NBER WORKING PAPER SERIES

ECONOMIC DEVELOPMENT, THE NUTRITION TRAP AND METABOLIC DISEASE

Nancy Luke
Kaivan Munshi
Anu Oommen
Swapnil Singh

Working Paper 29132
http://www.nber.org/papers/w29132

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2021

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Economic Development, the Nutrition Trap and Metabolic Disease
Nancy Luke, Kaivan Munshi, Anu Oommen, and Swapnil Singh
NBER Working Paper No. 29132
August 2021
JEL No. I15,O20

ABSTRACT

This research provides a single explanation for: (i) the persistence of malnutrition and (ii) the increased prevalence of metabolic disease (diabetes, hypertension, cardiovascular disease) among normal weight individuals with economic development. Our model is based on a set point for BMI or bodyweight that is adapted to conditions of scarcity 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 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.

Nancy Luke
Pennsylvania State University
702 Oswald Tower
University Park, PA 16802
nkl10@psu.edu

Anu Oommen
Christian Medical College
Vellore
India
anuoommen@cmcvellore.ac.in

Kaivan Munshi
Department of Economics
Yale University
37 Hillhouse Avenue
New Haven, CT 06511
and NBER
kaivan.munshi@yale.edu

Swapnil Singh
CEFER
Totoriu g. 4
Bank of Lithuania
Vilnius, LT-01121
Lithuania
ssingh@lb.lt

1 Introduction

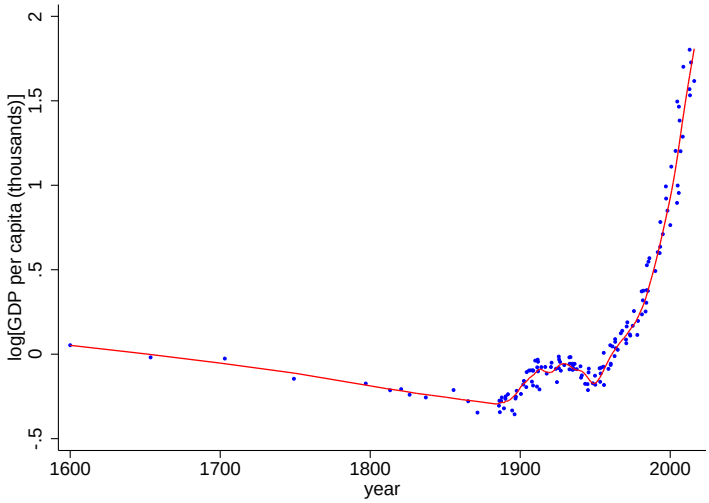
This research is motivated by two recently documented facts that run counter to the conventional wisdom that higher incomes lead to better health: First, the relatively weak relationship between nutritional status (BMI, height) and income in developing countries; both across countries (Deaton, 2007) and within country over time (Deaton and Drèze, 2009). Second, the increased prevalence of metabolic disease (diabetes, hypertension, cardiovascular disease) among *normal* weight individuals with economic development (Narayan, 2016, 2017).[1] Take India, for example, a country which has received much attention in the nutrition and health literatures. India has experienced substantial economic growth and sharp declines in the prevalence of poverty in recent decades. Nevertheless, a surprisingly large fraction of its population remains malnourished, while, simultaneously, the prevalence of metabolic diseases, including diabetes, has increased dramatically. 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, we will see below that a relatively small fraction of the Indian population is currently obese and that the risk of diabetes starts to increase at a BMI level that is well within the normal range.

The mathematical model that we propose to explain why malnutrition stubbornly persists, even as metabolic diseases emerge is based on the following verbal argument. The pre-modern economy was characterized by low and fluctuating food supply for centuries. There is theoretical support (discussed below) for the assumption that the population in this economy adapted biologically to the low-nutrition environment that it faced over many generations. With economic development, there is a substantial increase in wealth or permanent income, which we refer to henceforth as "income," and, with it, consumption. 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. It has been hypothesized 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 (Gluckman and Hanson, 2004; Narayan, 2016, 2017; Wells et al., 2016). We place additional structure on the mismatch and simultaneously explain the high rates of metabolic disease and malnutrition in these countries by tying the initial adaptation to a set point for body weight.

Experimental and non-experimental evidence in humans indicates that each individual is endowed with a set point for body weight or BMI, with metabolic and hormonal adjustments defending the set point against fluctuations in food supply. We postulate that the body will defend its inherited set point against the increases in consumption that accompany economic development, just as it responded to short-term fluctuations in food supply in the pre-modern economy. Once the mismatch between current and ancestral income, or consumption, crosses a threshold, however, the body will no longer be able to defend the set point; the individual's BMI will track more closely with current income, but because the metabolic load now exceeds the metabolic capacity, there will be in tandem an increased risk of metabolic disease (see

[1] We use the term "metabolic disease" to describe a group of related disorders. This should not be confused with "metabolic syndrome," which is associated with a cluster of conditions; e.g. high cholesterol, triglycerides, blood sugar, that are precursors to these disorders.

Figure 1: Evolution of Income in India



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

Wells et al. (2016) for a similar argument). During the process of development, the population will thus be partitioned into two distinct groups: (i) Individuals who remain at their set point, despite the increase in their consumption, are responsible (in part) for the weak relationship between nutritional status 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.

The pre-modern set point is not permanent. As discussed in the subsequent section, the available evidence indicates that the adaptation to pre-modern conditions is epigenetic (changing gene expression) and, hence, can be expected to persist for a finite number of generations. This might explain why European populations, which presumably underwent a similar transition more than a century ago, no longer display the same physiological traits. Although the friction that we incorporate in our analysis to explain the two stylized facts is biological, it thus has many features in common with economic models of institutional adaptation and persistence. For example, Munshi and Rosenzweig (2006) describe how caste-based networks, which emerged in response to labor market imperfections in the pre-modern economy, generated a dynamic inefficiency when they failed to adjust to subsequent structural change in the Indian economy. In the current analysis, the human body adapts to the environment in the pre-modern economy, which was stable for many centuries, but then fails to adjust to rapid economic development, resulting in the persistence of malnutrition and the emergence of metabolic disease.

Our inter-generational focus complements a voluminous intra-generational literature in economics, going back to Cunha and Heckman (2007, 2008) that examines how initial conditions determine subsequent investments over the lifecourse that, taken together, result in the formation of cognitive and non-cognitive skills. A parallel literature has studied how early-life circumstances determine (largely economic) outcomes in adulthood in developing countries (see Currie and Vogl (2013) for a survey). In our analysis, the initial condition or set point is determined by ancestral income in the pre-modern economy, which interacts

with the exogenous change in income that accompanies economic development to determine the outcomes of interest – nutritional status and metabolic disease – in the current generation. There is a pre-specified mapping from income to consumption in the model laid out below; i.e. households do not respond to their poor nutritional status or the risk of metabolic disease by changing their consumption. This simplifying assumption appears reasonable in this context because the biological relationships that underlie our analysis only emerge with economic development, still need to be scientifically investigated, and are certainly not common knowledge in developing-country populations.

If data on income, BMI, and metabolic disease were available for each dynasty (household) over many generations, going back to the pre-modern period, then we could test the proposed model 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 cross-sectional tests of the model 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, we can derive the following result at any point in time: (i) Although nutritional status 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.

We use nationally representative household data from the India Human Development Survey (IHDS) to test the cross-sectional implications of the model. Our main result is that the nutritional status-income relationship (separately for children and adults) and the metabolic disease-income relationship are precisely as predicted by the model.[2] 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 relationship between nutritional status and household income below the estimated threshold, which is located close to the median income level in the population, can explain (in part) the first stylized fact. The steep increase in the probability of metabolic disease with income above the same threshold, which corresponds to a BMI that is well within the normal range, helps explain the second stylized fact.

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. In line with the model, the results with the IFLS match what we obtain with Indian data. In contrast, there is a positive and continuous relationship between household income and nutritional status with the Ghanaian data (information on metabolic disease is not available in the GSPS). India and Indonesia are evidently at

[2]Nutritional status is measured by height-for-age for children and BMI for adults in the empirical analysis. Alternative measures, based on weight-for-age for children and height for adults, deliver similar results.

a stage of economic development where a substantial fraction of the population lies on either side of the threshold, resulting in the coexistence of malnutrition and metabolic disease. In contrast, the Ghanaian population appears to be largely at its pre-modern set point, which is why there is no discontinuity.

Although our model and the accompanying empirical tests provide a single explanation for the two stylized facts that motivate this research, we must still account for other independent determinants of nutritional status and metabolic disease. The estimating equations include a rich set of covariates, which control for the effect of son preference, food tastes, and the disease environment on these outcomes. In addition, we verify that two proximate determinants of nutritional status that are especially relevant in developing economies – nutrient intake and children's illness, including diarrhoea – do not exhibit the same discontinuous relationship with household income. More generally, we are unaware of any alternative non-biological explanation that would generate a slope discontinuity at the same income level, with both nutritional status and metabolic disease as outcomes.[3]

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 in the literature. (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. 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) Nutritional status is determined by ancestral income, which is associated with the set point, below the 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 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 nutritional status-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. 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, nutritional status and metabolic disease for a representative sample of households in rural Vellore district. The striking finding, obtained independently with IHDS, IFLS, and

[3] As discussed below, selective child mortality as emphasized by Deaton (2007) or poverty traps as in Dasgupta and Ray (1986); Galor and Zeira (1993); Banerjee and Newman (1993) cannot explain all the results that we obtain.

SICHS data is that pre-modern income determines nutritional status below the threshold, whereas current income determines nutritional status 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.

Having tested the predictions 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 in South Asia is lower than what would be predicted by GDP per capita, whereas the opposite is true for Africa. Moreover, 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, 2016, 2017). We show that these seemingly unrelated findings can be easily 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 malnutrition and the prevalence of metabolic disease. A comparison of counter-factual nutritional status and actual nutritional status (predicted by the estimated model) with IHDS data indicates that stunting among children and the fraction of underweight adults would have declined substantially, by 30% and 50% respectively, in the absence of a set point. To quantify the contribution of the set point to metabolic disease, we first show, based on the model, that the risk of disease will not respond to variation in BMI below a threshold, but will be increasing in BMI above the threshold. Estimates with IHDS data locate this threshold at a BMI that is 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 malnutrition and metabolic disease in other countries in the coming decades as they develop. Although the prognosis for the future thus seems bleak, there is a glimmer of hope. Recent studies indicate that intense and sustained nutrition supplementation through early childhood can permanently improve nutritional status (Ruel et al., 2008; Puentes et al., 2016) and reduce the risk of metabolic disease in adulthood (Ford et al., 2018). These findings imply, in our framework, that the early childhood intervention shifted the set point upward, suggesting a promising way forward that we return to in the concluding section.

2 Biological Foundations

This section briefly reviews the scientific literature on biological adaptation-persistence and a set point for body weight, which are the building blocks of our analysis. The available evidence, and the current gaps in the evidence, are also discussed.

2.1 Adaptation and Persistence

Developmental plasticity – the shaping of later life traits by early life environments – has been well documented in animals and humans (Lea et al., 2017). Two models of developmental plasticity have been proposed in the literature: (a) The *developmental constraints* model in which developing organisms in re-

source limited environments make tradeoffs to protect critical functions and improve survival in early life (Barker, 1995). In this model, the fetus adapts to the immediate availability of nutrition *in utero*. (b) The *predictive* model in which maternal cues *in utero* predict the adult environment and the organism evolves accordingly in anticipation of future conditions (Gluckman and Hanson, 2006). Theoretical modelling indicates that if the prediction is correct and confers fitness more than 50% of the time, then such anticipation is advantageous from an evolutionary perspective (Jablonka et al., 1995). This will be the case in a slowly fluctuating environment in which conditions in past generations are an accurate predictor of conditions in the current generation (Jablonka and Lamb, 1999; Lind and Spagopoulou, 2018).

The pre-modern economy was characterized by large short-term fluctuations in food supply across seasons and years, but remained essentially stagnant with growth rates close to zero for centuries (Galor and Weil, 2000). In this environment, predictive maternal cues, based on conditions over many past generations, would have provided better information about extrauterine conditions in a given generation than nutrition availability *in utero*. The pre-modern population would thus have been adapted to long-term conditions of low food supply, with the adaptation varying across space with fixed growing conditions (agricultural productivity). The exception to this argument would have been generations in which there was an acute environmental shock, such as a famine, or populations that faced chronic severe malnutrition. In those (special) cohorts or populations, the developmental constraints model would have applied.

The assumption in both 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 (Jablonka and Raz, 2009; Miska and Ferguson-Smith, 2016; Sales et al., 2017; Radford, 2018; Lind and Spagopoulou, 2018). In contrast with traditional genetic alterations, epigenetic changes persist for a limited number of generations. As noted, this allows us to explain why European populations, which were also under-nourished historically, no longer exhibit the traits we document in developing-country populations. The additional assumption in both models of developmental plasticity is that if there is a mismatch between the conditions to which individuals are adapted and the conditions they face in the extrauterine environment, then this will give rise to an increased risk of metabolic disease. This implies that a population that is adapted to the low nutrition environment in the pre-modern economy will be at increased risk of metabolic disease with economic development, potentially for multiple generations (Gluckman and Hanson, 2004). We explain the coexistence of metabolic disease and malnutrition in developing economies by tying the initial adaptation to a set point for body weight.

2.2 The Set Point

Set point theory was originally motivated by the observation that a typical individual's body weight is remarkably stable over time (Leibel, 2008). The basic assumption underlying the theory is that each individual inherits a set point for her bodyweight and that, given the physiological cost of weight-cycling (Brownell and Rodin, 1994) the body defends that set point against fluctuations in food supply by making metabolic and hormonal adjustments (Müller et al., 2010). In our model, the set point during the process of

development is determined by ancestral income (consumption) in the pre-modern period.[4] We postulate, in addition, that as long as current and pre-modern income remain sufficiently close to each other, the body will successfully defend its set point. Nutritional status will be determined by pre-modern rather than current income and the risk of metabolic disease will be low. 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. Nutritional status will now start to track current income and, in tandem, there will be an increased risk of metabolic disease (even among normal weight individuals) because the metabolic balance that maintained the set point has been disrupted.[5]

Although there is robust empirical support for the basic elements of set point theory, as discussed below, the theory has been criticized for its inability to explain the obesity epidemic in advanced economies (Speakman et al., 2011; Müller et al., 2018). In response to this criticism, set point models that apply more specifically to populations in these economies have been proposed: (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). The critique of the standard set point model does not apply to our analysis because (as seen below) a relatively small fraction of the population is obese in developing countries. The alternative models are nonetheless informative because they tell us what we might expect to observe in these countries when individuals escape the pre-modern set point. As described above, and in line with the alternative set point models, we assume that nutritional status for these individuals will now start to track with their (current) incomes.

2.3 The Evidence

The evidence for biological adaptation and persistence in humans is largely based on the developmental constraints model, or what is commonly referred to in the literature as the 'fetal origins' hypothesis (Hales and Barker, 1992; Barker, 1995). The robust finding from many studies that have tested this hypothesis is that a combination of low birth weight, generated accidentally by famine or some other adverse shock, and high adult BMI puts individuals at greatest risk of metabolic disease; e.g. Ravelli et al. (1998); Bhargava et al. (2004); Li et al. (2010). Providing support for epigenetic adaptation to the adverse initial conditions, the elevated risk of metabolic disease has been shown to persist for up to two generations among the descendants of individuals who experienced famines *in utero* (Lumey, 1992; Bygren et al., 2014; Li et al., 2017).

Statistical tests of the predictive model of developmental plasticity, which may be more germane to our analysis, are more difficult to implement because initial conditions are determined by fixed ancestral income, going back many generations, rather than by shocks in the current generation.[6] The best evidence to date in

[4] As noted, recent evidence from multiple studies indicates that intense nutrition supplementation throughout early childhood can shift the set point. The implicit assumption in our analysis, which we verify, is that the increase in nutrient intake that accompanies economic development is not sufficient to shift the set point in the same way.

