ECONOMIC DEVELOPMENT, UNDERNUTRITION AND DIABETES

By

Kaivan Munshi, Swapnil Singh, Nancy Luke,

and Anu Mary Oommen

September 2024

COWLES FOUNDATION DISCUSSION PAPER NO. 2407

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Economic Development, Undernutrition and Diabetes

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

September 17, 2024

Abstract

This research connects two seemingly unrelated facts that have recently been documented in developing countries, with important consequences for global health: (i) the weak association between nutritional status and income, and (ii) the elevated risk of diabetes among normal-weight individuals. The model that we develop to reconcile these facts is based on a set point for body size that is adapted to (low) pre-modern food supply, but subsequently fails to adjust to rapid economic change. During the process of development, some individuals thus remain at their low-BMI set point, despite the increase in their income (food consumption), while others who have escaped their set point (but are not necessarily overweight) are at increased risk of diabetes. The model is tested along different dimensions with multiple data sets. Our analysis indicates that many lean diabetics in developing country populations will be close to their individual-specific set point, suggesting a promising approach to diabetes control (reversal) that involves relatively little weight loss.

*Munshi: Yale University and Toulouse School of Economics, kaivan.munshi@yale.edu [corresponding author]. Singh: Bank of Lithuania and Kaunas University of Technology, ssingh@lb.lt. Luke: Pennsylvania State University, nkl10@psu.edu. Oomen: Christian Medical College, anuoommen@cmcvellore.ac.in. 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.

1 Introduction

Two recently documented facts run counter to the conventional wisdom that economic development leads to better health: first, the absence of a clear link between nutritional status and income in developing countries (Deaton, 2007; Swaminathan et al., 2019) and second, a surge in (type 2) diabetes in these countries (Diamond, 2011; Narayan and Kanaya, 2020). Our objective in this paper is to develop and test a model with three ingredients – adaptation, mismatch and a set point – that can reconcile these seemingly unrelated observations. Our model is also able to explain why a surprisingly large fraction of diabetics in developing-country populations are normal weight, as documented below for India, with resulting implications for their treatment.

The pre-modern (Neolithic) economy was characterized by wide short-term fluctuations in food supply, but had growth rates close to zero for centuries. This resulted in a population whose body size was adapted to long-term (low) food supply, with the adaptation varying across space with agroclimatic conditions (Pomeroy et al., 2019; Dalgaard et al., 2021). With economic development, there is a substantial increase in income and food consumption. The developmental origins of adult disease literature posits that the resulting mismatch between current and ancestral consumption (to which the population is adapted) has contributed to the high rates of diabetes in developing countries (Gluckman and Hanson, 2004; Wells et al., 2016; Narayan and Kanaya, 2020). Our model incorporates the mismatch hypothesis, while simultaneously explaining the persistence of undernutrition in these countries by characterizing the initial adaptation by a set point.

Many individuals have stable bodyweight throughout adult life (Leibel, 2008). This set point for bodyweight is part of a homeostatic (stabilizing) system that maintains the body's energy balance against fluctuations in food intake by making metabolic and hormonal adjustments (Müller et al., 2010, 2018).[1] We posit that the set point for a given dynasty (family) is determined by food supply in the pre-modern economy. While the adapted set point would have allowed pre-modern populations to maintain their energy balance, and to survive and reproduce in an environment characterized by low and fluctuating food supply, it becomes a liability if it persists for multiple generations after the onset of economic development.

A property of all – physical and biological – homeostatic systems is that they can only self-regulate within fixed bounds and will malfunction when those bounds are exceeded (Stebbing, 2009; Kültz, 2020). This implies that as long as current and pre-modern (ancestral) consumption or, equivalently, income remain sufficiently close to each other, the body will successfully defend its set point. Nutritional status will remain at its pre-modern level for such individuals, despite the increase in their consumption with economic development. Once the gap between current and pre-modern income crosses a threshold, however, the body will no longer be able to defend the set

[1] Homeostasis is a fundamental concept in biology, which describes how physiological systems maintain an equilibrium set point by counteracting environmental stresses. As discussed in Leibel (2008); Müller et al. (2010); Speakman et al. (2011), numerous studies indicate that when the energy balance is perturbed in either direction through a change in diet, the body returns to its original weight once the nutritional constraint is released. Furthermore, energy expenditures are modulated to resist the perturbation, indicating that the body is actively defending its set point.

point. Escape from the set point for these other individuals, who are not necessarily overweight, is associated with imbalance in energy regulation, which will be accompanied by imbalance in related (inter-linked) homeostatic systems. Failure of glucose homeostasis, in particular, is the proximate cause of diabetes.[2]

Although it may be appropriate to characterize the set point with respect to weight for a given individual, we account for possible variation in height across generations by specifying a common set point for members of a dynasty with respect to their BMI; i.e. weight conditional on height. This normalization is especially useful for our analysis because BMI is a standard measure of adult nutritional status and is also associated with the risk of diabetes.[3] Based on the discussion above, it follows that there will be two types of individuals in a developing economy: (i) Those individuals who remain at their pre-modern set point, despite the increase in their consumption, are partly responsible for the weak association between nutritional status, which we measure by BMI, and income. (ii) Those individuals who have escaped their set point, but are not necessarily overweight, are at increased risk of diabetes and accompanying metabolic disorders.

Economists are familiar with the concept of institutional adaptation and mismatch. For example, Greif (1994) in his pioneering contribution to the comparative institutions literature, posits that the informal networks that supported cooperation in the pre-modern economy may have prevented their members from taking advantage of new market opportunities with economic development, resulting in a dynamic inefficiency. Our analysis is concerned with biological adaptation to economic conditions (food supply) in the pre-modern period and the subsequent mismatch that accompanied economic development due to its persistence. If data on income, BMI, and diabetes were available for each family (dynasty) over many generations, going back to the pre-modern era, then we could test the set-point based argument directly. For a given dynasty, we would expect to observe a discrete increase in BMI in a particular generation in which the gap between current and ancestral income exceeded a threshold, with an accompanying increase in the risk of diabetes. In the absence of such multi-generational household-level data, our analysis proceeds as follows:

First, by characterizing the evolution of income in the population across generations during the process of development, the dynamic model laid out in Section 2 generates cross-sectional implications at any point in time that do not require knowledge of ancestral income: (i) Although BMI is increasing in current household income at all levels, there is a discontinuous increase in the slope of this association at a particular income threshold. (ii) The risk of diabetes is constant below the same threshold and increasing in current income above the threshold. Viewed through the lens of the model, these cross-sectional associations across households are informative about underlying causal relationships within households (dynasties) over generations. However, such causal interpretations

[2]In related research, Taylor and Holman (2015) hypothesize that there is an individual-specific weight threshold above which the risk of diabetes increases discontinuously, but they do not provide evolutionary foundations for the threshold nor do they incorporate a role for a set point (which is necessary for our analysis).

[3]Height is another common measure of nutritional status and archaeological evidence indicates that Neolithic populations adapted to low food supply by adjusting their stature (Pomeroy et al., 2019). Although this is not the focus of our analysis, we provide evidence supporting this complementary adaptation in Section 3.

are only appropriate if the model is correctly specified and, thus, much of the analysis will be devoted to validating the model.

Second, we verify the cross-sectional implications of the model in Section 3, using Hansen's (2017) slope-threshold test, with nationally representative household data from the India Human Development Survey (IHDS). The weak association between BMI and household income below the estimated threshold, which is located close to the median income level in the Indian population, explains (in part) the persistence of undernutrition in that population. The steep increase in the risk of diabetes with income above the same threshold, which corresponds to a BMI that is in the middle of the normal range, helps explain the second stylized fact. Our interpretation of these twin findings is that BMI and the risk of diabetes increase simultaneously and independently when an underlying homeostatic system (maintaining a low BMI) malfunctions. In contrast, we find that nutrient intake (food consumption) is increasing continuously in household income in Section 3, as assumed in the model.

Third, we empirically validate the model along different dimensions in Section 4. We begin by verifying not only that a set-point threshold is present, but also the specific structure that is imposed on the BMI-income relationship in the model. An accompanying quantification exercise tells us that the fraction of underweight adults in India, who comprise 20% of the population, would decline by 24% in the absence of the BMI set points. Next, we test an additional implication of the model, which is that the positive association between the risk of diabetes and BMI is also characterized by a slope discontinuity. For this exercise, we use data from the IHDS and the Indian DHS, which provides information on BMI and diabetes with biomarkers (but not income) for a large number of individuals. We precisely estimate a discontinuous increase in the risk of diabetes at a BMI below 23, which is well within the normal range, with both data sets.[4] Finally, we replicate the cross-sectional tests of the model with data from the Indonesia Family Life Survey (IFLS). In line with the observation that the gap between current income and historical (ancestral) income is greater in Indonesia than in India, the location of the precisely estimated income threshold with IFLS data indicates that three-quarters of the population has escaped its set point in that country.

Fourth, we empirically examine the mechanism underlying our model: (a) adaptation implies that BMI for individuals below the income threshold we have estimated will be determined by their ancestral income (and not by current income). (b) The mismatch hypothesis implies that the risk of diabetes will be increasing in the gap between current income and ancestral income, but only for individuals who have escaped their set point and thus must be above the estimated threshold. We test these implications of the adaptation-mismatch mechanism in Section 5 by constructing exogenous measures of ancestral (pre-modern) per household income at (i) the village level, using data on the agricultural revenue tax that was collected by the British colonial government in 1871, based on its

[4] Diabetes is self-reported and, hence, under-reported in the IHDS. For all analyses that utilize self-reported health data in this paper, we thus construct a composite variable, which we refer to as "metabolic disease" that indicates whether an individual has been diagnosed with diabetes or with either of two highly correlated comorbidities: hypertension and cardiovascular disease. This variable tracks direct measures of diabetes, based on biomarkers, that are obtained from the DHS, across the range of BMI's.

independent assessment of local agricultural productivity, and (ii) at the district level with FAO-GAEZ crop suitability data, using a method suggested by Galor and Özak (2016). The village-level measures, which are available for villages in the modern Indian state of Tamil Nadu, are merged with data from the South India Community Health Study (SICHS) which we directed and which provides information on income, BMI and diabetes for a representative sample of households in rural Vellore district. The district-level measures of ancestral income 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.

