threshold test.
Cluster bootstrapped 5% critical values are used to bound the threshold location.

Table B12: Piecewise Linear Equation Estimates – Nutrition Status and Metabolic Disease (South India)

Source:	IHDS		SICHS
Dependent variable:	adult BMI	metabolic disease	adult BMI
	(1)	(2)	(3)
Slope below (β_L)	0.200**	0.001	0.079
	(0.112)	(0.005)	(0.369)
Slope above (β_H)	0.803**	0.029**	1.148**
	(0.125)	(0.008)	(0.382)
Threshold location (τ)	1.70	2.00	1.69
	[1.50, 1.95]	[1.75, 2.25]	[1.29, 2.07]
Threshold test $p-$value	0.000	0.000	0.002
Dep. var. mean	22.186	0.074	23.449
N	22,316	41,198	7,634

Source: India Human Development Survey (IHDS), South India Community Health Study (SICHS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, and (for IHDS) rural area, district and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
** significant at 5%, based on cluster bootstrapped confidence intervals

Figure B11: Assortative Matching on Historical Income



(a) Current generation

(b) Parents' generation

Source: South India Community Health Study (SICHS)
Historical income is measured by tax revenue per acre of cultivated land in 1871 in the individual's natal village.
The number of bins in the binned scatter plot is set equal to 20.

Table B13: Nutritional Status - Income relationship below and above the threshold (SICHS)

Dep. var.:	adult BMI					
Income measure:	current income		ancestral income (non-parametric $g(R)$ function)		ancestral income (quadratic $g(R)$ function)	
Sample:	Below	Above	Below	Above	Below	Above
	(1)	(2)	(3)	(4)	(5)	(6)
Income coefficient	0.072	0.773***	0.334***	0.170	0.374***	0.014
	(0.161)	(0.102)	(0.123)	(0.159)	(0.129)	(0.127)
Dep. var. mean	23.034	23.755	23.034	23.755	23.034	23.755
N	1810	3844	1810	3844	1810	3844

Source: South India Community Health Study (SICHS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender and caste group are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the village, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

C Selective Child Mortality

Suppose that there is a positive and continuous relationship between mean nutritional status and household income, with a fixed dispersion in nutritional status at each level of income, as in Figure C1. If children can only survive above a subsistence nutrition level, and this constraint only binds at lower income levels, then as observed in the figure there will be a discontinuous relationship between mean nutritional status and income. Although the nutritional status-income relationship now precisely matches the prediction of our model, notice that it is driven entirely by households at the lower end of the nutritional status distribution, at each income level. Child mortality is concentrated in the first five years and, hence, if the nutritional status-income relationship is distorted by child mortality, this will show up most clearly among the 5-19 year olds. Figures C2a and C2b report quantile regression estimates of the baseline slope coefficient (β_1) and the slope-change coefficient (β_2) in a piecewise linear equation with child (aged 5-19) BMI-for-age as the dependent variable. Coefficient estimates for the same equation, evaluated at the mean of the dependent variable rather than at each quantile, were reported earlier in Table 1, Column 1. It is evident from Figures C2a and C2b that those results were not driven by a small fraction of households at the bottom of the nutritional status distribution, as the alternative explanation based on selective child mortality would imply. We cannot statistically reject the hypothesis that the estimated coefficients at each quantile are equal to the corresponding conditional mean coefficient.

Figure C1: Child Nutritional Status with respect to Income (with selective child mortality)

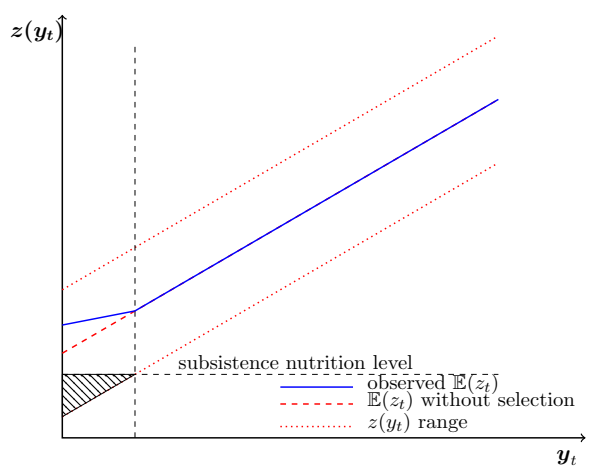
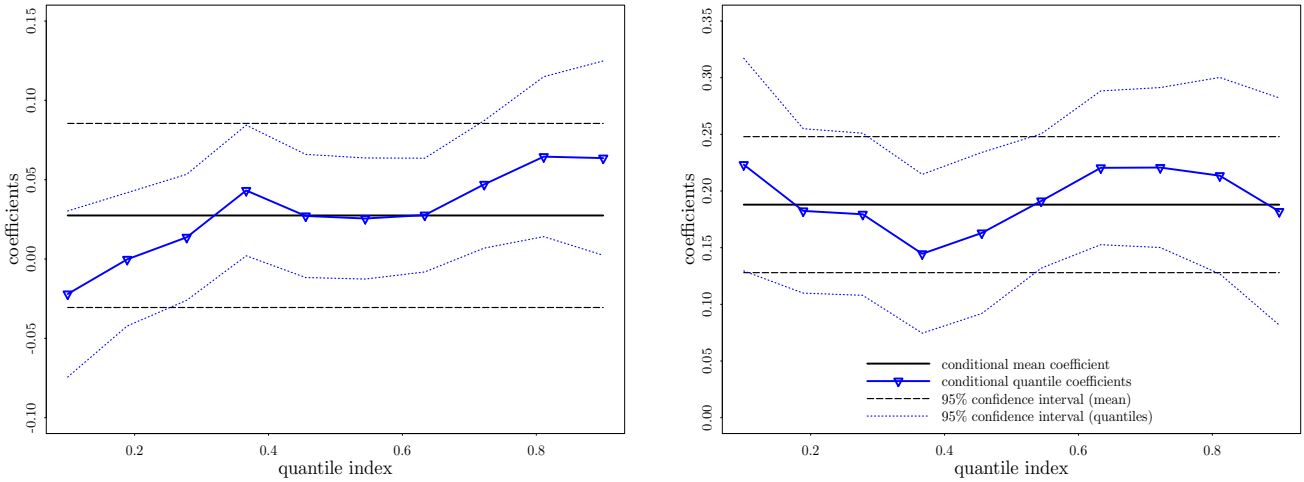


Figure C2: Conditional Mean and Conditional Quantile Coefficients (child nutritional status with respect to income)



(a) Baseline slope

(b) Slope change

Source: India Human Development Survey (IHDS)