[5] Wells et al. (2016) make the related argument that metabolic diseases increase with economic development (and increased consumption) because the metabolic load exceeds the inherited capacity.

[6] Both the developmental constraints model and the predictive model imply that a positive shock to income (and accompa-

support of long-term adaptation and subsequent multi-generational persistence comes from animal studies. For example, 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).[7]

In humans, the available evidence in support of adaptation-persistence is less direct and is based on the experience of migrants from developing countries residing in advanced economies, and their descendants. 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 the empirical evidence that migrants' nutritional status converges to the level of the native population very swiftly (Alacevich and Tarozzi, 2017). Nevertheless, their ancestral income should continue to determine the risk of metabolic disease, possibly for multiple generations, and inter-regional differences in historical income can thus be used to provide preliminary support for the predictive model. Immigrants from South Asia, a historically poor region, 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; Staimez et al., 2013; Kanaya et al., 2014). 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.[8]

The second building block of our analysis – the set point paradigm – is by now textbook material in the biological sciences. Providing support for the presence of a set point, experimental and non-experimental evidence, in animals and humans, indicates that when the system is perturbed in either direction through a change in diet, the body returns to its original weight once the nutritional constraint is released (Rothwell and Stock, 1979; Pasquet and Apfelbaum, 1994; Keesey and Hirvonen, 1997). Furthermore, energy expenditures are modulated to resist the perturbation, indicating that the body is actively defending its set point (Dulloo and Jacquet, 1998; Leibel, 2008).

Our analysis assumes, in addition, that the body will defend its inherited set point up to a threshold. Recent evidence on diabetes reversal through a weight loss program (Taylor and Holman, 2015) is consistent with this assumption; there is an individual-specific BMI threshold, which is independent of initial BMI, below which diabetes is reversed. However, the heterogeneity in the BMI threshold is not explained and similar experiments have not been conducted in developing countries (where we would expect the threshold to be determined by pre-modern income).

nying food consumption) later in life will increase the risk of metabolic disease. Providing support for this (shared) implication, Sekhri and Shastry (2019) exploit spatial variation in access to Green Revolution agricultural technology in rural India to document that a positive income shock associated with economic development is accompanied by an increased risk of metabolic disease.

[7] Although under-nourishment over many generations, followed by the restoration of normal food intake, was designed to mimic the development process, specific findings from this animal experiment (discussed below) are at odds with the experience in human populations. We will, nevertheless, extend the benchmark model to incorporate these findings, without altering its key implications.

[8] 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. Given the enormous differential between pre-modern income and current income for migrants to advanced economies, we might expect their set points to have shifted up. However, a model in which their set points are determined entirely by nutrition in early life would be unable to explain why the elevated risk of metabolic disease persists for multiple generations post-migration.

Summarizing the discussion in this section, there is broad theoretical and empirical support for biological adaptation-persistence and a set point for body weight. However, specific elements of our theoretical foundations remain to be verified in developing countries: (i) the assumption that the body defends its set point up to a threshold, and (ii) the assumption that the set point is determined by pre-modern (ancestral) income. The biological relationships that serve as the starting point for the model build directly on these assumptions and, hence, we will take particular care to validate these relationships.

3 The Model

3.1 Population and Income

The population consists of a large number of infinitely lived dynasties (families). Each dynasty consists of a single individual in each time period or generation, who is replaced by a single descendant in the period that follows. There is a fixed return on wealth in each period; 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 normalize so that the distribution of initial income is bounded below at zero, which corresponds to a subsistence level of consumption. We place no other restrictions on this distribution. We can think of the initial period as describing the pre-modern economy, while subsequent periods describe the process of development. Permanent income in an 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.

3.2 Biological Relationships

We now characterize the biological relationships between (i) nutritional status, measured by 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 arguments from the preceding section. We denote nutritional status in the model by BMI because the set point, which is a key ingredient in our analysis, is conventionally measured by bodyweight or BMI. However, we will experiment with alternative measures of nutritional status in the empirical analysis.

There is a positive and continuous relationship between consumption and income in all time periods. In addition, BMI is increasing continuously in consumption in the initial period. Dynasties with higher income

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

(and consumption) in the pre-modern economy will thus have a higher set point.[10] We specify the following relationship between initial BMI, z_0, or the set point, and initial income, y_0:

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

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 above, 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 the set point, z_0, determined in period 0, is assumed to be fixed across subsequent generations. 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.[11] Our model thus describes the relationship between nutritional status and income over a finite number of generations during the initial phase of economic development.

Notice also 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 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]

[10] In practice, epigenetic adaptation occurs over a long period of time. We can thus think of period 0 in the model as spanning multiple generations in the pre-modern era.

[11] In the multi-generational experiment with rats described above (Hardikar et al., 2015), altered epigenetic signatures after 50 generations of under-nutrition are partially (but not completely) reversed after normal nutrition is restored for two generations. We would thus expect these signatures to be erased after a sufficient number of generations.

[12] This is consistent with the conventional view that the regulation of body weight 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 body weight in a sample of rural Gambian women remained extremely stable (within 1.5 kg.) over a period of 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 generation. Such events have always been rare and are less relevant in the modern economy.

As long as consumption 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 consumption crosses the threshold, however, the metabolism can no longer maintain the energy balance and BMI starts to track with current income. As discussed in the preceding section, the accompanying mismatch between metabolic capacity and metabolic load simultaneously increases the risk of metabolic diseases. As in the developmental plasticity literature, this risk is specified to be increasing in the gap between current income, y_t, which determines current BMI (conditional on having crossed the threshold) and initial income, y_0, which determines the BMI set point. The relationship between the probability of metabolic disease, $P(D_t)$, and income can thus be characterized as follows:[14]

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

3.3 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 over time. 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 with current income. If data over many generations, going back to the pre-modern period, were available for each dynasty, then these predictions could be tested directly. In the absence of such multi-generational data, we proceed to derive the cross-sectional relationship between BMI and income, as implied by equation (3), when a dynasty-specific set point for body weight is present.

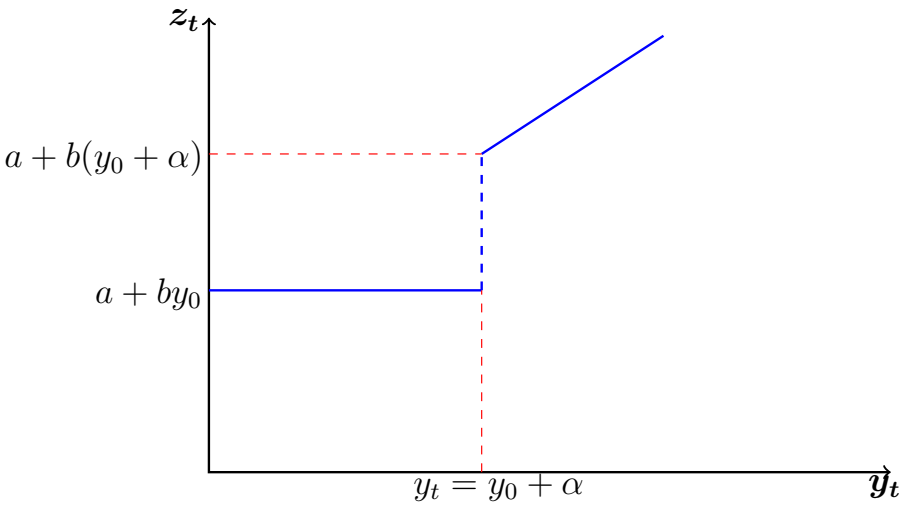
Recall that we normalize so that the initial income distribution is bounded below at zero. We also do not specify a lower threshold for the set point. It follows that all individuals with $y_t \leq \alpha$ must lie within their dynasty's set point threshold; some of these individuals will belong to dynasties 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. Given the assumed (normal) distribution of income shocks, mean BMI at any given level of income, y_t, for $y_t \leq \alpha$ is determined by the following expression:

$$\begin{aligned} \mathbb{E}(z_t | y_t, y_t \leqslant \alpha) &= \int_{-\infty}^{y_t} [a + b(y_t - U_t)] \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ &= a + b\left(y_t - e^L(y_t)\right), \quad e^L(y_t) = \frac{1}{\Phi(y_t; \mu_t, \sigma_t^2)} \int_{-\infty}^{y_t} U_t \phi(U_t; \mu_t, \sigma_t^2) \, \mathrm{d}\, U_t \end{aligned} \tag{5}$$

The implicit assumption when deriving the preceding expression is that y_0 is unbounded above, which is

[14] $\gamma_1 > 0$, $\gamma_2 > 0$ in equation (4). The implicit assumption, which is consistent with recent evidence discussed below, 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. This specification can also be compared and contrasted with Wells et al. (2016) who postulate that the risk of metabolic disease is increasing in the metabolic load and decreasing in metabolic capacity. They measure metabolic load by BMI and metabolic capacity by height. In our model, the risk of metabolic disease is also increasing in BMI (which is increasing in current income) but only to the right of the threshold. Moreover, metabolic capacity is determined by historical conditions, which can be measured by ancestral (but not current) height or, equivalently, by ancestral income.

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



why the range of integration extends to $-\infty$. Although the subsistence constraint in the pre-modern economy provides an obvious reason to bound the distribution below, there is no similar reason to impose an upper bound (although, in practice, extremely high values of y_0 would have very low probability). Nevertheless, we solve the model with an upper bound on y_0 in the Appendix. Although the results that we derive below cannot be obtained analytically, numerical solutions indicate that they continue to go through even when y_0 is bounded above.

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 within their thresholds. The expression for mean BMI at income level y_t, given that $y_t > \alpha$, thus includes both types of individuals,

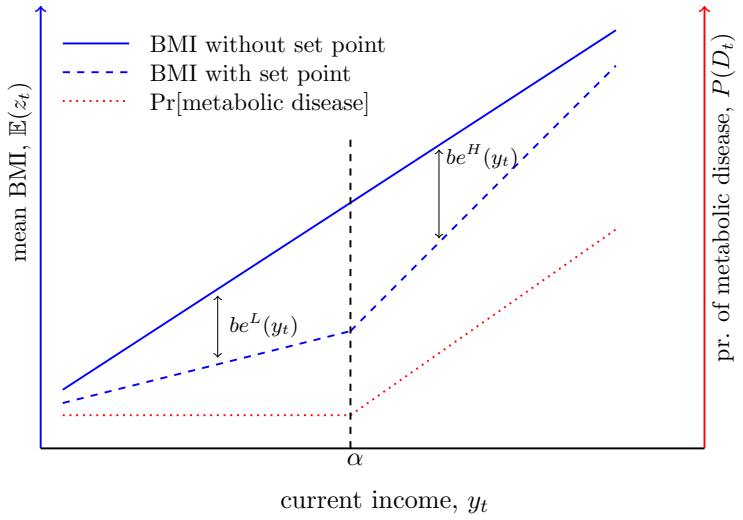
$$\begin{aligned}\mathbb{E}(z_t|y_t, y_t > \alpha) &= \int_{-\infty}^{\alpha} [a + b(y_t - U_t)] \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t + \int_{\alpha}^{y_t} [a + by_t] \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t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

As shown in the Appendix, 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:

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

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

Figure 3: Cross-Sectional Relationships



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 the Appendix):

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, when we derive the corresponding cross-sectional BMI-income relationship across dynasties at any point in time.

3.4 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 relationship 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 the Appendix) 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. Recall that

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: $y_t = \alpha$.[15]

3.5 Extensions to the Model

We close this section by considering alternative specifications of the biological relationships in the model. In our analysis, the population is adapted to long-term conditions in the pre-modern economy. Metabolic diseases emerge with economic development on account of the mismatch between current incomes (and consumption) and the conditions to which the population is adapted. While this reasoning is in line with the current paradigm in the developmental origins of adult disease literature, early contributions directly associated metabolic disease with adverse initial conditions, rather than the mismatch between initial and subsequent conditions. For example, Hales et al. (1991) and Barker (1995) report the unconditional negative correlation between the risk of metabolic disease and birth weight, without considering weight or BMI in adulthood. Hardikar et al. (2015) also find that under-nourished rats, and not just their better fed descendants, exhibit symptoms of metabolic disease, but it is worth noting that the treated rats were subjected to severe nutritional restrictions (less than half the food intake of control rats). In our view, research findings based on low birth-weight humans and severely under-nourished rats are informative on some dimensions, but they do not fully capture the development experience in humans. For example, the findings reported above are inconsistent with the observation that diabetes and related metabolic disorders are virtually nonexistent in human populations prior to economic development. We nevertheless allow for the possibility that initial conditions determine the risk of metabolic disease, independently of the mismatch, by augmenting equation (4): γ_1 is replaced by $\gamma_1 + \gamma_0 y_0$ in that equation, with $\gamma_0 < 0$, so that individuals with lower set points (y_0) are at greater risk, above and below the threshold. We cannot solve the model analytically when the disease-income relationship is augmented in this way, but numerical simulations reported in the Appendix indicate that Proposition 2 goes through unchanged, except that there will now be a negative slope below the threshold. Providing support for our preferred specification, empirical results reported below fail to detect an independent role for y_0 in determining the risk of metabolic disease.

We take a similar approach, guided by the received literature, as above, when considering alternative specifications of the BMI-income relationship (equation (3)). Following recent contributions to the set point literature, we assume that BMI tracks with current income once the individual escapes the nutrition trap. The implicit assumption is that BMI is now independent of ancestral income, y_0. In contrast, early studies in the developmental origins of adult disease literature; e.g. Barker et al. (2002) postulated that low birth weight was followed by subsequent compensatory growth. The Hardikar et al. (2015) experiment described above also finds that the descendants of under-nourished rats are heavier than control rats (once normal

[15] 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.

nutrition is restored). Although we have noted that evidence from low birth-weight humans or undernourished rats is not necessarily informative about the transition from the pre-modern to the developing economy, we nevertheless allow for compensatory growth by augmenting equation (3) above the threshold: $y_t = a+by_t+cy_t \cdot y_0$ if $U_t > \alpha$, where $c < 0$. This specification allows the BMI of individuals with low set points (y_0) to grow faster once they have escaped the nutrition trap. We cannot solve the model analytically with the augmented specification, but numerical simulations reported in the Appendix indicate that Proposition 1 goes through unchanged. While the robustness of our theoretical results to alternative specifications of the underlying biological relationships is reassuring, we note that tests of the model's internal validity (reported below) will provide support for the benchmark specification, as described by equation (3).

4 Cross-Sectional Analysis

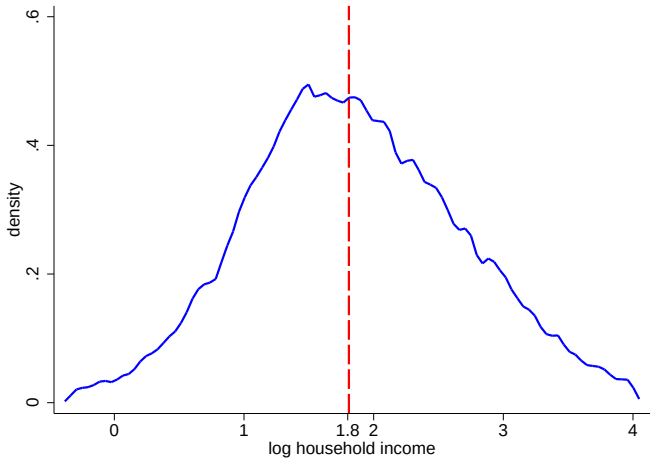
4.1 Descriptive Statistics

The key variables in the model are income, nutritional status, and the risk of metabolic disease. Although there is a single individual in each generation in our model, multiple individuals will reside in a household. Income will thus be measured at the household level. Nutritional status is measured for each (available) member of the household; by height-for-age for children and BMI for adults in the benchmark specifications. The model's implications for metabolic disease, in contrast, only apply to adults.