The analysis described above provides direct support for the long shadow of historical consumption on contemporary diabetes. Early contributions to the developmental origins of adult disease literature, going back to Barker (1995), focussed on the mismatch between conditions in adulthood and in *utero* as determinants of diabetes. Many empirical studies have tested this variant of the mismatch hypothesis, exploiting shocks caused by wars or famine to establish the link between adult diabetes and early-life conditions; e.g. Ravelli et al. (1998); Li et al. (2010). However, Wells et al. (2016) is the only previous study that we are aware of that tests the mismatch hypothesis with respect to more distant ancestral consumption. As discussed in Section 5, their analysis is subject to alternative interpretations. Franck et al. (2022) provide more credible evidence in support of a long-term mismatch, but in the context of historical adaptation to infectious diseases and its consequences for autoimmune and inflammatory conditions today. We complete the analysis of adaptation and mismatch by considering alternative explanations, such as unobserved changes in diet or lifestyles, poverty traps, or adaptation to conditions in *utero* rather than to the distant past, showing that none of them can be reconciled with the specific patterns that we uncover in the data.

How long do we expect the mismatch we have uncovered, with its consequences for diabetes, to persist? The assumption in many evolutionary models is that the initial adaptation is epigenetic; i.e. it involves changes in gene expression and, hence, will persist for a limited number of generations (Jablonka and Raz, 2009; Lind and Spagopoulou, 2018). This would explain why European populations, which were also under-nourished historically, no longer exhibit the traits we document in this paper.[5] At the same time, if the health experience of migrants from developing countries to substantially wealthier advanced economies is any indication, then we would expect to observe elevated risks of diabetes in developing-country populations for many generations. For example, Alacevich and Tarozzi (2017) document that the nutritional status of immigrants from South Asia to the U.K. converges to the level of the native population very swiftly, presumably because they have escaped their set points. Given the persistence in these underlying set points, South Asian immigrants residing in the U.K. and the U.S. nevertheless remain many times more likely to have diabetes, conditional on their BMI, than the native population (McKeigue et al., 1991; Oza-Frank and Narayan, 2010).

[5] In an advanced economy, the pre-modern set point is no longer in place and food supply is not a constraint. Wealthier individuals will have healthier diets and lifestyles. As a result, both BMI and the risk of diabetes are declining with wealth, in contrast with what we observe in developing economies.

CDC statistics indicate that 9.5% of diabetics in the U.S. are normal weight (with a BMI below 25). Using the same criterion, we find that 63% of diabetics in the 2015-16 round of the India DHS are normal weight. Our model provides an explanation for this difference, based on low-BMI set points that are specific to developing-country populations and which result in a relatively high fraction of normal weight individuals in the population *and* a high rate of diabetes among these individuals. Many (lean) diabetics in these populations, who have recently escaped their individual-specific BMI thresholds and are thus close to their set points, could potentially reverse their condition with relatively little weight loss. This implies that behavioral interventions could be especially effective, and we will return to this observation in the concluding section.

2 The Model

2.1 Population and Income

The population consists of a large number of infinitely lived dynasties (families). Each dynasty consists of a single individual in each generation, who is replaced by a single descendant in the generation that follows. There is a fixed return on wealth in each generation; i.e. a permanent income flow, which is consumed, so that the stock is passed on (without depletion) to the next generation. Wealth or (permanent) income, we will use these terms interchangeably in the discussion that follows, is the same in each generation during the pre-modern era in which adaptation takes place, but subsequently evolves. Denote the logarithm of the dynasty's initial income by y_0. Permanent income in the modern economy is well approximated by the log-normal distribution (Battistin et al., 2009). We thus assume that each dynasty receives a permanent, additive and independent income shock u_τ in each subsequent period or generation 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 $U_t = \sum_{\tau=1}^{t} u_\tau \sim N(t\mu, t\sigma^2)$. For ease of exposition, we will denote $t\mu$ by μ_t and $t\sigma^2$ by σ_t^2.

2.2 Biological Relationships

We now characterize the biological relationships between (i) BMI and income, and (ii) the risk of diabetes and income, during the process of economic development. Although these relationships apply more proximately with respect to (food) consumption, they can be specified with respect to income under the assumption that there is a positive and continuous association between consumption and income, as verified in Section 3. Focussing first on the initial period in which the set point is determined, it follows that nutritional status, which we measure by BMI z_0, is increasing continuously in pre-modern income y_0, as specified below:

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

In subsequent periods, each descendant's body will defend their dynasty's set point z_0 in the face of changes in consumption that arise due to the permanent income shocks. However, as noted, the body can only respond up to a point to deviations in income from the initial level, y_0, that determined the set point. There is thus a threshold α, such that BMI in period t,

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

Equation (3) imposes the restriction that the (linear) relationship between BMI and income is the same, below and above the threshold; what changes is the relevant measure of income, from y_0 to y_t. In Section 3, we will validate the structure we have imposed on equation (3) by separately estimating the b parameter, below and above the (estimated) threshold.[6]

Notice that we do not specify a lower threshold for the set point. Given low levels of food supply in the pre-modern era, the population would have been adapted to defend the set point especially vigorously against downward fluctuations in consumption.[7] Although mean income is increasing with economic development 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.[8]

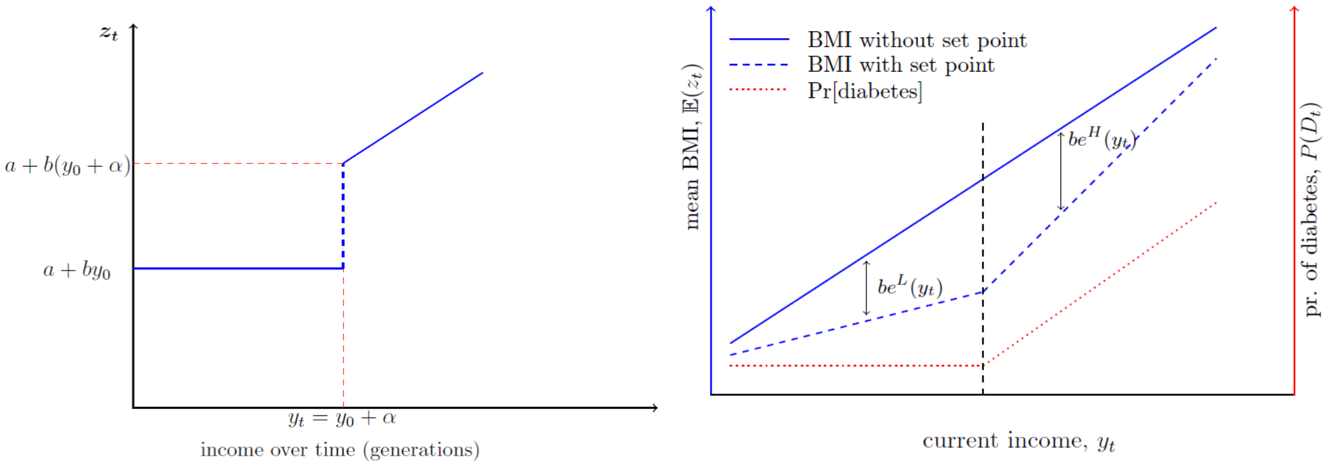
As long as income remains within the threshold associated with the dynasty's set point, metabolic and hormonal adjustments ensure that the increases in consumption that accompany the increases in income due to economic development do not translate into increases in BMI. Once income crosses the threshold,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

[6] While we focus on adaptation with respect to body size, in line with the modern evolutionary biology literature, the "thrifty genotype" hypothesis (Neel, 1962) posits that body weight in historically undernourished populations will be more responsive to the increase in food consumption that accompanies economic development. This implies that there should be an additional $y_o \cdot y_t$ term above the threshold in equation (3), with a negative coefficient. If that were the case, however, then we would fail the validation test that follows in Section 4.

[7] This is consistent with the conventional view that the regulation of bodyweight is more responsive to weight loss than to weight gain (Müller et al., 2010). For example, despite repeated weight cycling in response to seasonal fluctuations in food supply, minimal bodyweight in a sample of rural Gambian women remained extremely stable (within 1.5 kg.) over a period of 10 years (Prentice et al., 1992).

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

Figure 1: BMI, Diabetes and Income



(a) Within a dynasty over time

(b) Across households in the cross-section

income can thus be characterized as follows:[9]

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

2.3 Cross-Sectional BMI-Income Association

Figure 1a describes the evolution of BMI across multiple generations (time periods) for a single dynasty, based on the biological relationship specified above. For expositional convenience, we assume that the dynasty only receives positive income shocks. Starting from an initial income, y_0, the dynasty's income thus increases monotonically across generations. However, its members' BMI will remain at the dynasty's set point, $z_0 = a + by_0$, until y_t exceeds $y_0 + \alpha$. At that point in time, there will be a discrete increase in BMI, after which BMI will track y_t. If data over many generations, going back to the pre-modern period, were available for each dynasty, then these implications could be tested directly. In the absence of such multi-generational data, we proceed to derive the cross-sectional association between BMI and income, as implied by equation (3), when a dynasty-specific set point for BMI is present.

We normalize so that the initial income distribution is bounded below at zero. We also do not specify a lower threshold for the set point. It follows that all individuals with $y_t \leq \alpha$ must be at their set point; some of these individuals will belong to dynasties that had initial incomes below α and which subsequently increased their income by relatively little, whereas others will belong to dynasties whose income has drifted down over time. Mean BMI at any given level of income $y_t \leq \alpha$

[9] $\gamma_1 > 0$, $\gamma_2 > 0$ in equation (4). The implicit assumption, which is consistent with recent evidence on diabetes reversal; e.g. Lean et al. (2018) is that the risk of diabetes can change in both directions over time as the individual's BMI shifts on either side of the threshold.

can then be characterized by the following expression:

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

where $P(U_t|y_t)$ is the conditional density function of U_t given y_t. As shown in Appendix A, our distributional assumptions together with a simplifying (empirically validated) analytical approximation allow us to express the preceding equation as follows:

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

where $e^L(y_t) = \frac{1}{\Phi(y_t;\mu_t,\sigma_t^2)} \int_{-\infty}^{y_t} U_t \phi(U_t; \mu_t, \sigma_t^2) \, \mathrm{d}\, U_t = \mu_t - \sigma_t \Lambda\left(\frac{y_t - \mu_t}{\sigma_t}\right)$ and $\Lambda(\cdot)$ is the inverse Mills ratio.

For individuals with $y_t > \alpha$, some will have crossed their set point threshold, while others (who started with a higher initial income) will remain at their set point. The expression for mean BMI at a given level of income $y_t > \alpha$ thus includes both types of individuals. Incorporating the same analytical approximation and distributional assumptions as above:

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

where $e^H(y_t) = \frac{1}{\Phi(y_t;\mu_t,\sigma_t^2)} \int_{-\infty}^{\alpha} U_t \phi(U_t; \mu_t, \sigma_t^2) \, \mathrm{d}\, U_t = \frac{\mu_t \Phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right) - \sigma_t \phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right)}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)}$

Given the specifications of the $e^L(y_t)$, $e^H(y_t)$ functions, we can derive the following result (the proof is in Appendix A):

Proposition 1 *(i) The slope of the BMI-income association 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 association, but no level discontinuity, at* $y_t = \alpha$.