As discussed, the set point is typically associated with body weight or BMI, which is why we measure nutritional status (in later generations) by BMI in the model. However, biological adaptation to pre-modern conditions has been characterized more broadly in terms of a particular body type; e.g. (Narayan, 2016). When the body escapes the nutrition trap, this would imply that there is a change in both height and weight. Height is determined by nutrition in childhood, whereas BMI reflects food intake over the life-course. In the model, each generation (for a given dynasty) has a single level of income, in which case height and BMI could be used interchangeably to measure nutritional status. In practice, however, income (and consumption) will vary substantially within a generation in a developing economy. BMI may then be a better measure of adult nutritional status, which, in turn, is associated with the risk of metabolic disease. As a robustness test, we will nevertheless examine the cross-sectional implications of the model with height as an alternative measure of adult nutritional status; the additional advantage of this exercise is that it connects our analysis more directly to Deaton's (2007) analysis of the relationship between adult height and income. A similar motivation leads us to use height-for-age as the primary measure of age-specific nutritional status for children. While we verify that the results are robust to using weight-for-age to measure children's nutritional status, the core analysis uses height-for-age to link more directly to the literature on stunting; e.g. Jayachandran and Pande (2017).

The primary tests of the model are implemented with Indian data. This is because the rapidly developing Indian economy is simultaneously characterized by high levels of malnutrition and a high prevalence of metabolic disease; the two stylized facts that motivate our research. The core data set that we use for the analysis 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,

Figure 4: Household Income Distribution



Source: India Human Development Survey (IHDS)

nutritional status for children and adults residing in the household at the time of the survey, and the prevalence of metabolic diseases (diabetes, hypertension, and cardiovascular disease) 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.[16]

Figure 4 describes the distribution of household income in the IHDS data, measured as the log of monthly income in thousands of Rupees, averaged over the two survey rounds.[17] The vertical dashed line in Figure 4 denotes the median income, which is 1.8 in the nationally representative sample of households. Our tests of a slope-change, reported below, will locate an income threshold close to the median income, which tells us that it is not just the poorest who remain in the nutrition trap in this economy.

Figure 5a describes the nutritional status of children in the IHDS, separately for children aged 0-59 months and 5-19 years. Nutritional status, measured by the height-for-age, is reported as a z-score, based on child growth standards provided by the WHO.[18] We see that a substantial fraction of Indian children are stunted; with a z-score less than -2. Figure 5b describes the corresponding distribution of adult nutritional status, measured by the BMI.[19] The vertical dotted line in the figure denotes a BMI of 18.5, which is a cutoff

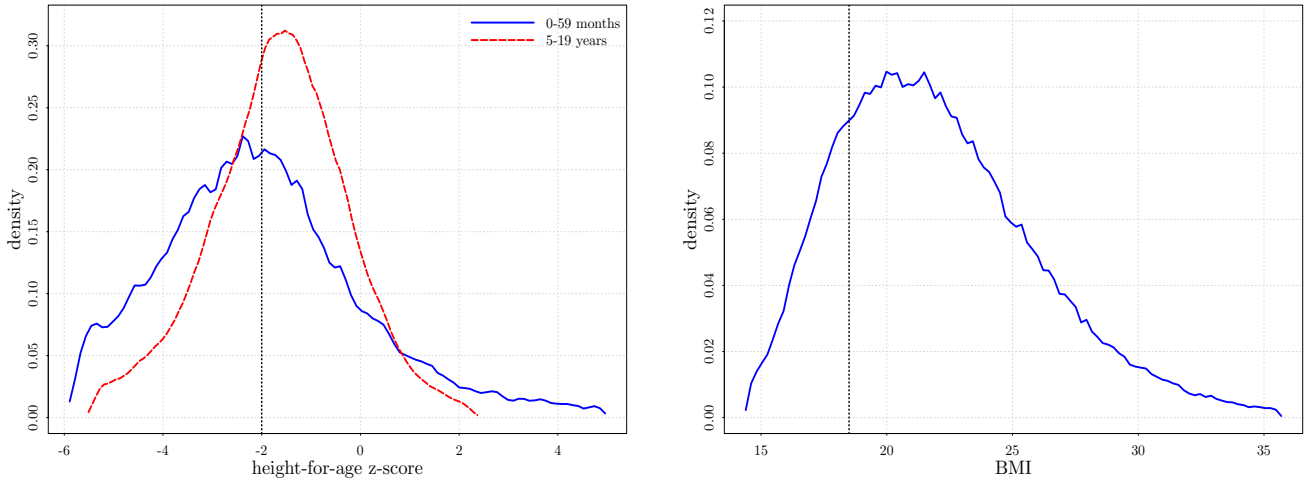
[16]The Demographic Health Survey (DHS), which is used by Deaton (2007) and Jayachandran and Pande (2017) also contains many of these variables. However, the DHS is not suitable for our purposes because it only collects indicators of asset ownership, which must then be converted into a crude wealth statistic 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.

[17]Household income 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.

[18]Given that the nutritional status measures are age-specific, information from both survey rounds is separately included for those children who appear in both rounds. The growth standard for children aged 0-59 months is based on the Multicentre Growth Reference Study (MGRS), conducted between 1997 and 2003. For children aged 5-19, we use the 2007 WHO Reference, which is a reconstruction of the 1977 National Center for Health Statistics (NCHS) growth standard. Following the recommendation of the WHO, height-for-age z-scores outside the (-6,6) interval are dropped from the analysis.

[19]The BMI is defined as the weight in kilograms divided by the square of the height in meters. Height and weight was

Figure 5: Nutritional Status Distribution - children and adults



(a) Children (b) Adults

Source: India Human Development Survey (IHDS)

conventionally associated with being underweight. We see that a substantial fraction of the Indian population remains below this cutoff, despite the economic progress of the past decades. By international standards, individuals are underweight if their BMI is below 18.5, the normal range is 18.5-25, the overweight range is 25-30, and obesity is defined by a BMI above 30. Based on this convention, most Indians are underweight or normal weight, and only a small fraction are obese. BMI that is too low or too high is physiologically damaging, but the latter is evidently less of a problem in India. We will see below that diabetes and related metabolic disorders, which are commonly associated with obesity in advanced economies, largely affect normal weight individuals in India.

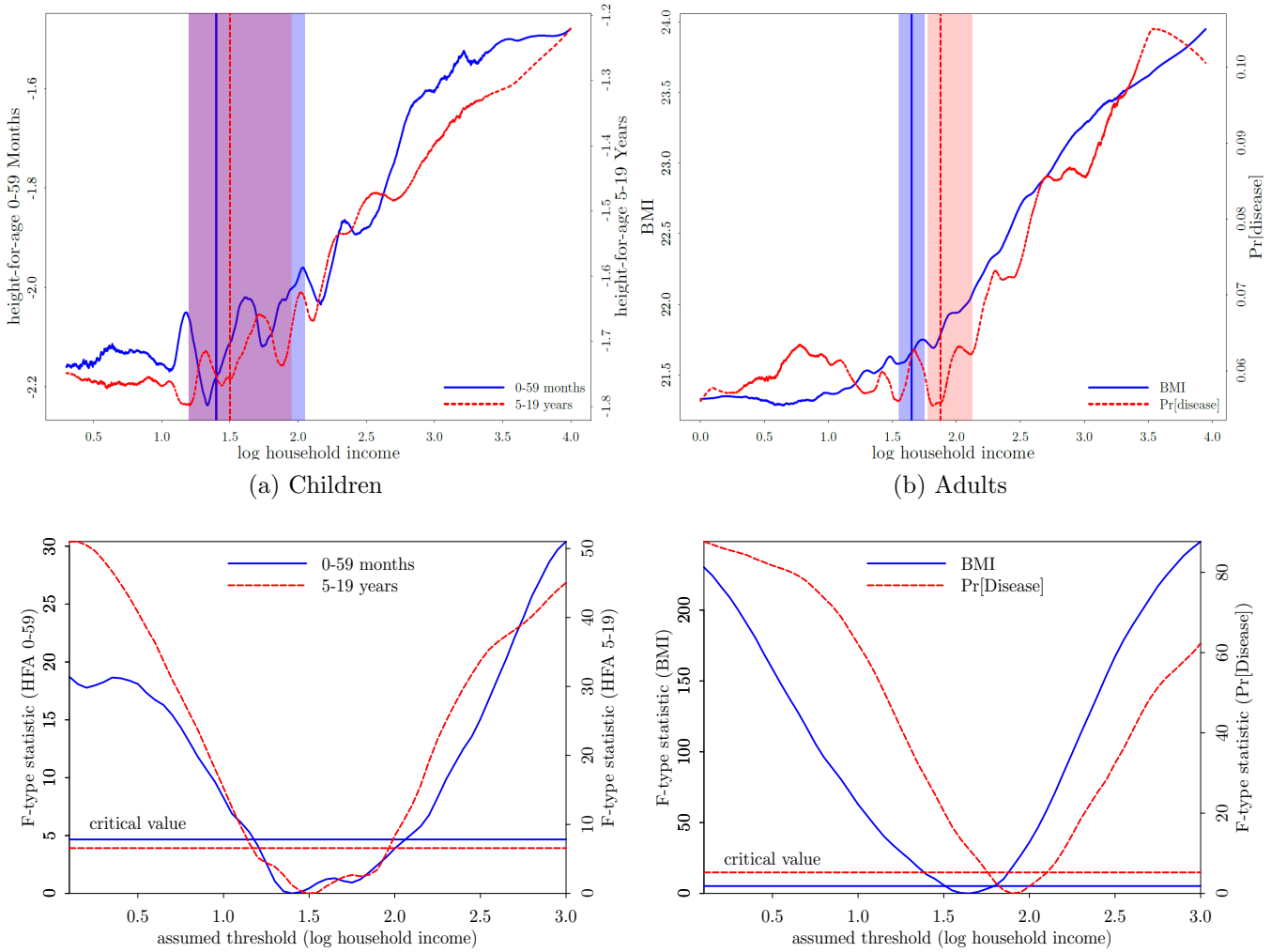
4.2 Cross-Sectional Tests

Proposition 1 derives the cross-sectional relationship between nutritional status 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 relationship 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 with data from the India Human Development Survey (IHDS) by separately estimating the relationship between income and both nutritional status and the probability of metabolic disease. Household income is measured as the average over the 2004 and 2012 survey rounds. 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 in the

measured for men and women in the 2011-2012 round, but only for a small fraction of men in the 2004-2005 round. As with the children, we include the BMI statistic separately from the two survey rounds when it is available for an adult.

Figure 6: 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: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies 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.

benchmark analysis is measured by BMI for the household head and his spouse and by height-for-age for their children, with individual information from both survey rounds included in the estimation sample when available. Metabolic disease is constructed as an individual-specific binary variable that indicates whether the household head and his spouse have been diagnosed with diabetes, hypertension, or cardiovascular disease.[20]

[20]Hypertension can be controlled by changes in lifestyle and recent evidence indicates that diabetes can be reversed with sufficient weight loss (Taylor, 2013). The model implicitly assumes that metabolic disease is reversible and we allow for this possibility in the empirical analysis by measuring metabolic disease in each survey round. As a robustness test, we also estimate the metabolic disease equation at a single point in time, with data from the final (2011-2012) survey round.

Figure 6a nonparametrically estimates the relationship between the nutritional status of the children and household income. Figure 6b repeats this exercise with nutritional status and the probability of metabolic disease among adult members of the household as outcomes.[21] Although our analysis focuses on the relationship with income, other individual and household characteristics could also 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: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies.[22] 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 Spears et al. (2013), Duh and Spears (2017), and 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 6a and 6b.

The vertical lines in Figure 6a and 6b 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 four outcomes, that the relationship with income is relatively weak below the estimated threshold, and much stronger above the threshold. The threshold location matches closely for the 0-59 month and the 5-19 year old children, with an almost complete overlap of the 95% confidence intervals. Despite the fact that we are using different measures of nutritional status for children and adults, the estimated threshold for the adults in Figure 6b, with BMI as the dependent variable in the estimating equation, is very close to what we obtain for the children in Figure 6a, with height-for-age as the dependent variable. Although the estimated threshold location with the probability of metabolic disease as the outcome is just slightly higher than the corresponding income level with adult BMI as the outcome, there is no overlap in the 95% confidence intervals. These confidence intervals are very precisely estimated, and we will see below that the threshold locations for adult BMI and the risk of metabolic disease are even closer to each other (and statistically indistinguishable) with alternative estimation samples from South India and Indonesia.

The threshold locations and confidence intervals in Figures 6a and 6b 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. nutritional status, y_i is household $i's$ income, τ is the location of the income 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. 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

[21] 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.

[22] Age is measured in years, except for the analysis with 0-59 month children where it is measured in months. The birth order is top coded at 3.

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

Dependent variable:	HFA 0-59	HFA 5-19	adult BMI	metabolic disease
	(1)	(2)	(3)	(4)
Baseline slope (β_1)	-0.049	0.024	0.239**	0.002
	(0.072)	(0.044)	(0.057)	(0.002)
Slope change (β_2)	0.365**	0.206**	0.940**	0.028**
	(0.073)	(0.045)	(0.066)	(0.003)
Threshold location (τ)	1.40	1.50	1.65	1.90
	[1.20, 2.05]	[1.20, 1.95]	[1.55, 1.75]	[1.80, 2.05]
Threshold test $p-$value	0.000	0.000	0.000	0.000
Mean of dependent variable	-1.991	-1.649	22.002	0.074
N	21,633	48,845	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: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies 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.

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.

Figures 6c and 6d plot the F-type statistic across the range of assumed thresholds for children's nutritional status 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 Figures 6a and 6b 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 6c and 6d 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

[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.

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 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 3) but not children (Columns 1-2), 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 4, in contrast with the strong positive relationship above the threshold.[24]

The estimated threshold location ranges from 1.4 to 1.9 for the four outcomes, with some amount of overlap in the confidence intervals between any pair of outcomes (the only exception is adult disease and nutritional status, as discussed above). Recall that the median income in our nationally representative sample of households is 1.8. Based on our model, all households with income to the left of the threshold remain in the nutrition trap, as do some households to the right of the threshold. This implies that a substantial fraction of the Indian population remains in the nutrition trap at this stage of economic development, with this group being partly responsible for the weak relationship between nutritional status and income that has been documented in the literature. Among the households to the right of the threshold, those that have escaped the nutrition trap are at elevated risk of metabolic disease. The micro-evidence we have provided can thus explain the co-existence of malnutrition and a high prevalence of diabetes and other metabolic disorders at this stage in India's economic development, as a consequence of underlying predetermined dynasty-specific set points in the population.[25]

We complete this section by verifying the robustness of this evidence in a number of ways: (i) We show in Appendix Table A1 that the results are robust to including period-specific income in place of average income (over the two survey rounds). (ii) We show in Appendix Table A2 that the results continue to be obtained when the outcomes are restricted to the 2011-2012 survey round. (iii) 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 Table A3.[26] While these variables could independently determine feeding practices, health seeking behavior, and other decisions that determine

[24]The number of observations in Column 4 is substantially greater than in Column 3 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]The positive relationship between health and income is a robust and well documented fact, within developed and developing countries (Cutler et al., 2006). In contrast, the positive relationship between the risk of metabolic disease and income that we document is specific to the development transition and we expect this relationship to be reversed in the long run (as observed in advanced economies).

[26]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.

nutritional status and health outcomes, we see that the results are robust to their inclusion. (iv) We show in Appendix Table A4 that the results for adults, with both nutritional status and metabolic disease as outcomes, are robust to separating men and women. (v) We show in Appendix Table A5 that the results continue to be obtained with alternative measures of nutritional status; weight-for-age for the children and height for adults. (vi) We show in Appendix Figure A1 and Appendix Table A6 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.

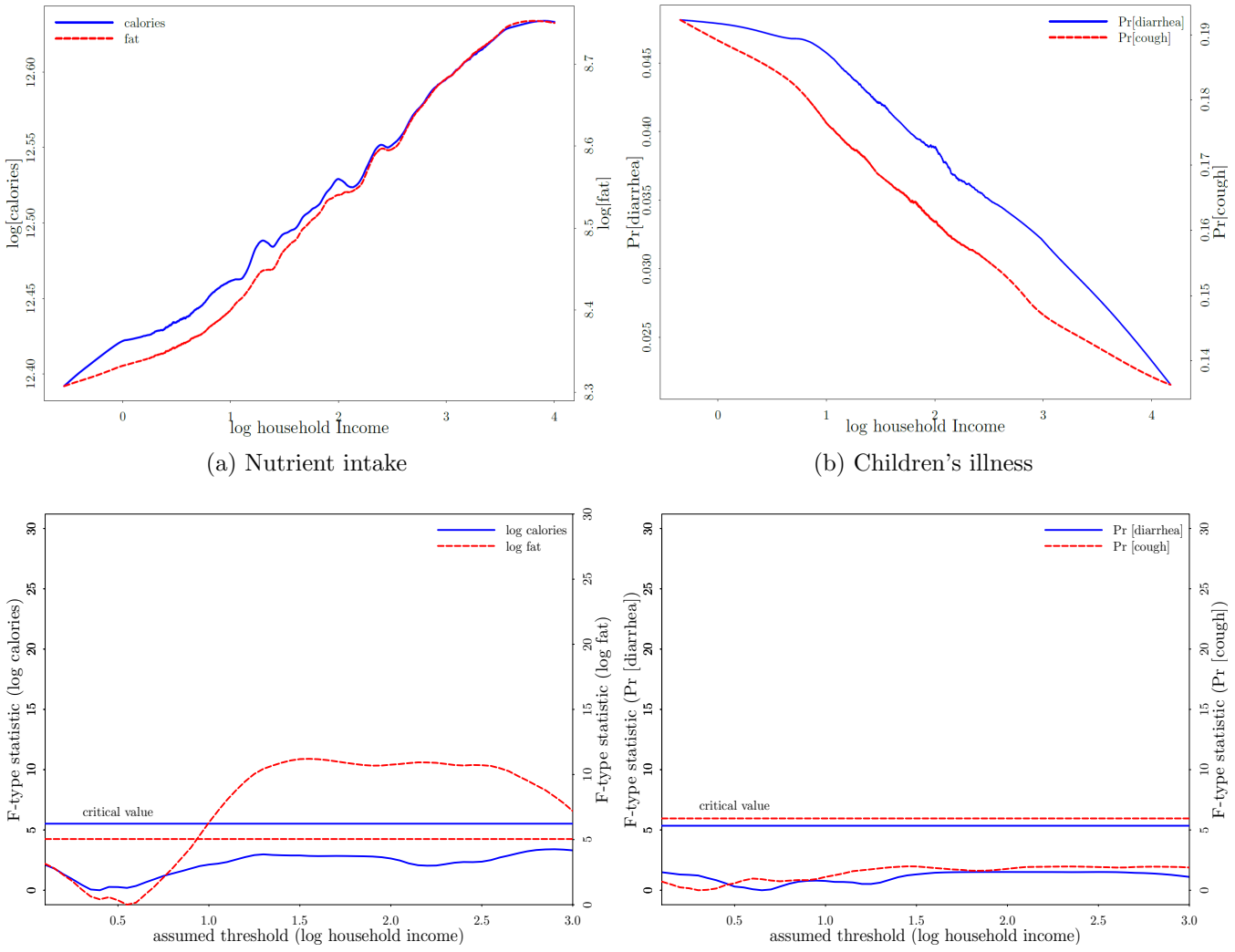
4.3 Alternative Explanations

The additional covariates that we include in the estimating equations account for two independent determinants of nutritional status in India: gender bias and a culturally determined preference for particular foods. The district dummies and the rural-urban dummy will also subsume spatial variation in the infectious disease environment and the availability of health services, which is especially relevant in a country such as India that is undergoing the epidemiological transition. However, such controls may not be complete. In the analysis that follows, we thus examine the possibility that important proximate determinants of nutritional status during the process of development – nutrient intake and children's illness, particularly diarrhoeal disease – vary with household income in a way that independently generates our results.[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. Suppose, instead, that the nutrient intake-income relationship strengthens discontinuously above an income threshold. Alternatively, suppose that there is a discontinuous change in the children's illness-income relationship. Either way, the nonlinear nutritional status-income relationship that we estimate could be obtained without a set point.

To assess the validity of these alternative explanations, we nonparametrically estimate the nutrient intake-household income relationship in Figure 7a and the children's illness-household income relationship in Figure 7b 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 (aged 0-19) 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. We see that there is a positive and continuous relationship between the intake of calories and fat and household income in Figure 7a. In contrast, there is a negative and continuous relationship between the incidence of both diarrhea and cough with household income in Figure 7b. In neither figure do we observe a discontinuous slope-change at any income level. Indeed, Hansen's test cannot place bounds on the threshold

[27] Deaton (2007) also considers energy expenditure (physical activity) as a determinant of nutritional status. 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. We will thus incorporate the type of work activity in the analysis that follows.

Figure 7: Nutrient Intake and Children's Illness with respect to Household Income



(a) Nutrient intake

(b) Children's illness

(c) Nutrient intake

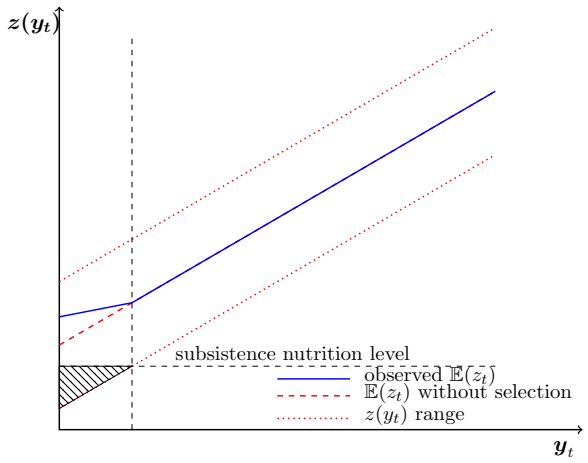
(d) Children's illness

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: dummies for the number of children, adults, and teens in the household, occupation dummies, caste group, rural-urban dummy, district dummies, survey year dummy, and reported local price of rice, wheat, cereals and their derivative products, pulses, meat, sugar, oil, eggs, milk and its derivative products, and vegetables.
For the children's illness figures, the standard set of covariates: gender, age (linear, quadratic, and cubic terms), birth order, caste group, rural-urban dummy, and district dummies 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.

location and, hence, fails to locate a slope-change at any assumed threshold in Figure 7c and Figure 7d.[28] The same result (not reported) is obtained with other measures of nutrient intake – sugar consumption –

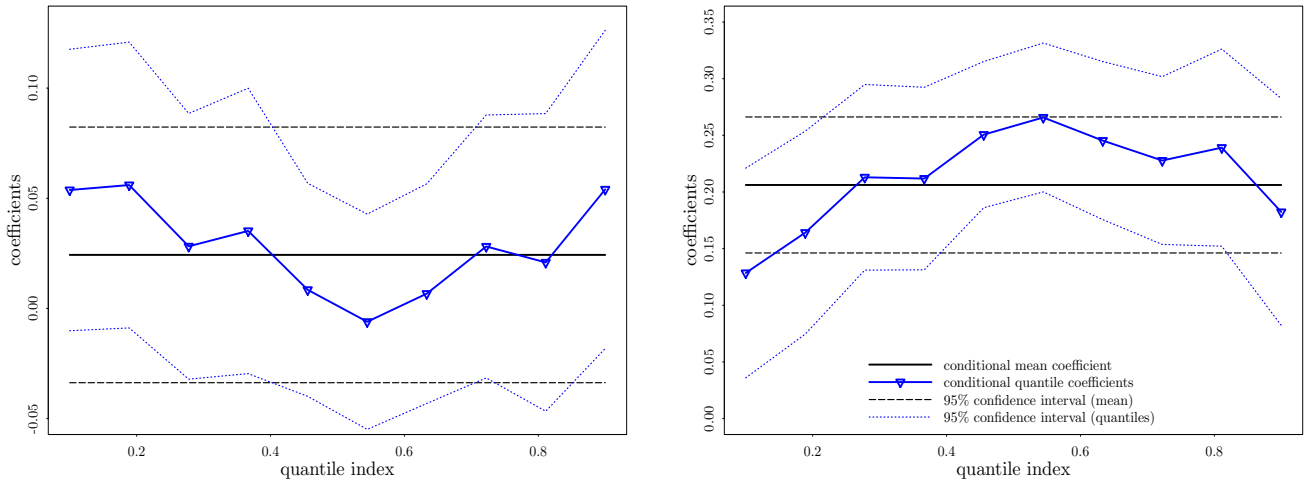
[28] We include all children aged 0-19 when estimating the children's illness-income relationship to increase the sample size and, hence, the likelihood of detecting a threshold. The corresponding tests (not reported) that separate the children into 0-5 year olds and 5-19 year olds also fail to detect 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.

Figure 8: Child Nutritional Status with respect to Income (with selective child mortality)



and children's illness – the incidence of fever. The discontinuous relationship between income and both nutritional status and metabolic disease is not being driven by an underlying discontinuous relationship between income and either food intake or children's disease (particularly diarrhoeal disease).

Figure 9: 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)

Although the proximate determinants of nutritional status do not vary discontinuously with household income, could the observed nonlinearity be generated by selective child mortality?[29] 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 8. If children can only survive above

[29]For example, Deaton (2007) considers the possibility that variation in child survival with income could explain the weak nutritional status-income relationship that he documents across countries.

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 9a and 9b 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) height-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 2. It is evident from Figures 9a and 9b 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.

Finally, could poverty trap models generate the results that we obtain? 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 malnutrition; 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. Indeed, Subramanian and Deaton (1996) go even further and argue, based on data from rural India, that there is no evidence that nutrition constrains income.[30]

Although we are unable to come up with an alternative explanation for the results that are obtained, some caveats are in order. First, we use coarse measures of nutrient intake – calories, fat, sugar – measured at the household level in our analysis. Food intake at the individual level is difficult to measure and recent evidence (Forouhi et al., 2014) indicates that nutrient-types must be measured at an extremely fine level to accurately predict the risk of diabetes. Second, while our nutritional status measures, based on weight and height, are directly measured, metabolic diseases (although diagnosed) are self-reported. It is possible that

[30]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. In contrast with the implications of the nutrition-based poverty trap model described above, 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 the model. Based on the nonparametric estimates for India in Figure 6 and the corresponding estimates for Indonesia that follow, this concern does not appear to be relevant in practice.

wealthier households are more likely to visit the doctor and, hence, to be diagnosed with these conditions. However, for such differential reporting, mis-measurement of food intake, or any other omitted variable to explain all of our results, it must explain the discontinuity in the relationship between household income and both nutritional status and the risk of metabolic disease, as well as the fact that the threshold is located at (or close to) the same income level for both outcomes. There is no obvious reason why this would be the case.

4.4 Internal Validity

As noted, 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 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). The analysis that follows thus validates the threshold assumption and the specific structure we have imposed on the threshold function in the nutritional status-income relationship.

Based on the assumed distribution of income shocks, equation (3) implies the following cross-sectional $z_t - y_t$ relationship, as specified in equations (5) and (6):

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

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

Closed-form 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 nutritional status-income relationship is correctly specified, 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

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

Dep. variable:	HFA 5-19		adult BMI	
Specification:	without adjustment (1)	with adjustment (2)	without adjustment (3)	with adjustment (4)
Slope below threshold (β_L)	0.011	0.132***	0.223***	0.735***
	(0.028)	(0.019)	(0.048)	(0.035)
Slope above threshold (β_H)	0.221***	0.166***	1.140***	0.797***
	(0.015)	(0.033)	(0.035)	(0.084)
$F-$statistic ($\beta_L = \beta_H$)	44.69	0.78	234.45	0.45
	[0.000]	[0.374]	[0.000]	[0.502]
Imposed threshold	1.50	1.50	1.65	1.65
N	48,846	48,846	76,949	76,949

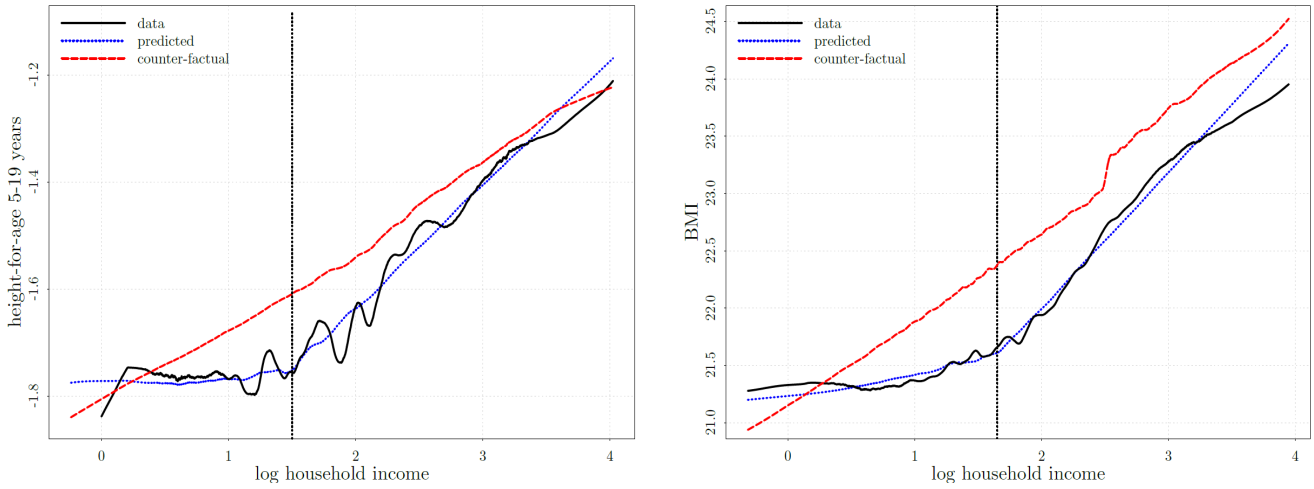
Source: India Human Development Survey (IHDS)
Logarithm of household income is the independent variable.
The standard set of covariates: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies are included in the estimating 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%

estimate the μ and σ parameters.[31]

Table 2 reports coefficient estimates from a piecewise linear equation, using IHDS all-India data, with child (aged 5-19) height-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

[31]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.

Figure 10: Predicted Nutritional Status and Counter-factual Simulations



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

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

than (greater than) b, below (above) the threshold, as implied by Proposition 1.