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

The preceding implication is robust to alternative specifications of the set point. Although an epigenetically determined set point may be heritable, it will ultimately cease to be relevant once a changed economic environment has been in place for a sufficient number of generations. Our model thus describes the relationship between nutritional status and income over a finite number of generations during the initial phase of economic development. During this phase, we assume that the set point, z_0, determined in period 0, is fixed. However, an alternative specification would allow the set point to adjust gradually across generations until it is no longer relevant. For example, the set point could be specified as a weighted average of y_0 and y_t, with the weight on y_t increasing over time. Alternatively, the set point could be determined by conditions in *utero* in each generation. Since income does not vary within periods in our setup, the set point in period t with this alternative specification will then be parental income, y_{t-1}. As shown in Appendix A, the alternative specifications generate the same qualitative predictions as Proposition 1. What distinguishes the benchmark specification in equation (3) from the alternatives, as verified empirically in Section 5, is that BMI below the estimated current-income threshold is determined exclusively by y_0.

2.4 Cross-Sectional Diabetes-Income Association

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

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

The proof in Appendix A follows the same steps as the proof of Proposition 1. The $P(D_t) - y_t$ association specified by Proposition 2 is described graphically in Figure 1b. This association is qualitatively the same as the $\mathbb{E}(z_t) - y_t$ association, except that the slope is zero below the threshold. This is because the risk of diabetes is constant (and the same) for all individuals who remain at their set point and because all individuals below the income threshold are at their set point. Above the threshold, in contrast, the risk of diabetes 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 $P(D_t) - y_t$ associations will exhibit a slope discontinuity at the same income level: $y_t = \alpha$.[10]

Proposition 1 indicates that BMI is increasing with income at all levels, more steeply above a threshold, while Proposition 2 indicates that the risk of diabetes is only increasing in income above the same threshold. Bringing the two implications together, it follows that there will be no association between the risk of diabetes and BMI up to a BMI threshold (which corresponds to the underlying income threshold) and a positive association thereafter. Although the cross-sectional tests of the model that follow in Section 3 focus on the BMI-income and diabetes-income associations, we will examine this additional implication of the model in Section 4.

[10] 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.

3 Cross-Sectional Tests

3.1 Nutritional Status and Diabetes with respect to Income

The primary data set that we use to test the model is the India Human Development Survey (IHDS), which was conducted in 2004-2005 and 2011-2012. Although a dynasty consists of a single individual in each generation in our model, multiple individuals will reside in a household in practice. Income is thus measured at the household level, as the average over the 2004 and 2012 rounds. 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.[11] Nutritional status is measured by BMI, derived from the weight and height of the household head and his spouse in each survey round. Diabetes is self-reported and, hence, under-reported in the IHDS. We thus construct a composite variable, "metabolic disease," which indicates whether a given individual has been diagnosed with diabetes or with either of two highly correlated comorbidities: hypertension and cardiovascular disease (Petrie et al., 2018). This indicator, which is validated in Section 4, is constructed for the household head and his spouse in each survey round, consistent with the implicit assumption in the model that diabetes is reversible, and with recent experimental evidence (Lean et al., 2018).[12]

We test the cross-sectional implications of the model by nonparametrically estimating the BMI-income and metabolic disease-income associations using the variables described above. Although our analysis focuses on the association with income, other individual and household characteristics, which are omitted from the model for expositional convenience, could independently determine BMI and the risk of diabetes. For example, both outcomes could vary with age and gender. There could also be spatial variation in food tastes, as emphasized by Atkin (2013, 2016), or in the disease environment, as documented by Dandona et al. (2017). All of the estimating equations in our analysis thus include the following standard set of covariates: age in years (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district and survey-round. These covariates are partialled out using the Robinson (1988) procedure prior to the nonparametric estimation reported in Figure 2a.[13]

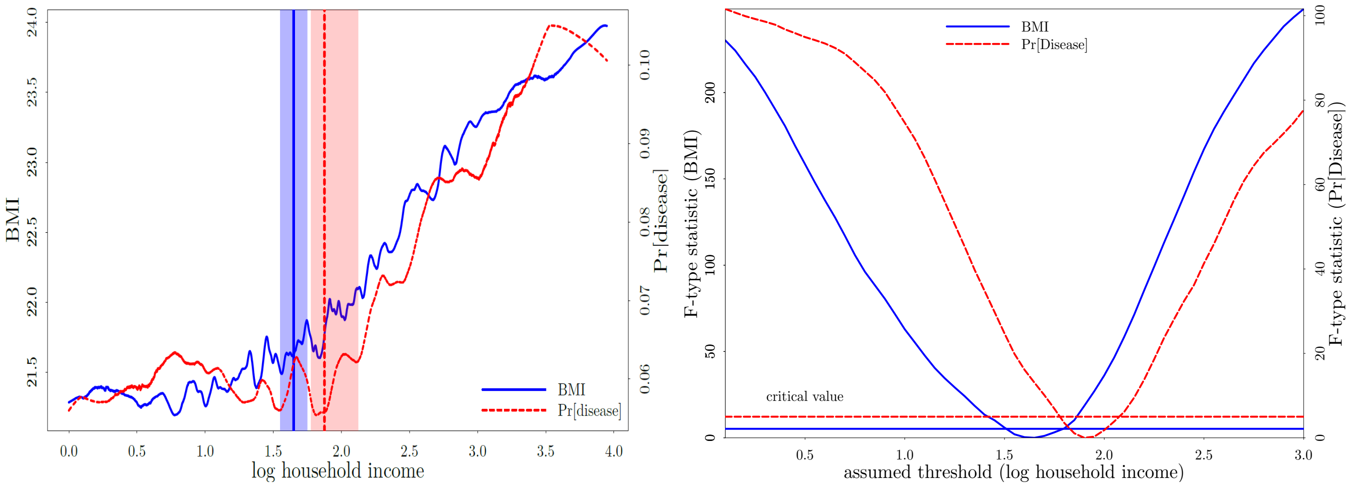
The vertical lines in Figure 2a mark the point where we locate an income threshold, based on the statistical test described below. The shaded area around each line marks the 95% confidence interval for the threshold location, based on the same test. It is evident with each outcome that the

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

[12] The construction of our composite variable is motivated by the observation that diabetics have a substantially elevated risk of hypertension and cardiovascular disease (Petrie et al., 2018). However, these diseases can also occur independently, which adds noise to our measure of diabetes. We verify in Section 4 that this variable, nevertheless, tracks closely with direct measures of diabetes obtained from the DHS.

[13] The Robinson procedure is described in Appendix B. Observations in the top and bottom 1% of the outcome distribution are excluded from the estimation sample in all of our analyses. This ensures that the estimation results are not driven by extreme outliers.

Figure 2: Nutritional Status and Metabolic Disease with respect to Household Income



(a) Nonparametric association

(b) Threshold test

Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are 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.

association with income is relatively weak below the estimated threshold, and much stronger above the threshold. A slope discontinuity is not visually apparent with BMI as the outcome in Figure 2a. However, we can detect its presence with a high degree of statistical confidence, and sharper discontinuities will be observed with other datasets (IFLS, SICHS). Notice also that the estimated threshold location is slightly lower with BMI as the outcome. Such minor differences are to be expected, given that BMI is directly measured, whereas metabolic disease (although diagnosed) is self reported. Nevertheless, this discrepancy is not observed in the robustness tests that follow and in the subsequent analyses with South Indian (IHDS) and Indonesian (IFLS) data.

The threshold locations and confidence intervals in Figure 2a are estimated using a procedure developed by Hansen (2017). This procedure involves sequential estimation of the following piecewise linear equation:

$$z_i = \beta_0 + \beta_1 y_i + \beta_2 (y_i - \tau) \times \mathbb{I}(y_i - \tau > 0) + x_i \lambda + \epsilon_i, \tag{7}$$

where z_i is an outcome of interest; e.g. BMI or diabetes, 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 (the same covariates that are partialled out prior to nonparametric estimation). This equation is estimated at different assumed income thresholds (values of τ), starting at a very low income level and then covering the entire income range in small increments. An F-type statistic is computed at each assumed threshold, based on a comparison of the sum of squared residuals at that assumed threshold and the minimized value

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

Dependent variable:	BMI (1)	metabolic disease (2)
Baseline slope (β_1)	0.239**	0.002
	(0.057)	(0.002)
Slope change (β_2)	0.940**	0.028**
	(0.066)	(0.003)
Threshold location (τ)	1.65	1.90
	[1.55, 1.75]	[1.80, 2.05]
Threshold test $p-$value	0.000	0.000
Mean of dependent variable	22.002	0.074
N	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. BMI is measured for adults present in the household at the time of the survey.
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
** significant at 5%, based on cluster bootstrapped confidence intervals.

across all assumed thresholds. This statistic will have a minimum value of zero by construction, and the assumed income threshold corresponding to that value is thus our best estimate of the true threshold. If there is indeed a slope-change, then the F-type statistic will increase steeply as the assumed threshold moves away (on either side) from the income level at which it is minimized.

Figure 2b plots the F-type statistic across the range of assumed thresholds for each outcome. The assumed threshold (income level) at which the statistic is minimized corresponds to the location of the threshold in Figure 2a. The confidence interval for each threshold location in that figure is determined by the points of intersection between the F-type statistic and the 5% critical value line for the corresponding outcome in Figure 2b. The F-type statistic increases steeply as the assumed threshold moves away from the income level at which it is minimized for both outcomes, allowing us to locate the thresholds with a high degree of statistical confidence. Note that the threshold location is accurately estimated under the assumption that the nonlinear association between each outcome and income, as specified in the model, can be approximated by a linear spline function. If this approximation was not justified, then we would fail the test of internal validity that follows in Section 4 even if the model were correctly specified.

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 Figure 2b 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.[14] Moreover, a similar

[14]Following Hansen (2017) and Roodman et al. (2019), a coefficient's significance at the 5% level is determined by

bootstrap procedure can be used to test our statistical model with a slope change at an income threshold, as described in equation (7), 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. Although this does not rule out the possibility that the true relationship is actually (highly) nonlinear, without a discontinuity, the test of internal validity in Section 4 will provide additional statistical support for the specific structure we have imposed on the model.

The reported point estimates of the baseline slope coefficient (β_1) and the slope-change coefficient (β_2) in Table 1 are obtained at our best estimate of the true threshold, τ. As implied 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). Moreover, the slope to the left of the threshold is positive and significant with BMI, but not with the risk of diabetes (measured by metabolic disease) as the outcome.[15] The estimated threshold location ranges from 1.65 to 1.9 for the two outcomes and the median income in our nationally representative sample of households is 1.8. This implies that the lower half of the income distribution in India remains at its pre-modern BMI set point, whereas the upper half is at risk of diabetes.