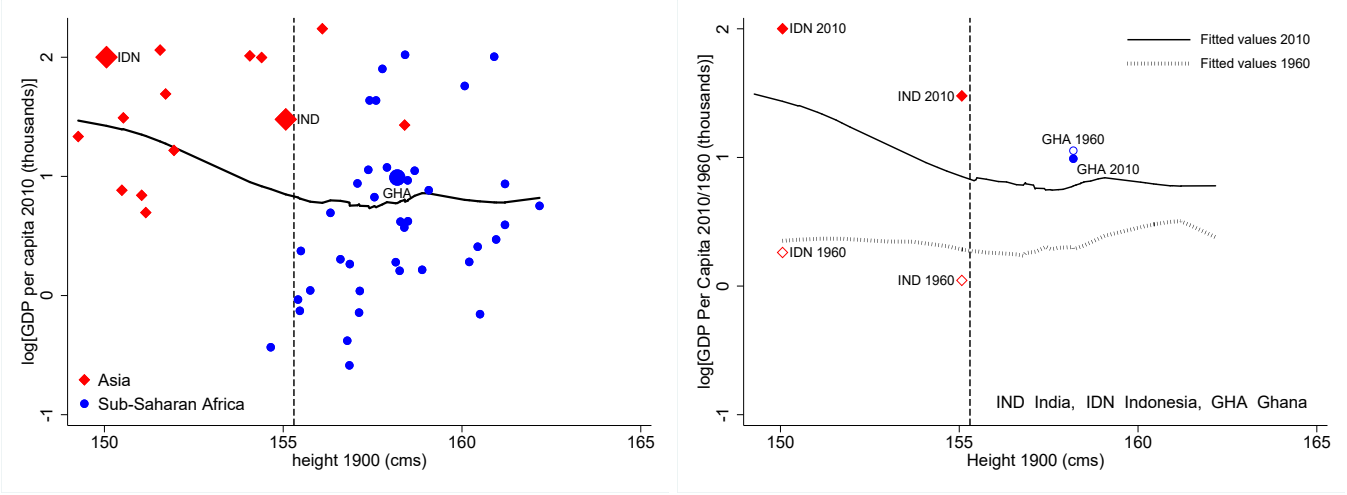
One benefit of the structural estimation is that it allows us to validate particular functional form assumptions in the model. An additional benefit is that it allows us to quantify the consequences of the nutrition trap. If the set point is irrelevant, there will be a linear relationship between household 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 10a reports actual height-for-age, predicted height-for-age (based on the model), and the counter-factual height-for-age (in the absence of the nutrition trap) for children aged 5-19. Figure 10b reports the corresponding relationships with adult BMI as the outcome. The standard set of covariates are partialled out, and the dotted vertical line in each figure marks the location of the income threshold. Based on these estimates, the fraction of stunted children (with a z-score below -2) would decline by 30% and the fraction of underweight adults (with a BMI below 18.5) would decline by 50% if the set point were absent.[32] 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, and we will return to this point in the concluding section of the paper.

4.5 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

[32]These statistics are based on a comparison of predicted and counter-factual malnutrition, taking account of the independent impact of the covariates.

Figure 11: 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 1900.

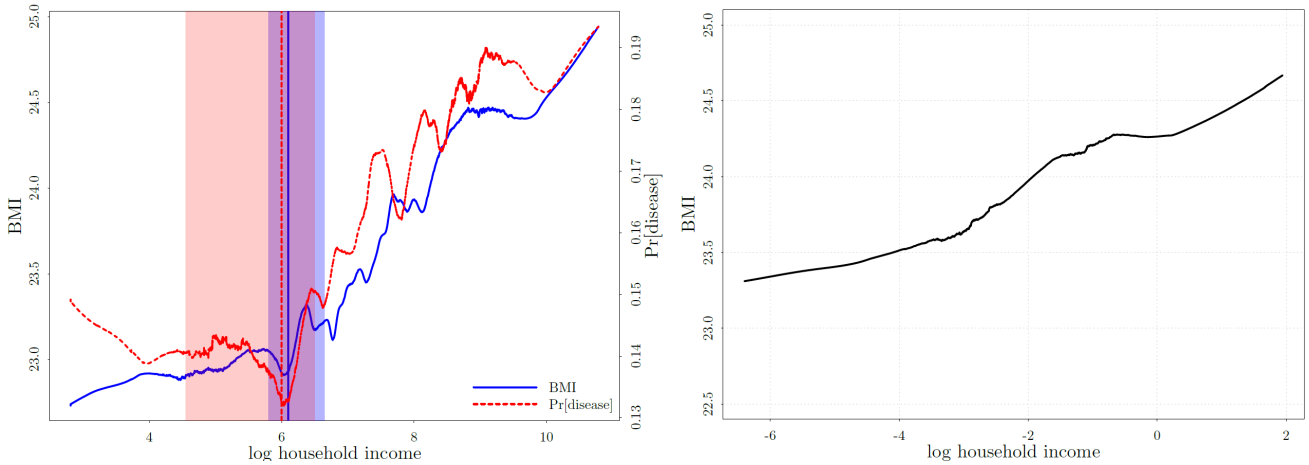
status. (iii) Indicators of metabolic disease. (iv) Household composition and detailed geographical indicators. The additional requirement is that a large sample is needed to locate a slope-change with precision. A search of publicly available data sets from other countries recovered two data sets that 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.[33] We thus proceed to test the model with these two data sets, just as we did with the IHDS for India.

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? Although income data only go back to 1960, adult height is available for many developing countries as far back as the nineteenth century. It is standard practice to use adult height as a proxy for income, and the standard of living, in historical

[33]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 is that it does not collect direct measures of income.

Figure 12: 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: gender, age (linear, quadratic, and cubic terms), ethnicity (Indonesia) or tribe (Ghana), rural-urban dummy, and district dummies are partialled out prior to nonparametric estimation.
The vertical line marks the threshold location and the shaded region demarcates the cluster bootstrapped confidence interval.

research (Steckel, 1995).[34] We thus use adult height in 1900 to measure historical income. Note that our use of historical height as a proxy for historical income does not contradict Deaton's (2007) observation that height and income are weakly correlated in developing countries today. Recall from the model that nutritional status, which can be measured by height, 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. Figure 11a plots the relationship between per capita GDP in 2010 and adult height in 1900 for a number of developing countries, including India, Indonesia, and Ghana.[35] The first point to take away from the figure is that there has been a reversal of 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 11b plots the relationship between per capita GDP, in 2010 and 1960, and height in 1900. 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 data were available a few decades prior to 1960. Even if we restrict attention to the 1960-2010 period, it is evident that the

[34] As noted by Steckel (1995) and Deaton (2007), genes are important determinants of individual height (and nutritional status more generally) but cannot explain variation across populations.

[35] 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.

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 12a 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 A2a 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 A7, there is a weak relationship 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 12b 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 12b. Formal statistical support for this observation is provided in Appendix Figure A2b, where the Hansen test is unable to detect an income threshold. As reported in Appendix Table A7, there is a positive and statistically significant relationship 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 fact 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.

5 The Mechanism

As noted, two assumptions that underlie the biological relationships specified in our model have not been previously verified in developing country populations: (i) the body will defend its inherited set point up to a threshold, and (ii) the set point is determined by conditions in the pre-modern economy, which we denote by ancestral income. We thus take particular care to validate these relationships: (a) Nutritional status 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. The test of the model's internal validity provides support for the particular structure, with a threshold, that we have imposed on the nutritional status-income relationship. We next proceed to independently validate the biological relationships described above by constructing exogenous measures of ancestral income. The threshold location for this exercise is derived from the cross-sectional tests of the model. 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 (with which a threshold can also be located).

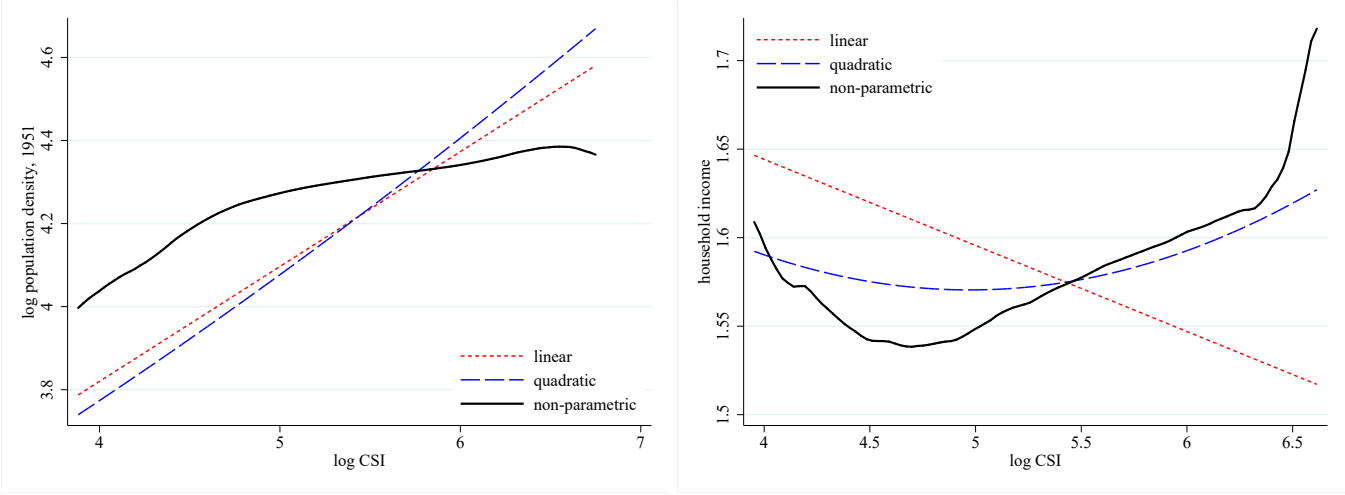
An appealing feature of the cross-sectional tests of the model is that they do not require knowledge of the set point, y_0. This allowed us to include rural populations and urban populations (which include a large share of relatively recent migrants) in the analysis. When testing the biological relationships, 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 these 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.

5.1 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 (including the impact of pests, disease, and weeds). They are provided for

Figure 13: Relationship between Population Density, Household Income and Caloric Suitability Index (CSI)



(a) Population density

(b) Household income

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

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 (Galor and Weil, 2000). If the CSI is a good measure of pre-modern aggregate wealth, then it should be closely related to historical population density. Figure 13a verifies this hypothesis by estimating the relationship between 1951 population density and CSI, at the district level, in India. 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. We see in Figure 13a that the CSI is positively associated with population density, although the relationship is nonlinear.

While the preceding analysis provides empirical support for our measure of historical aggregate wealth, it also indicates that the (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, $U_t \equiv y_t - y_0$.[36]

Figure 13b reports the relationship between current household income and CSI, as described by equation (10). In contrast with Figure 13a, this relationship is nonmonotonic, on account of the endogenous population response, and while the linear specification of the $f(CSI)$ function does a poor job of matching the data, the quadratic and nonparametric specifications track quite closely. 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, with the residual providing us with a measure of U_t.[37]

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 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 but not above it. Although the current income coefficient is also significant below the threshold, it is substantially smaller than the historical 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.[38]

[36]The residual 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. Equation (10) is at odds with the Malthusian model, which predicts that the endogenous population response would push per household ancestral income down to its subsistence level, regardless of aggregate wealth (CSI). It is, however, consistent with historical statistics reported in Galor and Weil (2000), which indicate that the population did not track one-for-one with total output in the pre-modern economy.

[37]We include state fixed effects when describing the relationship between CSI and both population density and household income, and also when estimating equation (10). These fixed effects are partialled out, using the Robinson procedure, in Figures 13a and 13b, and when $f(CSI)$ is specified as a nonparametric function in equation (10), in which case they are subsequently added back to estimate y_0. The state fixed effects will include a contemporaneous component, but this does not alter our interpretation of the predicted y_t as measuring historical conditions (y_0) because the state fixed effects are also included in the second-stage when we estimate the biological relationships.

[38]Equation (10) is estimated with regency fixed effects, which can also be included when estimating the biological relationships

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: gender, age (linear, quadratic, and cubic terms), caste group in India and ethnicity in Indonesia, state fixed effects in India and regency fixed effects in Indonesia 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.

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 benchmark specification of the disease-income relationship 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 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 benchmark specification of the metabolic disease -income relationship, we note 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

in Table 3 (as in the cross-sectional analysis) because CSI is measured at the sub-regency level.

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: gender, age (linear, quadratic, and cubic terms), caste group in India and ethnicity in Indonesia, district fixed effects in India and regency fixed effects in Indonesia 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.

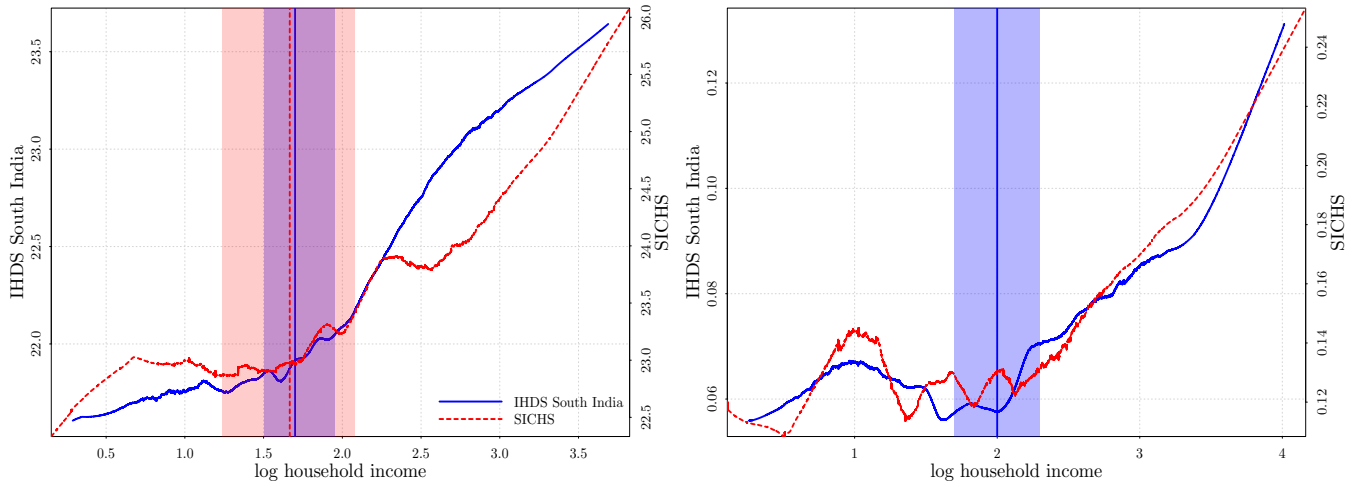
with ancestral income.

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, with adult BMI as the outcome in Appendix Table A8 and the risk of metabolic disease as the outcome in Appendix Table A9. (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 with adult BMI as the outcome in Appendix Table A10 and the risk of metabolic disease as the outcome in Appendix Table A11.

5.2 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

Figure 14: 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: gender, age (linear, quadratic, and cubic terms), caste group, and (for IHDS) rural-urban dummy and district dummies 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 95% confidence intervals.

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.[39] 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 14a reports the relationship 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

[39]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*.

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 demographic characteristics.[40] As a basis for comparison, we thus report the corresponding nonparametric plot obtained with IHDS data, for the South Indian states, in Figure 14a. 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 14b.

The estimated relationships, with SICHS and IHDS South India data, match very closely across the income distribution in both figures.[41] The vertical lines mark the spot where Hansen's test (shown in Appendix Figure A3) 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.[42] The tests of the mechanism underlying the model 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. 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 (Sales et al., 2017; Lind and Spagopoulou, 2018). 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.