We verify the robustness of the preceding evidence in a number of ways in Appendix B: (i) We include measures of household composition, which could independently determine decisions and behaviors that are relevant for nutritional status and health outcomes as additional covariates in the estimating equation. (ii) We construct a nonparametric shift-share instrument for household income, following Newey et al. (1999) and Goldsmith-Pinkham et al. (2020), that accounts for measurement error in the permanent income variable, as well as for possible reverse causality; i.e. the effect of BMI or metabolic disease on household income (Thomas and Strauss, 1997). (iii) We separate men and women. (iv) We separately examine the components of BMI (height, weight) and metabolic disease (diabetes, hypertension, cardiovascular disease). Although height is not the focus of our analysis, archaeological evidence indicates that stature was also adapted to pre-modern food supply (Pomeroy et al., 2019). We would thus expect the cross-sectional implications of the model to apply to height as well and this is indeed what we observe. There is a set point for height, which can explain, in part, Deaton's (2007) observation that there is a relatively weak association between height and income in developing countries.

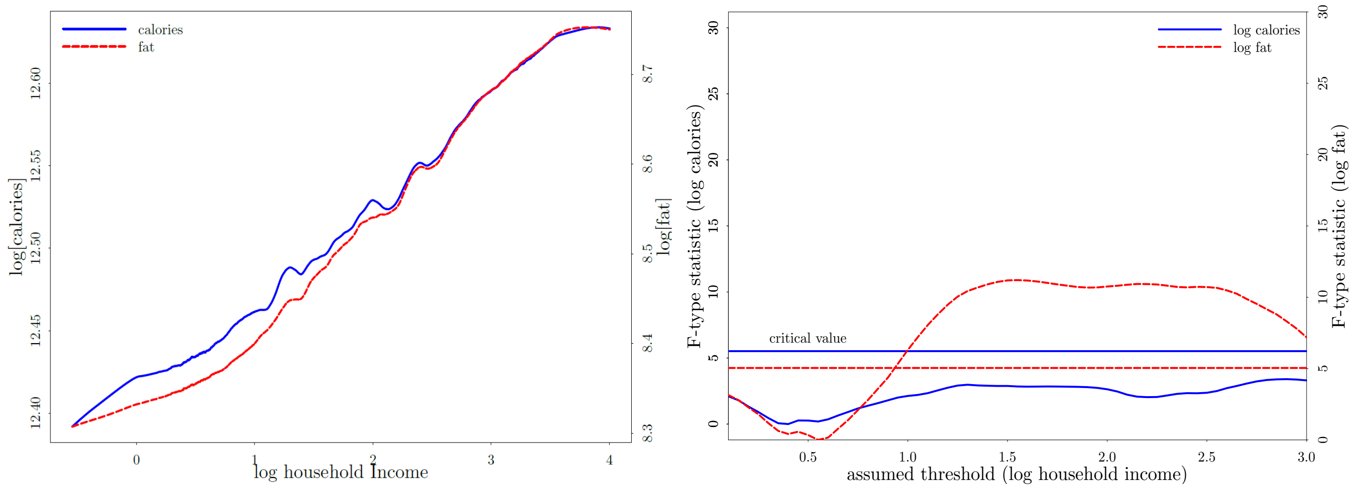
3.2 Nutrient Intake with respect to Income

The model assumes that there is a positive and continuous association between food consumption and income. To test this assumption, we report nonparametric estimates of the nutrient intake-household income association with IHDS data in Figure 3a. Nutrient intake is measured by the

cluster bootstrapped 95% confidence intervals. For ease of exposition we report cluster bootstrapped standard errors for each coefficient.

[15]The number of observations in Column 2 is substantially greater than in Column 1 for two reasons: (i) BMI, based on height and weight, can only be measured for adult 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.

Figure 3: Nutrient Intake with respect to Household Income



(a) Nonparametric association

(b) Threshold test

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

consumption of calories and fat (in grams) at the household level. The standard set of covariates, plus household composition and the number of adults engaged in physical labor are partialled out prior to estimation using Robinson's procedure. The additional covariates are included to control for energy expenditures, since energy (nutrient) intake net of these expenditures determines nutritional status. We see that there is a positive and continuous association between the intake of calories and fat and household income in Figure 3a. Moreover, Hansen's test fails to locate a slope-change at any assumed threshold in Figure 3b. In Appendix B, we examine the association between household income and expenditures on nine food categories: wheat, rice, cereals and derivative products, meat and eggs, milk and derivative products, pulses, vegetables, sugar and derivative products, and oil. Although a positive association is observed with each category, a slope discontinuity cannot be detected with any category.

4 Validating the Model

4.1 Internal Validity

The BMI-income relationship that we specify in equation (3) implies the following cross-sectional $z_t - y_t$ associations, below and above the threshold, respectively:

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

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

Closed-form expressions for the adjustment terms, $e^L(y_t)$, $e^H(y_t)$, as functions of y_t and the parameters of the model (α, $\mu_t \equiv t\mu$, and $\sigma_t^2 \equiv t\sigma^2$) are derived in equations (5) and (6). These expressions were used in Section 2 to prove Proposition 1, which describes the qualitative association between $\mathbb{E}(z_t)$ and y_t . If the parameter values can be independently obtained, then it is possible to implement a more stringent test, which is that once $e^L(y_t)$, $e^H(y_t)$ are subtracted from y_t in the equations above, the estimated income coefficient, which corresponds to the structural b parameter in the model, should be statistically indistinguishable below and above the threshold. Note that this condition will only be satisfied if the model is correctly specified. The analysis that follows thus empirically validates (i) the threshold structure we have imposed on the BMI-income relationship in equation (3), (ii) the normality assumption underlying the income generating process, and (iii) the appropriateness of the linear spline function that is used to identify a slope discontinuity in the BMI-income relationship and to locate the associated income threshold.

The value of the α parameter can be obtained directly from the estimated location of the income threshold in the cross-sectional tests. To determine the value of t, we see in Figure C.1, based on historical data over a long time span, that economic development in India commenced in the middle of the twentieth century. If each generation spans 30 years, then the grandparents of current working-age adults would have been the first generation to experience development; i.e. we are now in generation $t = 3$ of the model. To estimate the parameters of the distribution of income shocks, μ and σ^2, we require data on the income distribution over multiple time periods or generations. The distribution of pre-tax national income is available from the World Inequality Database from 1951 onwards for India (Chancel and Piketty, 2017). Assuming that each generation spans 30 years, as above, we use the (real) income distribution in 1951, 1981, and 2011 and, in particular, the change in these distributions, to estimate the μ and σ parameters.[16]

Table 2 reports coefficient estimates from a piecewise linear equation, using IHDS data, with BMI

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

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

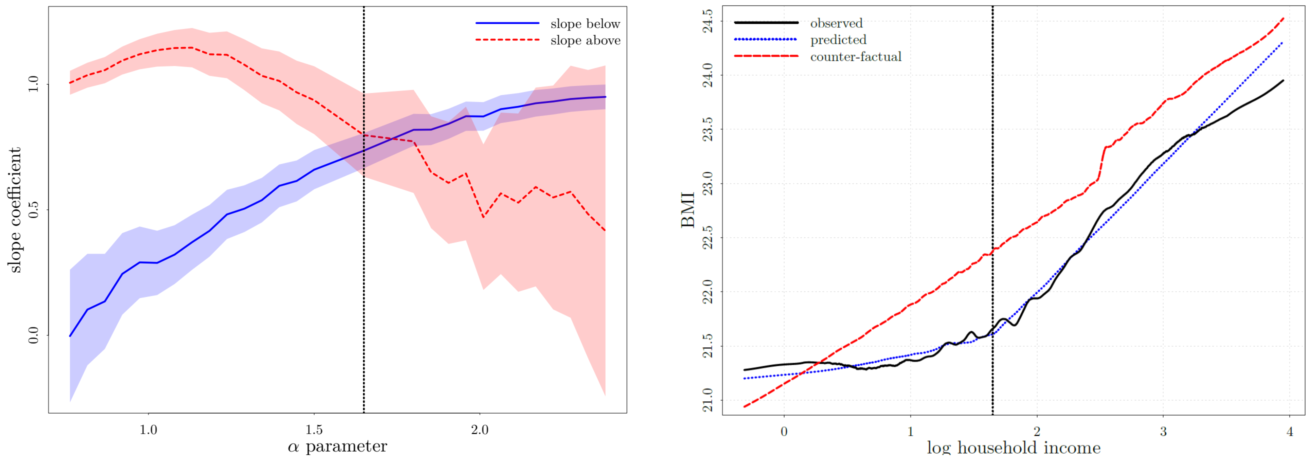
Dep. variable:	BMI	
Specification:	without adjustment (1)	with adjustment (2)
Slope below threshold (β_L)	0.223***	0.735***
	(0.048)	(0.035)
Slope above threshold (β_H)	1.140***	0.797***
	(0.035)	(0.084)
$F-$statistic ($\beta_L = \beta_H$)	234.45	0.45
	[0.000]	[0.502]
Imposed threshold	1.65	1.65
N	76,949	76,949

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

as the dependent variable. The standard covariates, in addition to household income, are included in each estimating equation and the slope-change is imposed at the income level where the threshold was previously located in the cross-sectional test. Column 1 reports 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). Column 2 reports 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 Column 1, without the adjustment, we cannot reject the null once the adjustment terms are included. Indeed, the point estimates of the slope coefficient are remarkably similar, below and above the threshold. A comparison of the point estimates indicates, in addition, that the slope without the adjustment term is less than (greater than) b, below (above) the threshold, as implied by Proposition 1.

Figure 4a examines the sensitivity of the slope coefficient estimates in Table 2, Column 2 to different values of the threshold, α, parameter. We see that the slope coefficients below (above) the specified threshold are increasing (decreasing) in α and coincide just around the value that we assign to that parameter in Table 2 (marked by the vertical lines in Figure 4a). Appendix Figure

Figure 4: Sensitivity of Slope Coefficients with respect to Parameter Values and Counter-factual Nutritional Status



(a) Sensitivity to parameter values (b) Counter-factual simulations

Source: India Human Development Survey (IHDS)
Panel (a) plots the estimated slope coefficients, below and above the threshold, with respect to the value of the α parameter. The vertical line marks the parameter value that we use for estimation in Table 2, which is based on the estimated income threshold in the cross-sectional test. Panel (b) plots observed, predicted and counter-factual BMI with respect to the logarithm of household income. The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are partialled out prior to nonparametric estimation.