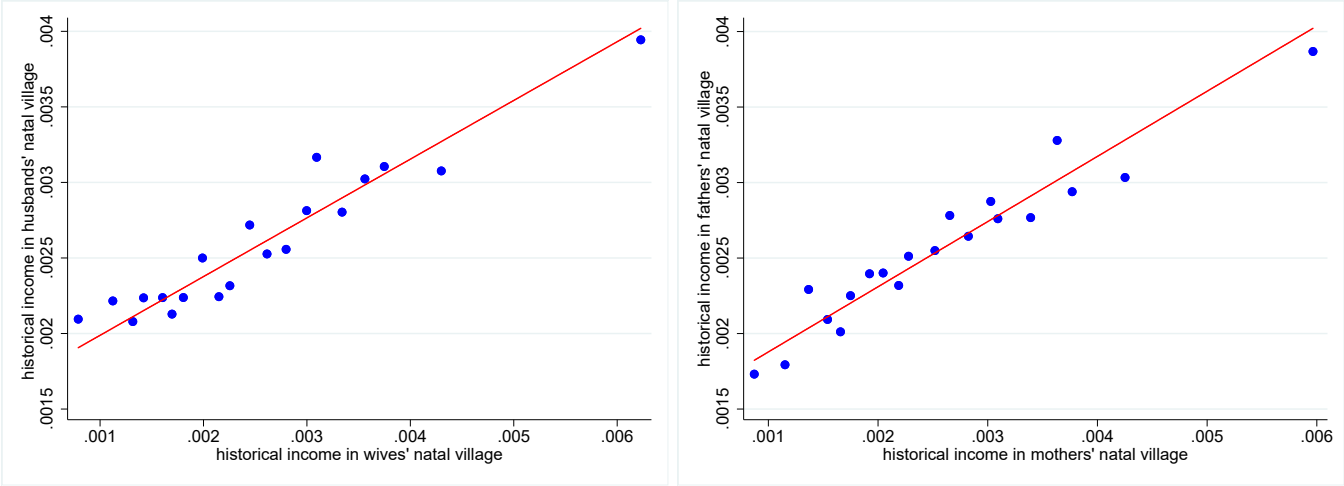
To construct a single measure of ancestral income, we take advantage of the fact that families in rural India match assortatively on wealth (permanent income) in the marriage market, as documented with SICHS data by Borker et al. (2018). Although ancestral income, y_0, will not match perfectly on the 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 Figure 15a and for their parents in Figure 15b, using the 1871 tax revenue in each

[40] Borker et al. (2018) 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, and labor force participation.

[41] Nutritional status 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 14). 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.

[42] 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 14, the piecewise linear equation estimates at the estimated thresholds are reported in Appendix Table A12. As with the Indonesian (IFLS) data, 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.

Figure 15: 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.

individual's natal village to measure y_0.[43] 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 14. 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.

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

[43]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: gender, age (linear, quadratic, and cubic terms), 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 primary sampling unit, are in parentheses.
$*$ significant at 10%, $**$ at 5%, $***$ at 1%, based on cluster bootstrapped confidence intervals.

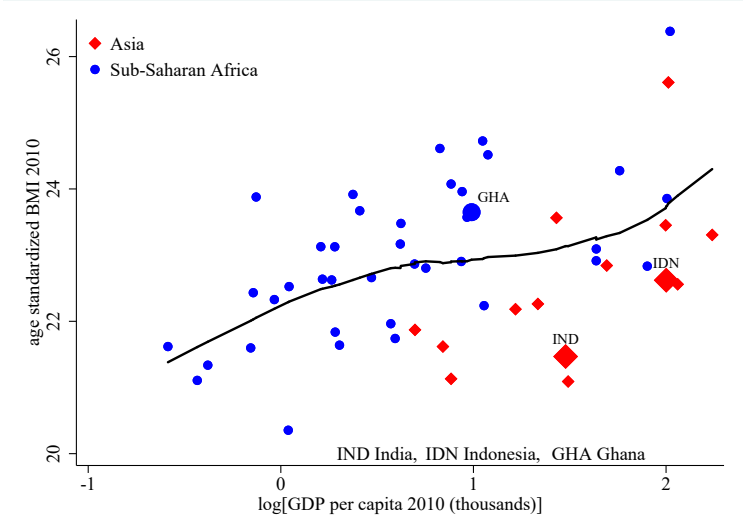
pre-modern conditions. Even if these conditions are exogenously determined, they could still be associated with factors that independently determine the outcomes of interest. For example, village-level tax revenue in 1871, which we use to construct ancestral income, is also associated with pre-modern 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 A13 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.[44]

6 Cross-Regional Implications

The nutrition-income puzzle that Deaton (2007) uncovered is that nutritional status in South Asia is lower than what would be predicted by GDP per capita, whereas the reverse is true for Africa. To explain this stylized fact through the lens of our model, consider a variant of the model that is adapted to a cross-country setting with aggregate data. We make the following assumptions: (i) A fixed fraction of the population, π,

[44]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.

Figure 16: Nutritional Status - Current Income Relationship Across Countries



Source: NCD-RisC and Penn World Table 9.0

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) Log income, $y_t^j \sim N(\mu_t^j, \sigma_t^2)$. (iii) Each dynasty in country j has the same income, y_0^j, in the initial period, 0. 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 the Appendix):

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

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 11, 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 hypothesis 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 hypothesis. Keeping this in mind, we plot average BMI against current GDP per capita in Figure 16. Drawing a vertical line through the figure at any level of current income, BMI is indeed higher in African countries than in Asian countries. The same result (not reported) would be obtained if we replaced adult BMI with the fraction of children that are (not) stunted or with adult height (the measure used by Deaton). Our model, based on a biological friction, is able to explain the well documented differences in nutritional status, conditional on income, between South Asia and Africa. Indeed, it can explain the wider difference between Africa and Asia, not just South Asia, as observed in Figure 16.

Although other mechanisms have been proposed to explain Deaton's puzzle, 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 during the process of economic development. 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 (Ramachandran, 2005; Narayan, 2016, 2017).

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

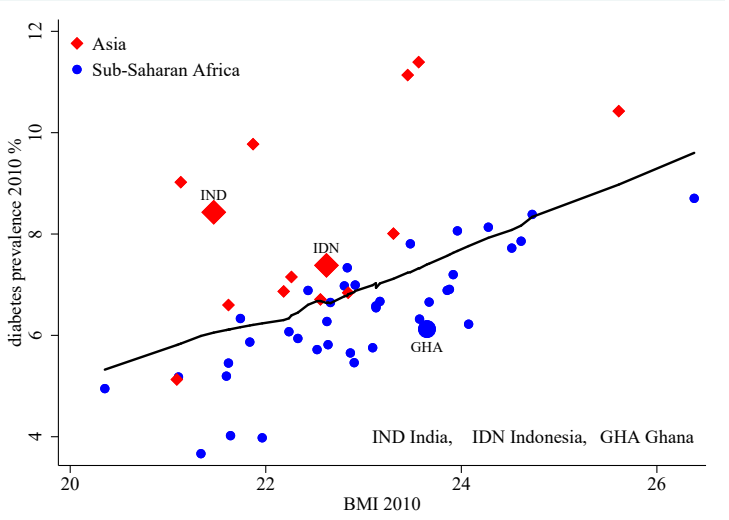
$$D_t^j = \gamma_1 + \gamma_2(1-\pi)\left[\mu_t^j + \sigma_t \frac{\phi\left[\Phi^{-1}(\pi;0,1);0,1\right]}{1-\pi} - y_0^j\right], \tag{12}$$

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 11 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 17 tests the preceding prediction by plotting diabetes prevalence against average BMI across countries. Drawing 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 nutritional status - income relationship, indicates that inter-regional differences in diabetes prevalence extend to the Asian continent as a whole.

Figure 17: Diabetes - BMI Relationship Across Countries.



Source: NCD-RisC

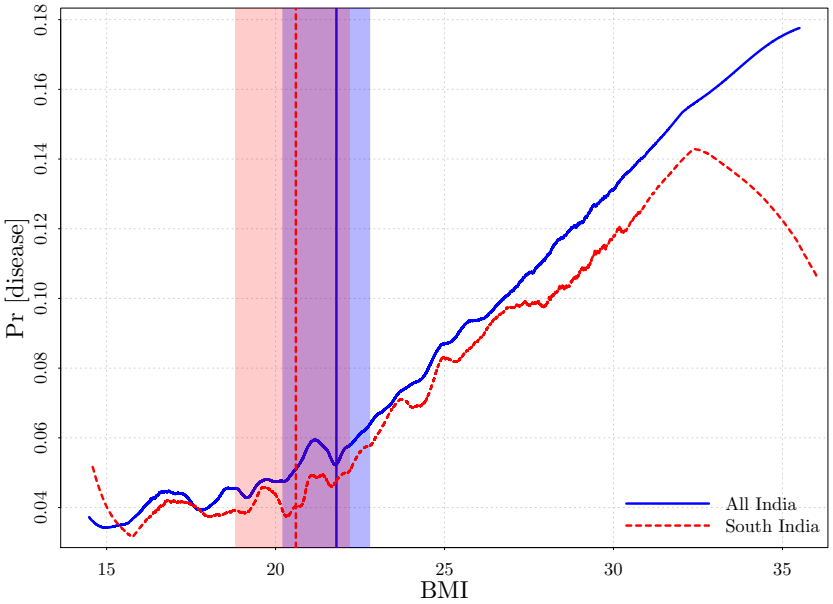
7 Conclusion

This paper provides a single explanation for two seemingly unrelated stylized facts: (i) the relatively weak relationship between nutritional status (BMI, height) and income in developing countries, and (ii) the increased prevalence of metabolic disease (diabetes, hypertension, cardiovascular disease) among normal weight individuals with economic development. Our explanation is based on a set point for body weight or BMI, which is adapted to economic 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 set point BMI, despite the increase in their consumption, and are (partly) responsible for the weak relationship between nutritional status and current income. Individuals in the second group, who have escaped the nutrition trap, but are not necessarily overweight, are the primary contributors to the increased risk 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 nutritional status-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, allowing us to simultaneously explain why nutritional status in Africa (Asia) is higher (lower) than what would be predicted by current GDP per capita, as well as why there is higher prevalence of diabetes, for given 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 stunting among 5-19 year olds would have declined by 30% and the fraction of underweight adults (with BMI below 18.5) in the population would have declined by

Figure 18: Metabolic Disease - BMI Relationship



Source: India Human Development Survey (IHDS)

50% in the absence of a set point. 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 among older children will also need to account for their potentially negative consequences for (future) chronic disease. A more promising approach, which has received empirical support (Ruel et al., 2008; Puentes et al., 2016; Ford et al., 2018), would be to invest in early childhood interventions. By shifting the set point, these interventions will have a larger effect on adult nutritional status, while simultaneously reducing the future risk of metabolic disease.

By heavily investing in early childhood nutrition programs, developing countries may be able to reduce the future burden of metabolic disease in younger cohorts. However, the prospects for older cohorts, who are locked into the pre-modern set point, are less promising. An increasing fraction of these cohorts will escape the nutrition trap in the coming decades, with an accompanying increase in the prevalence of metabolic disease. It is imperative that governments in developing countries take adequate steps to improve the prevention and treatment of these conditions. Screening will be an important component of these public health programs, 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 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. In our model, BMI is increasing with income at all levels, more steeply above a threshold, whereas the risk of metabolic disease is only increasing in income above the threshold. If income

heterogeneity is the source of forcing variation, then this implies that there will be no relationship between metabolic disease and BMI up to a threshold (which corresponds to an underlying income threshold) and a positive relationship thereafter. Figure 18 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. Our analysis indicates that the public health challenge faced by countries like India, which will need to successfully navigate the nutrition-disease tradeoff over the next couple of generations, may be even greater than what is currently envisaged.

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.

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.

Barker, D. J., J. G. Eriksson, T. Forsén, and C. Osmond (2002): "Fetal Origins of Adult Disease: Strength of Effects and Biological Basis," *International Journal of Epidemiology*, 31, 1235–1239.

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.

Bhargava, S. K., H. S. Sachdev, C. H. Fall, C. Osmond, R. Lakshmy, D. J. Barker, S. K. D. Biswas, S. Ramji, D. Prabhakaran, and K. S. Reddy (2004): "Relation of Serial Changes in Childhood Body-mass Index to Impaired Glucose Tolerance in Young Adulthood," *New England Journal of Medicine*, 350, 865–875.

Borker, G., J. Eeckhout, N. Luke, S. Minz, K. Munshi, and S. Swaminathan (2018): "Wealth, Marriage, and Sex Selection," Tech. rep., Working Paper, Cambridge University.

Brownell, K. D. and J. Rodin (1994): "Medical, metabolic, and psychological effects of weight cycling," *Archives of internal medicine*, 154, 1325–1330.

Bygren, L. O., P. Tinghög, J. Carstensen, S. Edvinsson, G. Kaati, M. E. Pembrey, and M. Sjöström (2014): "Change in paternal grandmothers early food supply influenced cardiovascular mortality of the female grandchildren," *BMC genetics*, 15, 12.

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

Cunha, F. and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47.

Cunha, F. and J. J. Heckman (2008): "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation," *Journal of Human Resources*, 43, 738–782.

Currie, J. and T. Vogl (2013): "Early-life health and adult circumstance in developing countries," *Annual Review of Economics*, 5, 1–36.

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

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.

Deaton, A. and J. Drèze (2009): "Food and Nutrition in India: Facts and Interpretations," *Economic and Political Weekly*, 42–65.

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

Duh, J. and D. Spears (2017): "Health and Hunger: Disease, Energy Needs, and the Indian Calorie Consumption Puzzle," *The Economic Journal*, 127, 2378–2409.

Dulloo, A. G. and J. Jacquet (1998): "Adaptive reduction in basal metabolic rate in response to food deprivation in humans: a role for feedback signals from fat stores," *The American journal of clinical nutrition*, 68, 599–606.

Ford, N. D., J. R. Behrman, J. F. Hoddinott, J. A. Maluccio, R. Martorell, M. Ramirez-Zea, and A. D. Stein (2018): "Exposure to improved nutrition from conception to age 2 years and adult cardiometabolic disease risk: a modelling study," *The Lancet Global Health*, 6, e875–e884.

Forouhi, N. G., A. Koulman, S. J. Sharp, F. Imamura, J. Kröger, M. B. Schulze, F. L. Crowe, J. M. Huerta, M. Guevara, J. W. Beulens, et al. (2014): "Differences in the Prospective Association Between Individual Plasma Phospholipid Saturated Fatty Acids and Incident Type 2 Diabetes: the EPIC-InterAct Case-Cohort Study," *The Lancet Diabetes & Endocrinology*, 2, 810–818.

Galor, O. and Ö. Özak (2016): "The agricultural origins of time preference," *American Economic Review*, 106, 3064–3103.

Galor, O. and D. N. Weil (2000): "Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond," *American economic review*, 90, 806–828.

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.

Hales, C. N. and D. J. Barker (1992): "Type 2 (Non-Insulin-Dependent) Diabetes Mellitus: the Thrifty Phenotype Hypothesis," *Diabetologia*, 35, 595–601.

Hales, C. N., D. J. Barker, P. M. Clark, L. J. Cox, C. Fall, C. Osmond, and P. Winter (1991): "Fetal and Infant Growth and Impaired Glucose Tolerance at Age 64." *BMJ: British Medical Journal*, 303, 1019–1022.

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 M. J. Lamb (1999): *Epigenetic inheritance and evolution: the Lamarckian dimension*, Oxford University Press.

Jablonka, E., B. Oborny, I. Molnar, E. Kisdi, J. Hofbauer, and T. Czaran (1995): "The adaptive advantage of phenotypic memory in changing environments," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 350, 133–141.

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.

Kanaya, A. M., D. Herrington, E. Vittinghoff, S. K. Ewing, K. Liu, M. J. Blaha, S. S. Dave, F. Qureshi, and N. R. Kandula (2014): "Understanding the High Prevalence of Diabetes in US South Asians Compared With Four Racial/Ethnic Groups: the MASALA and MESA Studies," *Diabetes Care*, 37, 1621–1628.

Keesey, R. E. and M. D. Hirvonen (1997): "Body weight set-points: determination and adjustment," *The Journal of nutrition*, 127, 1875S–1883S.

Lea, A. J., J. Tung, E. A. Archie, and S. C. Alberts (2017): "Developmental plasticity: bridging research in evolution and human health," *Evolution, medicine, and public health*, 2017, 162–175.

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

Li, J., S. Liu, S. Li, R. Feng, L. Na, X. Chu, X. Wu, Y. Niu, Z. Sun, T. Han, et al. (2017): "Prenatal exposure to famine and the development of hyperglycemia and type 2 diabetes in adulthood across consecutive generations: a population-based cohort study of families in Suihua, China," *The American journal of clinical nutrition*, 105, 221–227.