C.2 repeats this exercise for the three remaining parameters of the model: μ, σ, t. As with α, we see that the slope coefficients coincide just around the values that we assign to the μ, t parameters in Table 2 (the slope coefficients are largely insensitive to the value of σ). These results indicate that three parameter values need to line up precisely to equalize the slope coefficients in Table 2, which is especially striking given that these values are derived independently from different sources: the value of α is based on the income threshold location estimated with IHDS data, the value of μ is derived from the World Inequality Database, and t is based on the changes in per capita income over many centuries reported in Appendix Figure C.1.

One benefit of the structural estimation is that it allows us to validate our modeling assumptions. An additional benefit is that it allows us to quantify the consequences of the set point for nutritional status. If the set point is irrelevant, there will be a linear relationship between BMI and household income: $\mathbb{E}(z_t) = a + by_t$. Figure 4b reports the relationship between income and (i) observed BMI, (ii) predicted BMI based on the estimated model, and (iii) counter-factual BMI in the absence of a set point. The standard set of covariates are partialled out, and the dotted vertical line in the figure marks the location of the estimated income threshold. Despite the model's parsimonious structure, and the simplifying assumptions we need to make to estimate its parameters, we see that the model fits the data very well. In our data, 20% of adults are underweight (with a BMI below 18.5). Based on the counter-factual estimates, the fraction of underweight adults would decline by 24% if the set point were absent. The observed dampening of the nutritional status-current income relationship below the threshold, which we attribute to predetermined individual-specific set points, has important consequences for adult nutritional status in India.

4.2 Diabetes with respect to BMI

The focus of the analysis thus far has been on the association between BMI and diabetes with respect to income. However, the model also has implications for the association between the risk of diabetes and BMI.[17] The IHDS includes 76,000 observations on metabolic disease, our composite measure of diabetes, and BMI over two rounds. Although the Demographic Health Survey (DHS) does not contain fine-grained income data, and thus cannot be used for the tests of the model with respect to income, the 2015-16 round of the Indian DHS includes diabetes information (with biomarkers) and BMI for as many as 770,000 adults. Nonparametric estimates of the association between diabetes and BMI with these datasets are reported in Figure 5a, after partialling out the additional covariates in the estimating equation as usual. Focusing first on the DHS data, which measure diabetes more accurately, there is no association between the risk of diabetes and BMI up to a BMI-threshold of 22.6 and a positive and significant association thereafter. These estimates, which are consistent with the model, provide statistical support for the recommendation that the overweight range in Asian populations be reduced from 25 to 23 to account for their elevated risk of diabetes at lower BMI (Deurenberg-Yap et al., 2002; Pan et al., 2004).

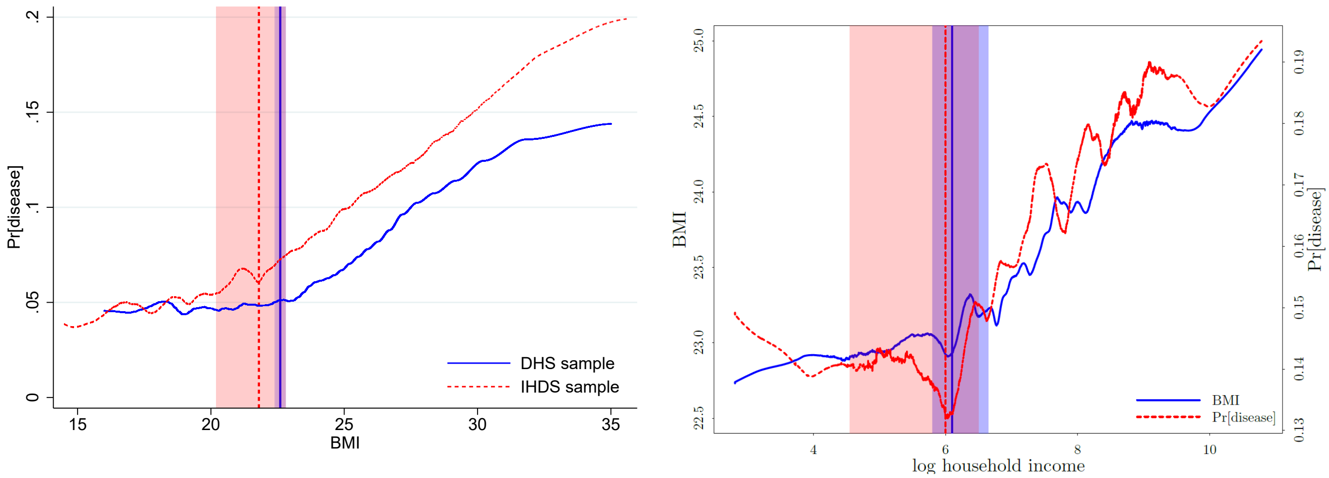
Next, we compare the diabetes measures constructed with DHS and IHDS data. As observed in Figure 5a, the alternative measures of diabetes track together across the range of BMI's. Moreover, we cannot reject the hypothesis that the threshold locations with the two data sets are statistically equal (see Appendix Table C.1). Diabetes with IHDS data is measured by a composite variable, which indicates whether a given individual has been diagnosed with diabetes, hypertension, or cardiovascular disease. In contrast, the analysis with DHS data is based on objective biomarkers for diabetes: blood sugar levels exceeding 140 mg/dL, which is the threshold that has been estimated for Indian populations with random glucose testing (Somannavar et al., 2009; Susairaj et al., 2019). The close match between these alternative measures of diabetes validates the composite measure of diabetes that we have used thus far in the analysis and which we will use with Indonesian data below.

4.3 External Validity

The core analysis focuses on the Indian population because it is simultaneously characterized by high levels of undernutrition and a high prevalence of diabetes. However, we expect the model to apply more generally. The data requirements for the cross-sectional tests are quite stringent and a search of representative data sets from other developing countries recovered just one – the Indonesia Family

[17] Other obesity indicators; e.g. waist circumference, waist-hip ratio have also been associated with diabetes. However, these indicators are highly correlated and meta-analyses indicate that the three indicators have similar associations with diabetes (Vazquez et al., 2007).

Figure 5: Nutritional Status and Metabolic Disease, India and Indonesia



(a) India (b) Indonesia

Source: Demographic and Health Survey (DHS), India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste (India) or ethnicity (Indonesia), rural area, regency (indonesia) or district (India), and survey-round are partialled out prior to nonparametric estimation.
The vertical line marks the threshold location and the shaded region demarcates the cluster bootstrapped confidence interval.

Life Survey (IFLS) – that is consistent with the IHDS and contains all the information that we need.

Figure 5b nonparametrically estimates the relationships between adult BMI, the risk 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 C.3 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 C.2, there is a weak association between household income and each outcome below the estimated threshold and a positive and significant slope-change above the threshold.

The fraction of the population that has escaped its pre-modern set point in a given country will depend on its stage in the process of development or, equivalently, the gap between current and historical (pre-modern) incomes. While roughly half the Indian population has escaped its set point, what would we expect in Indonesia? To answer this question, we compare current and historical incomes in the two countries. As documented in Appendix Table C.3, historical incomes were lower, but current incomes are higher in Indonesia. We would thus expect a larger fraction of the Indonesian population to have escaped its set point. Based on our estimates of the threshold location with respect to the income distribution, three-quarters of the Indonesian population has escaped its set point.

5 The Mechanism: Adaptation and Mismatch

Tests Based on Ancestral Income: In our model, each individual's set point is adapted to their ancestral income. Recall also that individuals with household income below the income threshold we estimate in the cross-sectional test remain at their set point. Adaptation thus implies that BMI below the threshold will be determined by ancestral income (and not by current income). The additional assumption in our model is that the risk of diabetes will be increasing in the mismatch (gap) between current income and ancestral income, but only for individuals who have escaped their set point (and thus must be above the income threshold). Our tests of these assumptions, described below, provide support for historical adaptation and the accompanying mismatch between current consumption and ancestral consumption that is central to the developmental origins of adult disease literature.[18]

Measuring Ancestral Income: If measures of ancestral income were available at the household (dynasty) level, then the preceding tests of adaptation and mismatch could be implemented directly. In practice, however, we have measures of historical income per unit of land at the village level and the district level. We map these measures into ancestral income in the following steps.

(i) *Historical income per unit of land*: Our first income measure is obtained from the British Library in London and is based on the agricultural revenue tax per acre of cultivated land that was collected by the colonial government in 1871 for each village in the modern Indian state of Tamil Nadu. The revenue tax was based on potential income, which was derived from a detailed assessment of crop suitability, soil quality, precipitation and other growing conditions. This village-level statistic can be merged with household data from the South India Community Health Study (SICHS), which we directed in one district (Vellore) of Tamil Nadu.[19] The SICHS includes a detailed survey of 5,000 representative households that is designed to be consistent with IHDS and IFLS data.

Our second, district-level, measure of historical income is based on food supply. Agriculture was

[18]Wells et al. (2016) is the only previous study that we are aware of that tests for historical adaptation and accompanying mismatch. In their analysis, ancestral consumption is measured by current height and current consumption is measured by weight. The risk of diabetes is shown to be increasing in weight, conditional on height. The assumption that height is historically determined is at odds with the results that we reported in Section 3. Moreover, their analysis essentially tells us that the risk of diabetes is increasing in current BMI (weight conditional on height, with a particular functional form). This well documented association is not informative about the mismatch hypothesis.

[19]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*.

the dominant activity in the pre-modern economy and income per unit of land would thus have been determined by crop productivity. Galor and Özak (2016) convert potential crop yields, obtained from the Food and Agriculture Organization Global Agro-Ecological Zones (FAO-GAEZ) project, 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 across countries. We use the same index, based on low-technology-rainfed agriculture, to measure pre-modern income per unit of land at the district level, except that we restrict attention to staple crops that dominated historical agricultural production in the developing countries that we consider: wheat and rice for India and rice for Indonesia. The CSI can be merged with household data from IHDS and IFLS.