Li, Y., Y. He, L. Qi, V. W. Jaddoe, E. J. Feskens, X. Yang, G. Ma, and F. B. Hu (2010): "Exposure to the Chinese famine in early life and the risk of hyperglycemia and type 2 diabetes in adulthood," *Diabetes*, 59, 2400–2406.

Lind, M. I. and F. Spagopoulou (2018): "Evolutionary Consequences of Epigenetic Inheritance," *Heredity*, 121, 205–209.

Lumey, L. H. (1992): "Decreased Birthweights in Infants After Maternal in Utero Exposure to the Dutch Famine of 1944–1945," *Paediatric and Perinatal Epidemiology*, 6, 240–253.

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.

Miska, E. A. and A. C. Ferguson-Smith (2016): "Transgenerational inheritance: Models and mechanisms of non–DNA sequence–based inheritance," *Science*, 354, 59–63.

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.

Müller, M. J., C. Geisler, S. B. Heymsfield, and A. Bosy-Westphal (2018): "Recent advances in understanding body weight homeostasis in humans," *F1000Research*, 7.

Munshi, K. and M. Rosenzweig (2006): "Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy," *American Economic Review*, 96, 1225–1252.

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

Narayan, K. V. (2016): "Type 2 Diabetes: Why We Are Winning the Battle But Losing the War? 2015 Kelly West Award Lecture," *Diabetes Care*, 39, 653–663.

——— (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.

Pasquet, P. and M. Apfelbaum (1994): "Recovery of initial body weight and composition after long-term massive overfeeding in men," *The American journal of clinical nutrition*, 60, 861–863.

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.

Puentes, E., F. Wang, J. R. Behrman, F. Cunha, J. Hoddinott, J. A. Maluccio, L. S. Adair, J. B. Borja, R. Martorell, and A. D. Stein (2016): "Early life height and weight production functions with endogenous energy and protein inputs," *Economics and Human Biology*, 22, 65–81.

Radford, E. J. (2018): "Exploring the Extent and Scope of Epigenetic Inheritance," *Nature Reviews Endocrinology*, 14, 345–355.

Ramachandran, A. (2005): "Epidemiology of Diabetes in India–Three Decades of Research," *JAPi*, 53, 34–38.

Ravelli, A. C., J. H. van der Meulen, R. Michels, C. Osmond, D. J. Barker, C. Hales, and O. P. Bleker (1998): "Glucose Tolerance in Adults After Prenatal Exposure to Famine," *The Lancet*, 351, 173–177.

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.

Rothwell, N. J. and M. J. Stock (1979): "A role for brown adipose tissue in diet-induced thermogenesis," *Nature*, 281, 31–35.

Ruel, M. T., P. Menon, J.-P. Habicht, C. Loechl, G. Bergeron, G. Pelto, M. Arimond, J. Maluccio, L. Michaud, and B. Hankebo (2008): "Age-based preventive targeting of food assistance and behaviour change and communication for reduction of childhood undernutrition in Haiti: a cluster randomised trial," *The Lancet*, 371, 588–595.

Sales, V. M., A. C. Ferguson-Smith, and M.-E. Patti (2017): "Epigenetic Mechanisms of Transmission of Metabolic Disease Across Generations," *Cell Metabolism*, 25, 559–571.

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.

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.

Spears, D., A. Ghosh, and O. Cumming (2013): "Open Defecation and Childhood Stunting in India: an Ecological Analysis of New Data From 112 Districts," *PloS One*, 8, e73784.

Staimez, L. R., M. B. Weber, K. Narayan, and R. Oza-Frank (2013): "A Systematic Review of Overweight, Obesity, and Type 2 Diabetes Among Asian American Subgroups," *Current Diabetes Reviews*, 9, 312–331.

Steckel, R. H. (1995): "Stature and the Standard of Living," *Journal of Economic Literature*, 33, 1903–1940.

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.

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 Appendix Figures and Tables

Table A1: Piecewise Linear Equation Estimates (period-specific income)

Dependent Variable:	HFA 0-59 (1)	HFA 5-19 (2)	adult BMI (3)	metabolic disease (4)
Baseline slope	-0.032	-0.009	0.183**	0.0021
	(0.060)	(0.039)	(0.034)	(0.0014)
Slope change (β_2)	0.334**	0.179**	0.856**	0.036**
	(0.062)	(0.038)	(0.052)	(0.004)
Threshold location (τ)	1.30	1.40	1.80	2.20
	[1.00, 1.85]	[1.20, 1.70]	[1.65, 1.85]	[2.05, 2.35]
Threshold test $p-$value	0.000	0.000	0.000	0.000
Mean of dependent variable	-1.991	-1.652	21.996	0.074
N	21,533	46,538	76,189	147,480

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, measured in each survey year, is the independent variable.
The standard set of covariates: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies 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.

Table A2: Piecewise Linear Equation Estimates (outcomes restricted to IHDS 2011-2012)

Dependent variable:	HFA 0-59 (1)	HFA 5-19 (2)	adult BMI (3)	metabolic disease (4)
Baseline slope (β_1)	-0.071	0.045	0.294**	-0.001
	(0.117)	(0.049)	(0.068)	(0.003)
Slope change (β_2)	0.345**	0.188**	0.862**	0.036**
	(0.115)	(0.048)	(0.078)	(0.005)
Threshold location (τ)	1.30	1.55	1.60	1.90
	[0.70, 1.95]	[1.20, 2.10]	[1.50, 1.80]	[1.65, 2.05]
Threshold test $p-$value	0.000	0.000	0.000	0.000
Mean of dependent variable	-1.900	-1.578	22.189	0.098
N	10,362	35,764	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: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies 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.

Table A3: Piecewise Linear Equation Estimates (adult education and household composition included as additional covariates)

Dependent Variable:	HFA 0-59	HFA 5-19	adult BMI	metabolic disease
	(1)	(2)	(3)	(4)
Baseline slope (β_1)	-0.022	0.039	0.281**	0.003
	(0.079)	(0.073)	(0.052)	(0.002)
Slope change (β_2)	0.264**	0.130**	0.516**	0.014**
	(0.080)	(0.071)	(0.069)	(0.003)
Threshold location (τ)	1.40	1.50	1.80	1.95
	[0.95, 2.25]	[1.00, 2.10]	[1.60, 1.95]	[1.75, 2.30]
Threshold test $p-$value	0.002	0.000	0.000	0.000
Mean of dependent variable	-1.991	-1.649	22.002	0.074
N	21,634	48,845	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: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies 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.

Table A4: Piecewise Linear Equation Estimates (nutritional status and metabolic disease, 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

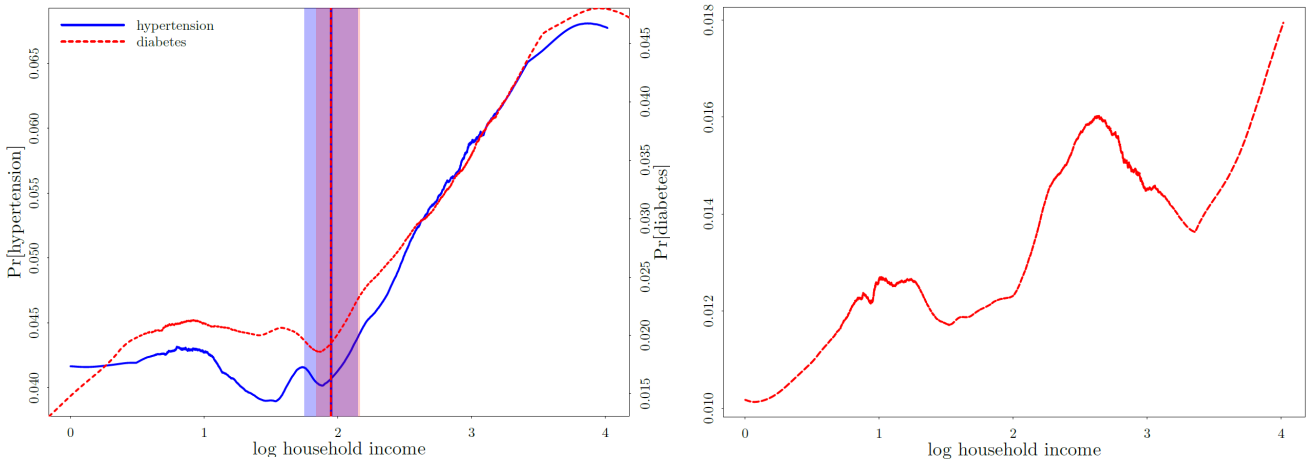
Source: India Human Development Survey (IHDS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
Logarithm of household income is the independent variable.
The standard set of covariates: gender, age (linear, quadratic, and cubic terms), caste group, rural-urban dummy, and district dummies 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.

Table A5: Piecewise Linear Equation Estimates (alternative nutritional status measures)

Dependent variable:	WFA 0-59 (1)	WFA 5-19 (2)	adult Height (3)
Baseline slope (β_1)	-0.053	-0.004	0.191
	(0.059)	(0.033)	(0.135)
Slope change (β_2)	0.287**	0.331**	0.836**
	(0.059)	(0.041)	(0.144)
Threshold location (τ)	1.30	1.75	1.45
	[1.05, 2.15]	[1.50, 2.00]	[1.30, 1.70]
Threshold test $p-$value	0.000	0.000	0.000
Mean of dependent variable	-1.512	-1.634	154.483
N	24,843	23,030	77000

Source: India Human Development Survey (IHDS)
Logarithm of household income is the independent variable.
The standard set of covariates: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste group, rural-urban dummy, and district dummies 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 A1: Metabolic diseases (separately) with respect to income



(a) Hypertension/Diabetes

(b) Cardiovascular disease

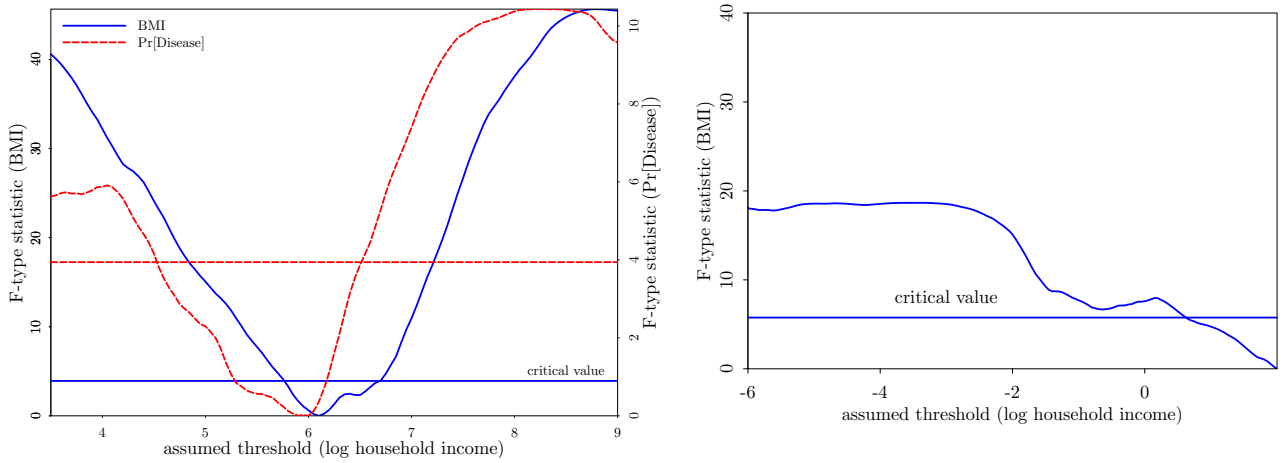
Source: India Human Development Survey (IHDS)
The standard set of covariates: gender, age (linear, quadratic, and cubic terms), caste group, rural-urban dummy, and district dummies are partialled out prior to nonparametric estimation.
The vertical lines mark the estimated threshold locations and the shaded regions demarcate the corresponding cluster bootstrapped 95% confidence intervals.

Table A6: Piecewise Linear Equation Estimates (hypertension and diabetes)

Dependent variable:	Hypertension	Diabetes
	(1)	(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: gender, age (linear, quadratic, and cubic terms), caste group, rural-urban dummy, and district dummies are included in the estimating equation and for the threshold test.
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 A2: 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: gender, age (linear, quadratic, and cubic terms), ethnicity (Indonesia) or tribe (Ghana), rural-urban dummy, and district dummies 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 A7: 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: gender, age (linear, quadratic, and cubic terms), ethnicity (Indonesia) or tribe (Ghana), rural-urban dummy, and district dummies 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.

Table A8: 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: gender, age (linear, quadratic, and cubic terms), caste group in India and ethnicity in Indonesia, state fixed effects in India and regency fixed effects in Indonesia 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 A9: 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: gender, age (linear, quadratic, and cubic terms), caste group in India and ethnicity in Indonesia, state fixed effects in India and regency fixed effects in Indonesia 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.
Joint $F-$statistic measures the joint significance of coefficients associated with Measure and Measure $\times 1\{income > \gamma\}$. $p-$values associated with joint $F-$statistics are in square brackets.
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 A10: 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

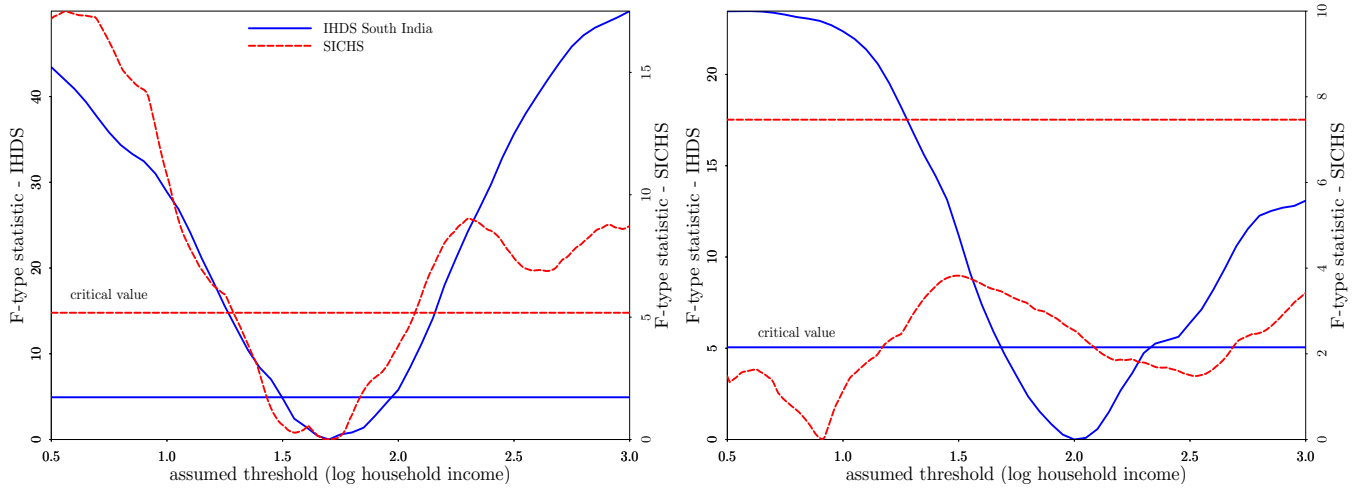
Source: India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: gender, age (linear, quadratic, and cubic terms), caste group in India and ethnicity in Indonesia, state fixed effects in India and regency fixed effects in Indonesia are included in the estimating 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 A11: 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: gender, age (linear, quadratic, and cubic terms), caste group in India and ethnicity in Indonesia, state fixed effects in India and regency fixed effects in Indonesia 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.
Joint $F-$statistic measures the joint significance of coefficients associated with Measure and Measure $\times 1\{income > \gamma\}$. $p-$values associated with joint $F-$statistics are in square brackets.
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 A3: 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: gender, age (linear, quadratic, and cubic terms), caste group and (for IHDS) rural-urban dummy and district dummies 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 A12: 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: gender, age (linear, quadratic, and cubic terms), caste group and (for IHDS) rural-urban dummy and district dummies 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

Table A13: Nutritional Status - Income Relationship (SICHS)

Dep. var.:	adult BMI					
Independent variable:	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)
Slope below	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	22.948	23.684	23.034	23.755	23.034	23.755
N	2426	5208	1810	3844	1810	3844

Source: South India Community Health Study (SICHS)
The following covariates: gender, age (linear, quadratic, and cubic terms), and caste group are included in the estimating equation.
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.