(ii) *Historical income per household*: Both historical income measures derived above are defined per unit of land. We convert these measures into per household income by estimating the following equation:

$$y_t = f(Y_0) + \epsilon_t, \tag{8}$$

where y_t is current household income, obtained from SICHS, IHDS, or IFLS and Y_0 is historical income per unit of land, as described above. Equation (8) can be compared with the income equation (1) in the model: predicted income in equation (8) corresponds to ancestral household income, y_0, and the residual in that estimating equation corresponds to the income mismatch, $U_t \equiv y_t - y_0$, up to a constant.[20]

Historical income per household is, by definition, the historical income per unit of land (Y_0) divided by the number of households per unit of land (N). The latter statistic or, equivalently, the population density, would also have been an increasing (continuous) function of agricultural productivity, measured by Y_0, in the pre-modern economy (Diamond, 1997; Ashraf and Galor, 2011). It follows that the historical income per household, $f(Y_0) \equiv Y_0/N(Y_0)$, will be a continuous, possibly non-monotonic, function of Y_0. To be as flexible as possible, our measure of y_0 will be predicted household income based on a nonparametric specification of the $f(Y_0)$ function in equation (8). Appendix D provides empirical support, with Indian data, for the use of CSI as a measure of pre-modern income, as well as the procedure used to map historical income per unit of land to historical income per household.[21]

(iii) *Ancestral income*: One complication that arises, when using the village-level measure of Y_0 to construct y_0, is that there are multiple ancestral villages to choose from. Marriage in India is

[20]The residual, ϵ_t in equation (8) is mean-zero by construction, whereas U_t in equation (1) 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.

[21]If the CSI is a valid proxy for the historical income per unit of land; i.e. agricultural productivity then it should be positively associated with pre-modern population density. We verify that this is indeed the case in Appendix Figure D.1, measuring population density in 1951, when the Indian economy was just starting to grow after centuries of stagnation. We also document a continuous and non-monotonic association between our measure of y_0 and Y_0 (CSI). Appendix Figure D.2 uses binned scatter plots to (separately) describe the relationships between household income, y_t, and our measures of y_0 and U_t. These relationships are linear, matching the structure of the income equation (1) in the model. Note that failure of the separability assumption in equation (8), which allows us to construct measures of y_0 and U_t, would lead to false rejection of the model and not the converse.

patrilocal, with women often leaving their natal (birth) village when they marry. This implies that there is a single ancestral village, which is the individual's natal village, on the male line, whereas there are (possibly) many different villages from which female ancestors are drawn.[22] To construct a single historical income measure, we take advantage of the fact that spousal ancestral incomes for a given dynasty, measured by 1871 tax revenues in the respective natal villages, are highly correlated, as documented with SICHS data in Appendix Figure D.3.[23] This implies that any ancestral village can be used to measure Y_0 and we thus (to be consistent) use 1871 tax revenue in the current village of residence, for both the household head and his spouse, to construct y_0.

A second complication that arises, when using the district-level measure of Y_0 to construct y_0, is that the IHDS and IFLS include rural and urban residents. An appealing feature of the cross-sectional tests is that they do not require knowledge of y_0. This allowed us to include both rural residents and urban residents (many of whom would be recent migrants from diverse rural areas) in the analysis. For the current analysis, however, we need measures of ancestral income and our measure of y_0 will only be appropriate if a household has remained in its place of residence for many generations. The tests of adaptation and mismatch with IHDS and IFLS data are thus restricted to rural households.

Locating a threshold: To test for adaptation and mismatch, we need to locate the income thresholds. These thresholds have already been estimated, for BMI and metabolic disease, with IHDS and IFLS data. We now proceed to estimate the thresholds with SIHS data by implementing the cross-sectional tests of the model.

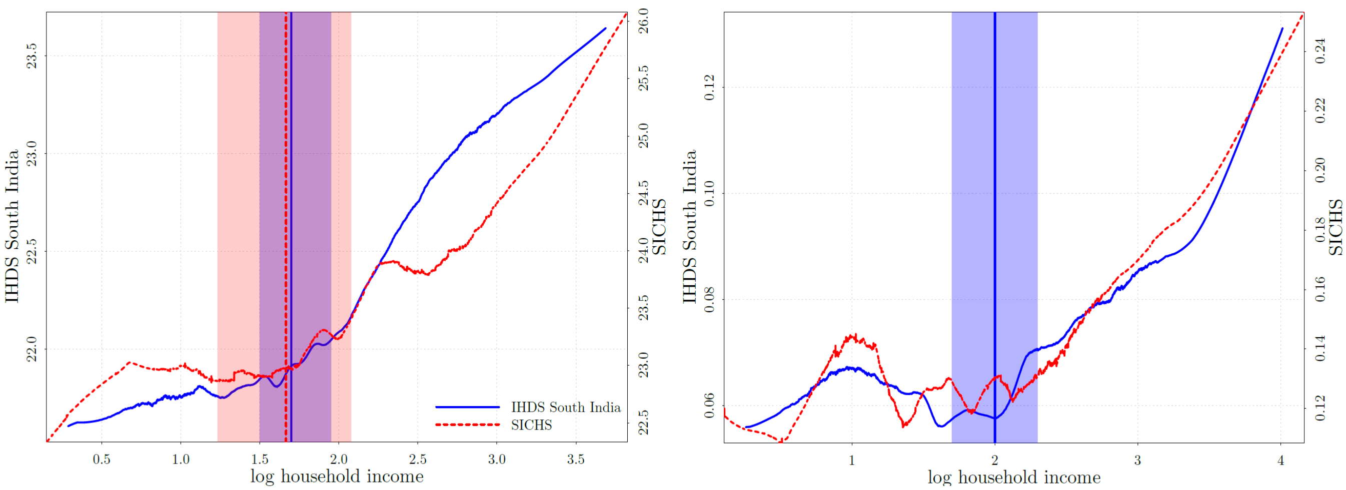
The SICHS covers a rural population of 1.1 million individuals residing in Vellore district in the South Indian state of Tamil Nadu. Two components of the SICHS are relevant for our analysis: a census of all 298,000 households residing in the study area, completed in 2014, and a detailed survey of 5,000 representative households, completed in 2016. The SICHS census collected each household's income in the preceding year. The SICHS survey collected information on marriages, as discussed above, and in addition covers all variables included in the analysis using IHDS and IFLS data. The SICHS study area was purposefully selected to be representative of rural South India, defined by the states of Tamil Nadu, Andhra Pradesh, Karnataka, and Maharashtra, with respect to socioeconomic and demographic characteristics. As a basis for comparison, we thus also implement the cross-sectional tests with IHDS data, restricting the sample to the southern states. As seen in Figures 6a and 6b, the estimated BMI-income and metabolic disease-income associations track very closely with SICHS and IHDS South India data, across the income distribution.[24] The vertical lines

[22]Epigenetic inheritance was traditionally assumed to occur along the female line; i.e. via the mother, although recent evidence indicates that paternal traits can also be transmitted epigenetically (Jablonka and Raz, 2009; Lind and Spagopoulou, 2018). We allow for both possibilities.

[23]The strong correlation in ancestral incomes that we document, separately for the household head and his wife and for their parents, 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.

[24]BMI and the risk of metabolic disease are systematically higher with SICHS data relative to IHDS South India

Figure 6: Nutritional Status and Metabolic Disease with respect to Income (IHDS and SICHS)



(a) BMI (b) Metabolic disease

Source: India Human Development Survey (IHDS), South India Community Health Study (SICHS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, and (for IHDS) rural area, district and survey-round are partialled out prior to nonparametric estimation.
The vertical lines mark the estimated threshold location and the shaded areas demarcate the corresponding 95% confidence intervals.

mark the spot where Hansen's test (shown in Appendix Figure D.4) locates an income threshold, with the shaded area demarcating the associated 95% confidence interval. The threshold locations with 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. This is because the sample size is much smaller with SICHS data and the threshold is more difficult to estimate with a binary outcome.[25] We will thus test for adaptation, but not mismatch, with SICHS data.

Evidence of Adaptation and Mismatch: Table 3 reports the relationship between BMI and both ancestral income, y_0, and current income, y_t, below and above the estimated income threshold with each data set. 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, where relevant, are included in the estimating equations. Columns 1-2 report results with SICHS data, Columns 3-4 with IHDS data and Columns 5-6 with IFLS data. The consistent finding with all three data sets is that ancestral income has a positive and significant effect on BMI below the

data (this can be observed by comparing the range of the Y-axes in Figure 6). 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.

[25]For those outcomes for which thresholds can be located in Figure 6, the piecewise linear equation estimates at the estimated thresholds are reported in Appendix Table D.1. In line with previous results, we cannot reject the hypothesis with South Indian (IHDS) data that the thresholds with BMI and metabolic disease as outcomes are located at the same income level.

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

Dependent variable:	BMI					
Country:	India				Indonesia	
Survey:	SICHS		IHDS		IFLS	
Sample:	below (1)	above (2)	below (3)	above (4)	below (5)	above (6)
Ancestral income	0.334***	0.170	0.899***	0.165	1.059***	0.464
	(0.124)	(0.150)	(0.243)	(0.283)	(0.254)	(0.337)
Current income	0.012	0.834***	0.185***	0.852***	-0.048	0.591***
	(0.190)	(0.119)	(0.040)	(0.047)	(0.119)	(0.064)
Threshold location	1.69	1.69	1.65	1.65	6.1	6.1
Dependent var. mean	23.033	23.755	20.482	21.851	22.317	23.021
N	1810	3844	27,164	20,296	3,182	10,610

Source: South India Community Health Study (SICHS), India Human Development Survey (IHDS), Indonesia Family Life Survey (IFLS)
In columns (1)–(2), the covariates include age (linear, quadratic, and cubic terms), gender, and caste group. In columns (3)–(4), the covariates include age (linear, quadratic, and cubic terms), gender, caste group, state, and survey-round. In columns (5)–(6), the covariates include age (linear, quadratic, and cubic terms), gender, ethnicity, regency, and survey-round. The rural-urban dummy is omitted because the sample is restricted to rural households. Standard errors are bootstrapped and clustered at the village level for the SICHS sample, and at the primary sampling unit for both the IHDS and the IFLS samples
Significance levels: $*$ for 10%, $**$ for 5%, $***$ for 1%, based on cluster bootstrapped confidence intervals.

threshold (where households are at their adapted set point) but not above it. In contrast, current income has a positive and significant effect on BMI above the threshold but not below it (with one exception).

Current income in equation (8) 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. Table 4 reports the relationship between the risk of metabolic disease and (separately) each income component, below and above the estimated income threshold (τ). Results with Indian (IHDS) data are presented in Columns 1-2 and with Indonesian (IFLS) data in Columns 3-4. As specified in the model, 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, the ancestral income coefficients in Columns 2 and 4 are insignificant, with one exception, and jointly insignificant in both columns, once again in line with the model.

Alternative Explanations: We complete the analysis by considering alternative explanations for the observed cross-sectional associations, with BMI and the risk of diabetes as outcomes. These explanations are generated by relaxing key assumptions of the model, which we list below:

1. *There is a continuous association between nutrient intake, net of energy expenditures, and*

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 × $\mathbf{1}\{\text{current income} > \tau\}$	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)
In columns (1)–(2), the covariates include age (linear, quadratic, and cubic terms), gender, caste group, state, and survey-round. In columns (3)–(4), the covariates include age (linear, quadratic, and cubic terms), gender, ethnicity, regency, and survey-round. The rural-urban dummy is omitted because 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.

household income. In support of this assumption, we did not observe any discontinuity between various measures of nutrient intake and household income in Section 3. Moreover, we accounted for energy expenditures by including the number of household members engaged in manual labor in the estimating equations. 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 and we are thus likely incorporating the major expenditures in the analysis. Nevertheless, if there is a discontinuous increase in the net nutrient intake at a particular income level for some unspecified reason, then the observed discontinuous association between BMI and income could be obtained without a set point.

2. The income distribution is log normal. Our test of internal validity in Section 4 provides empirical support for the distributional assumption in our model. Nevertheless, we consider the possibility that poverty traps, which shift the income distribution, could independently generate the results that we obtain. When poverty traps are caused by credit constraints and non-convexities, as in Galor and Zeira (1993) and Banerjee and Newman (1993), households with sufficiently low initial income will remain permanently at that level. This will change the distribution of current income, but will not give rise to a discontinuity in the cross-sectional BMI-income association. Poverty trap models generated by undernutrition; e.g. Dasgupta and Ray (1986) could, however, generate a discontinuity because of the feedback from BMI to income below a threshold.

3. Adaptation is based on ancestral consumption. As discussed in the Introduction, the observation that migrants to advanced economies retain an elevated risk of diabetes for multiple generations, despite the fact that their nutritional status converges very quickly to that of the native population, indicates that the underlying (ancestral) set point must be stable. Nevertheless, we consider the possibility that adaptation is based, instead, on conditions in *utero* in the current generation, as assumed by early contributions to the developmental origins of adult disease literature; e.g. Barker (1995). In the context of our model, this implies that the set point is determined by income in the previous generation rather than the initial generation. As discussed in Section 2, this alternative specification of the set point, in which it drifts over time, will also generate the cross-sectional implications of the model.

The discussion on alternative explanations thus far has focussed on the BMI-income association. We now turn our attention to the risk of diabetes. As implied by our model with a fixed set point, this risk increases discontinuously with respect to BMI, at a particular threshold, in Section 4. For the explanations based on unobserved changes in diet and lifestyles or poverty traps to generate this cross-sectional association, the discontinuous increase in BMI at a particular income level that we document would need to trigger an accompanying discontinuous increase in the risk of diabetes. There is no obvious reason why this should be the case, especially as there is no association between the risk of diabetes and BMI below the BMI threshold.

As we have noted, an alternative model in which the set point shifts across generations would also generate the cross-sectional associations that we document. However, and perhaps most conclusively, ancestral (historical) income plays no role in any of the alternative explanations that we consider. In particular, we are unaware of any mechanism, other than our own model, that can explain the specific patterns that we document in this section with respect to ancestral income; with BMI and the risk of diabetes as outcomes, above and below the estimated income threshold.

6 Conclusion

This research examines the health consequences of an individual-specific set point for BMI that is adapted to (low) food supply in the pre-modern economy, but which subsequently fails to adjust to economic development. Our structural estimates of the BMI-income relationship and accompanying counter-factual simulations indicate that the fraction of underweight adults in India, who comprise 20% of the population, would decline by 24% in the absence of a set point. At the same time, half the adult population who remain at their set point are protected from diabetes. While the health consequences of the set point are currently ambiguous, what is the prognosis for the future in India and other developing countries? If the experience of migrants from South Asia to advanced economies is any indication, then we would expect that these populations will rapidly escape their pre-modern set points in the coming decades and improve their nutritional status. However, the elevated risk of diabetes at relatively low BMI's could persist for multiple generations. The public health strategy would need to shift in that case, from the conventional focus on prevention, to screening and treatment. Our analysis, which documents a discontinuous increase in the risk of diabetes at a BMI below 23 in India, indicates that much of the adult population may need to be screened for this condition.

While the cost of screening may be greater than currently envisaged, the flip-side of this finding is that many individuals detected with diabetes will have relatively low BMI's. A natural question to ask is how these lean diabetics, who do not necessarily have unhealthy lifestyles, should be treated. As Taylor and Holman (2015) note, weight loss is the focus when treating obese diabetics, but is not usually considered for those with normal BMI. The recent medical literature has, perhaps for this reason, shifted focus away from BMI towards other risk factors that are seen to be correlated with diabetes in developing-country populations, such as low lean mass, low insulin secretion, and ectopic fat deposition (Pomeroy et al., 2019; Narayan and Kanaya, 2020). If the objective is to reverse diabetes, however, then our analysis indicates that this objective would be better served by correcting the fundamental cause of the problem, which is failure of an underlying homeostatic system and resulting energy imbalance. Evidence from a weight-loss program in the U.K. indicates that diabetes can be successfully reversed (Lean et al., 2018). In this group of mostly overweight and obese patients, the average weight loss was as much as 10 kg. In a developing-country population, we expect that the BMI threshold below which diabetes is reversed will be associated with the pre-modern set point. Many (lean) diabetics, who would have recently escaped their set point, will have BMI's that are not far from their threshold. This suggests a promising behavioral approach to diabetes control in such a population, involving relatively little weight loss, that we plan to explore in future research.

References

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

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

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

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

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

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

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.

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

Dalgaard, C.-J., J. B. Madsen, and H. Strulik (2021): "Physiological Constraints and the Transition to Growth: Implications for Comparative Development," *Journal of Economic Growth*, 26, 241–289.

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

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

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

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

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

Franck, R., O. Galor, O. Moav, and Ö. Özak (2022): "The shadow of the neolithic revolution on life expectancy: A double-edged sword," Tech. rep., National Bureau of Economic Research.

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

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

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

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik instruments: What, when, why, and how," *American Economic Review*, 110, 2586–2624.

Greif, A. (1994): "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies," *Journal of Political Economy*, 102, 912–950.

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

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.

Kültz, D. (2020): "Defining Biological Stress and Stress Responses based on Principles of Physics," *Journal of Experimental Zoology Part A: Ecological and Integrative Physiology*, 333, 350–358.

Lean, M. E., W. S. Leslie, A. C. Barnes, N. Brosnahan, G. Thom, L. McCombie, C. Peters, S. Zhyzhneuskaya, A. Al-Mrabeh, K. G. Hollingsworth, et al. (2018): "Primary care-led weight management for remission of type 2 diabetes (DiRECT): an open-label, cluster-randomised trial," *The Lancet*, 391, 541–551.

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

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.

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

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

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

Narayan, K. and A. M. Kanaya (2020): "Why are South Asians prone to type 2 diabetes? A hypothesis based on underexplored pathways," *Diabetologia*, 63, 1103–1109.

Neel, J. V. (1962): "Diabetes Mellitus: A "Thrifty" Genotype Rendered Detrimental by "Progress"?" *American Journal of Human Genetics*, 14, 353.

Newey, W. K., J. L. Powell, and F. Vella (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565–603.

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.

Petrie, J. R., T. J. Guzik, and R. M. Touyz (2018): "Diabetes, hypertension, and cardiovascular disease: clinical insights and vascular mechanisms," *Canadian Journal of Cardiology*, 34, 575–584.

Pomeroy, E., V. Mushrif-Tripathy, T. J. Cole, J. C. Wells, and J. T. Stock (2019): "Ancient origins of low lean mass among South Asians and implications for modern type 2 diabetes susceptibility," *Scientific Reports*, 9, 1–12.

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.

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.

Somannavar, S., A. Ganesan, M. Deepa, M. Datta, and V. Mohan (2009): "Random capillary blood glucose cut points for diabetes and pre-diabetes derived from community-based opportunistic screening in India," *Diabetes Care*, 32, 641–643.

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.

Stebbing, A. (2009): "Interpreting 'Dose-Response' Curves using Homeodynamic Data: with an Improved Explanation for Hormesis," *Dose-Response*, 7, 221–233.

Susairaj, P., C. Snehalatha, A. Raghavan, A. Nanditha, R. Vinitha, K. Satheesh, D. G. Johnston, N. J. Wareham, and A. Ramachandran (2019): "Cut-Off value of random blood glucose among Asian Indians for preliminary screening of persons with prediabetes and undetected type 2 diabetes defined by the glycosylated haemoglobin criteria," *Journal of Diabetes and Clinical Research*, 1, 53.

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

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

Thomas, D. and J. Strauss (1997): "Health and wages: Evidence on men and women in urban Brazil," *Journal of Econometrics*, 77, 159–185.

Vazquez, G., S. Duval, D. R. Jacobs Jr, and K. Silventoinen (2007): "Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis," *Epidemiologic Reviews*, 29, 115–128.

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.

Online Appendix

A The Model

A.1 Proofs of Propositions

Proof of Proposition 1: At any given level of income $y_t \leq \alpha$,

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

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

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

In the absence of any prior knowledge about the distribution of pre-modern income, we make the simplifying assumption that initial income is uniformly distributed; i.e. $f(\cdot)$ is constant. It follows that

$$\mathbb{E}(z_t|y_t) = \int_{-\infty}^{y_t} [a + b(y_t - U_t)] \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \,\mathrm{d}\, U_t = a + b\left(y_t - e^L(y_t)\right) \tag{A.1}$$

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

Since the uniform distribution has bounded support, the lower range of integration should extend to $y_t - \overline{y}_0$, where $\overline{y}_0$ is the right support of the initial income distribution. The advantage of extending the range to $-\infty$ is that we can solve the model analytically and derive a closed-form expression for $e^L(y_t)$, with simulations reported below in Appendix A.2 indicating that this approximation has no discernable effect on predicted BMI (and the risk of metabolic disease) except in the right tail of the y_t distribution.

Making the same approximation as above, at any given level of income $y_t > \alpha$:

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

We next derive closed-form expressions for $e^L(y_t)$, $e^H(y_t)$, which are given as

$$e^L(y_t) = \mu_t - \sigma_t \Lambda\left(\frac{y_t - \mu_t}{\sigma_t}\right) \tag{A.3}$$

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

where $\Lambda(\cdot)$ is the Inverse Mill's ratio. Focusing on the numerator of the $e^L(y_t)$ expression in (A.1) 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 (A.1) gives us the closed-form expression for $e^L(y_t)$ in equation (A.3). The corresponding expression for $e^H(y_t)$ in equation (A.4) 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 (A.3) in equation (A.1) and differentiate with respect to y_t. Given the properties of the inverse Mill's ratio, the slope at $y_t \leqslant \alpha$ is given as

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

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 (A.4), 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 $\frac{\mathrm{d}\,\mathbb{E}(z_t|y_t)}{\mathrm{d}\,y_t} > b$ for $y_t > \alpha$.