B Mathematical Appendix

B.1 Proofs of Propositions

Proof of Proposition 1: We first derive closed-form expressions for $e^L(y_t)$, $e^H(y_t)$. Focusing on the numerator of the $e^L(y_t)$ expression in (5), 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,

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

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 $\frac{\mathrm{d}\,\mathbb{E}(z_t|y_t,y_t>\alpha)}{\mathrm{d}\,y_t} > b$.

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

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

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),

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

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

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

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

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

From equation (B.1), $\frac{\mathrm{d}\, P(D_t|y_t, y_t \leqslant \alpha)}{\mathrm{d}\, y_t} = 0$ and from equation (B.2), $\frac{\mathrm{d}\, P(D_t|y_t, y_t > \alpha)}{\mathrm{d}\, y_t} > 0$ 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 (B.2) eliminates the term inside square brackets, implying that there is no level discontinuity at $y_t = \alpha$.

B.2 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

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

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:

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

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:

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

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

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

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 $y_t \leqslant \alpha$,

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

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,

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

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

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

which is independent of y_t.

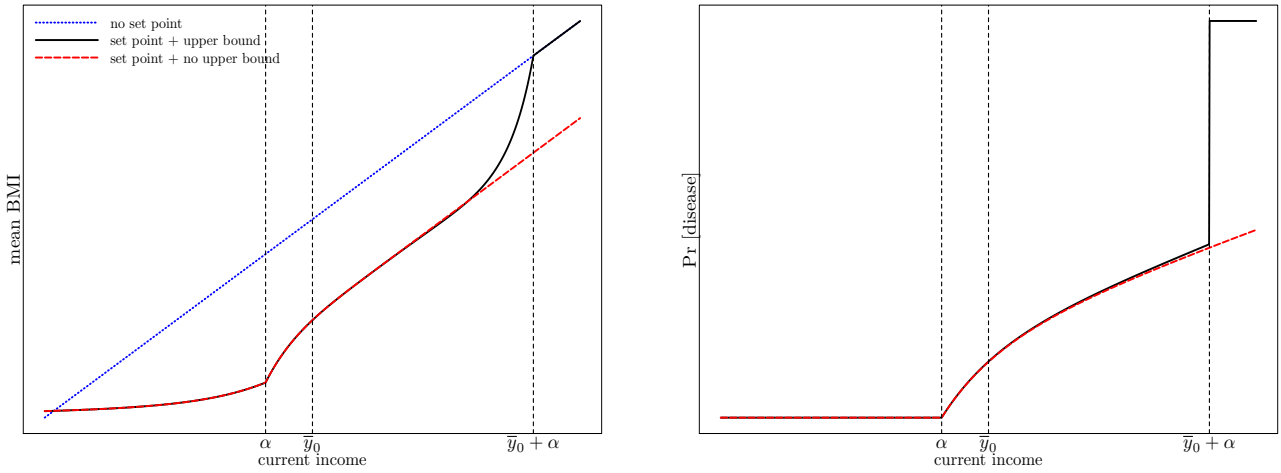
Although analytical results can no longer be derived as in Propositions 1 and 2, expressions (7), (8), (B.3), (B.4), (B.5) and (B.6) can be used to simulate the relationship between current income and both BMI and the probability of metabolic disease. We use the actual income from the IHDS and the estimates of μ_t, σ_t from the structural estimation exercise for the simulation. The left panel in Figure B1 plots the relationship between BMI and current income, with and without the upper bound on y_0. In both cases, we observe a slope discontinuity at $y_t = \alpha$. However, when there is an upper bound on y_0, we observe another slope discontinuity at $y_t = \overline{y}_0 + \alpha$. To the right of this point, BMI varies linearly with current income with slope b, and to the left the slope is greater than b. The right panel in Figure B1 plots the relationship between the probability of metabolic disease and current income, with and without the upper bound on y_0. There is a slope discontinuity at $y_t = \alpha$ in both cases. With an upper bound on y_0, there is, in addition, a level discontinuity at $y_t = \overline{y}_0 + \alpha$ and no relationship with y_t thereafter. In practice, we do not observe a second discontinuity, at a high income level, with either BMI or the risk of metabolic disease as outcomes.

B.3 Augmented metabolic disease - income relationship

Instead of being characterized by equation (4), suppose that the risk of metabolic disease is decreasing in ancestral income, below and above the threshold, independent of the income mismatch. The relationship between the probability of metabolic disease and income can then be characterized as

$$P(D_t) = \begin{cases} \gamma_1 + \gamma_0 y_0 & \text{if } U_t \leqslant \alpha \\ \gamma_1 + \gamma_0 y_0 + \gamma_2 (y_t - y_0) & \text{if } U_t > \alpha \end{cases} \tag{B.7}$$

Figure B1: Simulated Cross-Sectional Relationships (with and without Upper Bound on y_0)



(a) BMI (b) Metabolic disease

where $\gamma_0 < 0$. The probability of metabolic disease for $y_t \leqslant \alpha$ can be written as

$$\begin{aligned} P(D_t|y_t, y_t \leqslant \alpha) &= \int_{-\infty}^{y_t} (\gamma_1 + \gamma_0 y_0) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ &= \gamma_1 + \gamma_0 \int_{-\infty}^{y_t} (y_t - U_t) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ &= \gamma_1 + \gamma_0 y_t - \gamma_0 \int_{-\infty}^{y_t} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \end{aligned}$$

We use the following identity

$$\int_{x_1}^{x_2} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \equiv \frac{\sigma_t \left[\phi\left(\frac{x_1 - \mu_t}{\sigma_t}; 0, 1\right) - \phi\left(\frac{x_2 - \mu_t}{\sigma_t}; 0, 1\right) \right] + \mu_t \left[\Phi\left(\frac{x_2 - \mu_t}{\sigma_t}; 0, 1\right) - \Phi\left(\frac{x_1 - \mu_t}{\sigma_t}; 0, 1\right) \right]}{\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)} \tag{B.8}$$

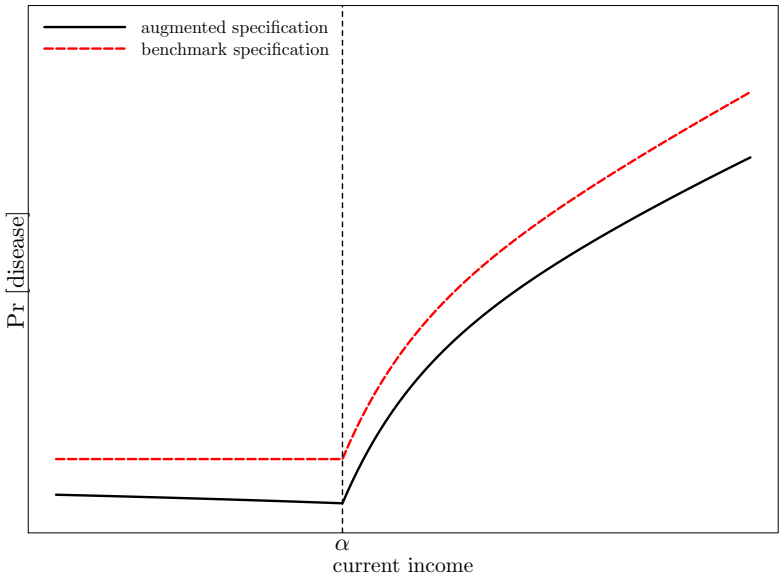
to write $P(D_t|y_t, y_t \leqslant \alpha)$ as

$$P(D_t|y_t, y_t \leqslant \alpha) = \gamma_1 + \gamma_0 y_t - \gamma_0 \left[\mu_t - \sigma_t \frac{\phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)}{\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)} \right] \tag{B.9}$$

For $y_t > \alpha$, $P(D_t|y_t, y_t > \alpha)$ can be written as

$$\begin{aligned} P(D_t|y_t, y_t > \alpha) &= \int_{-\infty}^{\alpha} (\gamma_1 + \gamma_0 y_0) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t + \int_{\alpha}^{y_t} (\gamma_1 + \gamma_0 y_0 + \gamma_2 (y_t - y_0)) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ &= \gamma_1 + \gamma_0 y_t - \gamma_0 \int_{-\infty}^{y_t} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} + \gamma_2 \int_{\alpha}^{y_t} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \end{aligned}$$

Figure B2: Simulated Metabolic Disease-Income Relationship



Using the identity (B.8), the expression for $P(D_t|y_t, y_t > \alpha)$ can be simplified to

$$P(D_t|y_t, y_t > \alpha) = \gamma_1 + \gamma_0 y_t - \gamma_0 \left[\mu_t - \sigma_t \frac{\phi\left(\frac{y_t-\mu_t}{\sigma_t}; 0, 1\right)}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t}; 0, 1\right)} \right] + \gamma_2 \left[\frac{\sigma_t \left\{ \phi\left(\frac{\alpha-\mu_t}{\sigma_t}; 0, 1\right) - \phi\left(\frac{y_t-\mu_t}{\sigma_t}; 0, 1\right) \right\} + \mu_t \left\{ \Phi\left(\frac{y_t-\mu_t}{\sigma_t}; 0, 1\right) - \Phi\left(\frac{\alpha-\mu_t}{\sigma_t}; 0, 1\right) \right\}}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t}; 0, 1\right)} \right] \tag{B.10}$$

Differentiating equation (B.9) with respect to y_t, and given the properties of the inverse Mills ratio, it is straightforward to verify that $P(D_t|y_t, y_t \leq \alpha)$ is decreasing in y_t if $\gamma_1 < 0$. Comparing equations (B.9) and (B.10), $P(D_t|y_t, y_t \leqslant \alpha) = P(D_t|y_t, y_t > \alpha)$ for $y_t = \alpha$. Hence, there is no mean discontinuity. Although we cannot analytically derive the remaining elements of Proposition 2, we can use equations (B.9) and (B.10) to simulate the relationship between probability of metabolic disease and current income. Figure B2 plot the result from this simulation. For comparison we plot two specifications. The black solid line refers to the simulation when $\gamma_1 = 0.01, \gamma_0 = -0.02$ and $\gamma_2 = 0.02$. The red dashed line refers to the same specification except that we set $\gamma_0 = 0$ as in the benchmark specification. The relationship between the probability of metabolic disease and current income under the two specifications is almost identical except for the mild negative slope to the left of the threshold with the augmented specification and the level difference. Notice that both specifications generate a slope discontinuity at $y_t = \alpha$.

B.4 Augmented nutritional status-income relationship

Instead of being characterized by equation (3), where the nutritional status-income relationship is independent of ancestral income below the threshold, suppose that nutritional status is increasing more steeply in current income for dynasties with lower ancestral income. The relationship between nutritional status and income relationship can then

be characterized as

$$z_t = \begin{cases} a + by_0 & \text{if } U_t \leqslant \alpha \\ a + by_t + cy_t y_0 & \text{if } U_t \leqslant \alpha \end{cases} \tag{B.11}$$

where $c < 0$. The mean BMI at $y_t \leqslant \alpha$ is given as

$$\mathbb{E}[z_t | y_t, y_t \leqslant \alpha] = a + b(y_t - e^L(y_t)) \tag{B.12}$$

For mean BMI at $y_t > \alpha$, we can write,

$$\mathbb{E}[z_t | y_t, y_t > \alpha] = \int_{-\infty}^{\alpha} (a + by_0) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t + \int_{\alpha}^{y_t} (a + by_0 + cy_0 y_t) \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t$$

Substituting $y_0 = y_t - U_t$, we get

$$\begin{aligned} \mathbb{E}[z_t | y_t, y_t > \alpha] = a + by_t + cy_t^2 \int_{\alpha}^{y_t} \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t - b \int_{-\infty}^{\alpha} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \\ - cy_t \int_{\alpha}^{y_t} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, \mathrm{d}\, U_t \end{aligned}$$

We use the identity (B.8) to write

$$\begin{aligned} \mathbb{E}[z_t | y_t, y_t > \alpha] = a + by_t + cy_t^2 \left[1 - \frac{\Phi(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1)}{\Phi(\frac{y_t - \mu_t}{\sigma_t}; 0, 1)} \right] - b \left[\frac{\mu_t \Phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right) - \sigma_t \phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right)}{\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)} \right] \\ - cy_t \left[\frac{\sigma_t \left[\phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right) - \phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right) \right] + \mu_t \left[\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right) - \Phi\left(\frac{\alpha - \mu_t}{\sigma_t}; 0, 1\right) \right]}{\Phi\left(\frac{y_t - \mu_t}{\sigma_t}; 0, 1\right)} \right] \end{aligned} \tag{B.13}$$

The BMI-income relationship to the left of the threshold is unaffected by the augmented specification. Note also that at $y_t = \alpha$, $\mathbb{E}[z_t | y_t, y_t \leqslant \alpha] = \mathbb{E}[z_t | y_t, y_t > \alpha]$ implying that there is no mean discontinuity. However, we cannot analytically derive the remaining elements of Proposition 1, in particular, the BMI-income relationship to the right of the threshold. Using equations (B.12) and (B.13) we thus simulate the relationship between nutrition status and current income. Figure B3 plots the result from this simulation for two different values of $c = \{-0.5, 0\}$. The black solid line refers to the augmented specification with $c = -0.5$ and the dashed red line refers to the benchmark specification with $c = 0$. In both specification, there is slope discontinuity instead of mean discontinuity at $y_t = \alpha$, with a steeper slope to the right of the threshold as in Proposition 1.

B.5 Aggregate BMI equation derivation

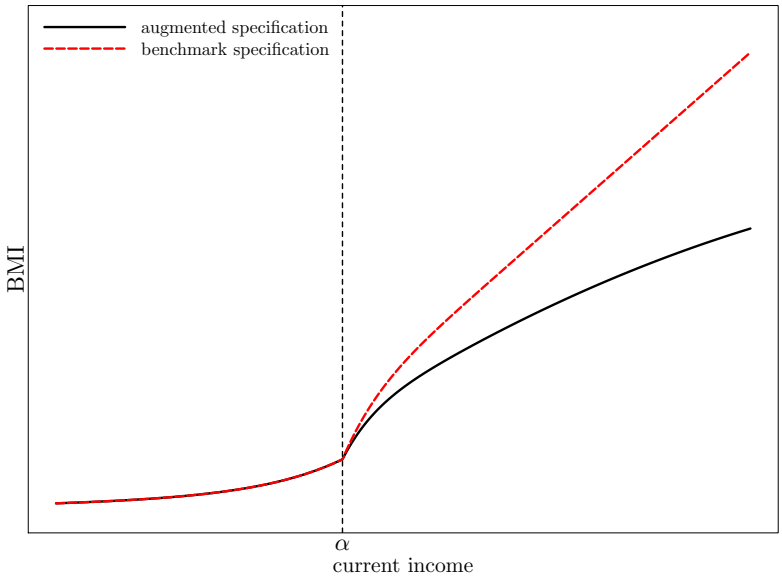
Let $\underline{y}_t^j$ denote the income threshold above which households escape their set point.

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

By the property of the normal distribution,

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

Figure B3: Simulated Nutritional Status-Income Relationship



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:

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