Note, from equations (A.3) and (A.4), that $e^L(y_t) = e^H(y_t)$ at $y_t = \alpha$, and thus, from equations (A.1) and (A.2), there is no level discontinuity at the threshold. To prove that there is, nevertheless, a slope discontinuity at the threshold, $y_t = \alpha$, we need to show that

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

From equations (A.1) and (A.2), 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 (A.3) and (A.4), 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 (A.3), $e^L(y_t) = \frac{\mathcal{L}(v_t)}{\Phi(v_t;0,1)}$, where $\mathcal{L}(v_t) = \mu_t\Phi(v_t;0,1) - \sigma_t\phi(v_t;0,1)$. From equation (A.4), $e^H(y_t) = \frac{\mathcal{L}(\overline{v})}{\Phi(v_t;0,1)}$ 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: The relationship between the probability of diabetes, $P(D_t)$, and income is given as

$$P(D_t) = \begin{cases} \gamma_1 & \text{if} \quad U_t \leqslant \alpha \\ \gamma_1 + \gamma_2(y_t - y_0) & \text{if} \quad U_t > \alpha \end{cases} \tag{A.5}$$

Hence, for any given $y_t \leqslant \alpha$, making the same analytical approximation and distributional assumptions as above:

$$P(D_t|y_t) = \int_{-\infty}^{y_t} \gamma_1 \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t = \gamma_1 \tag{A.6}$$

and for any given $y_t > \alpha$,

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

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

Following the same steps that we used to derive the expression for $e^L(y_t)$ in (A.3), we can write for any given $y_t > \alpha$,

$$P(D_t|y_t) = \gamma_1 + \gamma_2 \left[\mu_t - \sigma_t \Lambda\left(\frac{y_t - \mu_t}{\sigma_t}\right) - \frac{\mu_t \Phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right) - \sigma_t \phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right)}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)}\right] \tag{A.7}$$

From equation (A.6), $\frac{\mathrm{d}\,P(D_t|y_t)}{\mathrm{d}\,y_t} = 0$ for $y_t \leqslant \alpha$, and from equation (A.7), $\frac{\mathrm{d}\,P(D_t|y_t)}{\mathrm{d}\,y_t} > 0$ for $y_t > \alpha$ because $\Lambda'(\cdot) < 0$ and $\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)$ is increasing in y_t. This also establishes that there is a slope discontinuity at $y_t = \alpha$. Further, substituting $y_t = \alpha$ in equation (A.7) eliminates the term inside square brackets, implying that there is no level discontinuity at $y_t = \alpha$.

A.2 Placing an upper bound on y_0

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

$$\begin{aligned}\mathbb{E}(z_t|y_t) &= \int_{y_t-\overline{y}_0}^{y_t} [a + b(y_t - U_t)] \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\,U_t \\ &= a + by_t - b\int_{y_t-\overline{y}_0}^{y_t} U_t \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\,U_t \\ &= a + b(y_t - \overline{e}^L(y_t)) \qquad \text{(A.8)}\end{aligned}$$

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:

$$\overline{e}^L(y_t) = \mu_t - \sigma_t \frac{\left[\phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right) - \phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right]}{\left[\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right) - \Phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right]} \qquad \text{(A.9)}$$

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 given $y_t \in [\alpha, \overline{y}_0 + \alpha]$ is thus described by the following expression:

$$\begin{aligned}\mathbb{E}\left(z_t|y_t\right) &= \int_{y_t-\overline{y}_0}^{\alpha} [a + b(y_t - U_t)] \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\,U_t \\ &\quad + \int_{\alpha}^{y_t} [a + by_t] \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\,U_t \\ &= a + by_t - b\int_{y_t-\overline{y}_0}^{\alpha} U_t \frac{\phi(U_t;\mu_t,\sigma_t^2)}{\Phi(y_t;\mu_t,\sigma_t^2) - \Phi(y_t - \overline{y}_0;\mu_t,\sigma_t^2)} \,\mathrm{d}\,U_t \\ &= a + b(y_t - \overline{e}^H(y_t)) \qquad \text{(A.10)}\end{aligned}$$

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

$$\overline{e}^H(y_t) = \frac{\mu_t\left[\Phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right) - \Phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right] - \sigma_t\left[\phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right) - \phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)\right]}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right) - \Phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)} \qquad \text{(A.11)}$$

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

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

Diabetes-income relationship: For any given $y_t \leqslant \alpha$,

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

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

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

Solving the integral,

$$P(D_t|y_t) = \gamma_1 + \gamma_2 \frac{\mu_t\left[\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)-\Phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right)\right]-\sigma_t\left[\phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)-\phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right)\right]}{\Phi\left(\frac{y_t-\mu_t}{\sigma_t};0,1\right)-\Phi\left(\frac{y_t-\overline{y}_0-\mu_t}{\sigma_t};0,1\right)} \tag{A.12}$$

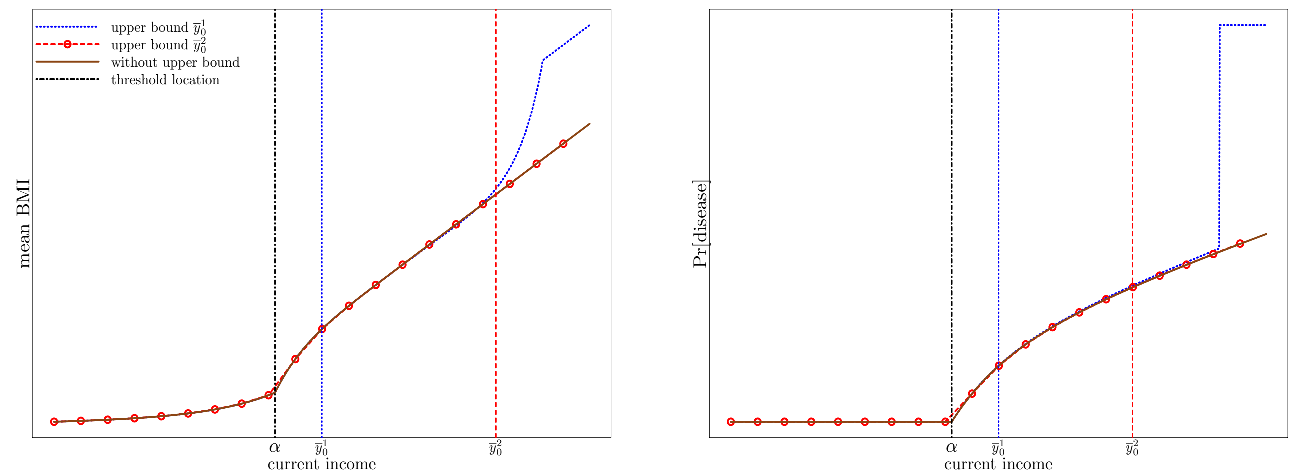
For any given $y_t > \overline{y}_0 + \alpha$, as everyone has escaped their set point, we can write,

$$\begin{aligned} P(D_t|y_t) &= \int_{\alpha}^{\infty} (\gamma_1+\gamma_2 U_t) \frac{\phi(U_t;\mu_t,\sigma_t^2)}{1-\Phi(\alpha;\mu_t,\sigma_t^2)}\,\mathrm{d}\,U_t \\ &= \gamma_1 + \gamma_2\left[\mu_t + \sigma_t \frac{\phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right)}{1-\Phi\left(\frac{\alpha-\mu_t}{\sigma_t};0,1\right)}\right] \end{aligned} \tag{A.13}$$

which is independent of y_t.

Although analytical results can no longer be derived as in Propositions 1 and 2, expressions (A.8), (A.9), (A.11), (A.10), (A.12) and (A.13) can be used to simulate the relationship between current income and both BMI and the probability of metabolic disease. We use the actual income from the IHDS and the estimates of μ_t, σ_t from the structural estimation exercise for the simulation. The left panel in Figure A1 plots the relationship between BMI and current income, with and without the upper bound on y_0. The right panel plots the corresponding relationships between metabolic disease and income. For the upper bound we choose two values of $\overline{y}_0$. The first value $\overline{y}_0^1$, marked by the blue dotted vertical line, is close to the threshold α whereas the second value $\overline{y}_0^2$, marked by the red dashed line, is further to the right. The simulated BMI-income and metabolic disease-income relationships track together, almost exactly, with the three specifications, except in the right tail

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



(a) BMI

(b) Metabolic disease

of the income distribution where we observe a second discontinuity with $\overline{y}_0^1$. In our data, we do not observe a second discontinuity, at a high income level, with either BMI or the risk of metabolic disease as outcomes.

A.3 Alternative Specifications for the Set Point

Set point determined by ancestral and current income

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

$$z_t = \begin{cases} a + b[r_t y_0 + (1 - r_t) y_t] & \text{if } \; y_t - [r_t y_0 + (1 - r_t) y_t] \leqslant \widetilde{\alpha} \\ a + b y_t & \text{if } \; y_t - [r_t y_0 + (1 - r_t) y_t] > \widetilde{\alpha} \end{cases} \tag{A.14}$$

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

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

where $e^L(y_t)$ is defined in (A.3). Similarly, for any given $y_t > \frac{\widetilde{\alpha}}{r_t}$, we can write

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

where the expression for $\widetilde{e}^H(y_t)$ is the same as in equation (A.4) when α is replaced by $\frac{\widetilde{\alpha}}{r_t}$.

Set point determined by previous generation income

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

$$z_t = \begin{cases} a + by_{t-1} & \text{if } \; y_t - y_{t-1} \leqslant \overline{\alpha} \\ a + by_t & \text{if } \; y_t - y_{t-1} > \overline{\alpha} \end{cases} \tag{A.15}$$

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

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

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

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

B Cross-Sectional Tests

B.1 Robinson Procedure

Consider the following semi-parametric estimating equation:

$$y_i = f(Z_i) + X_i\beta + \epsilon_i$$

where y_i is an outcome such as BMI or the risk of diabetes for individual i, Z_i is their household income, X_i is the standard vector of covariates that needs to be partialled out prior to nonparametric estimation of the $y_i - Z_i$ association and ϵ_i is a mean-zero disturbance term. The Robinson Robinson (1988) procedure is implemented as follows:

Step 1. Separately regress y_i and each element of the X_i vector nonparametrically on Z_i.

Step 2. Regress the residuals from the first equation, $\hat{\xi}_y$, on the residuals from the other equations, $\hat{\xi}_X$, using a linear specification without a constant term to estimate $\hat{\beta}$.

Step 3. Nonparametrically regress $y_i - (X_i - \overline{X})\hat{\beta}$ on Z_i, where $\overline{X}$ is the sample mean of each element in the vector of covariates.

B.2 Robustness Tests

We complete the cross-sectional tests with IHDS data by verifying the robustness of the results in the following ways:

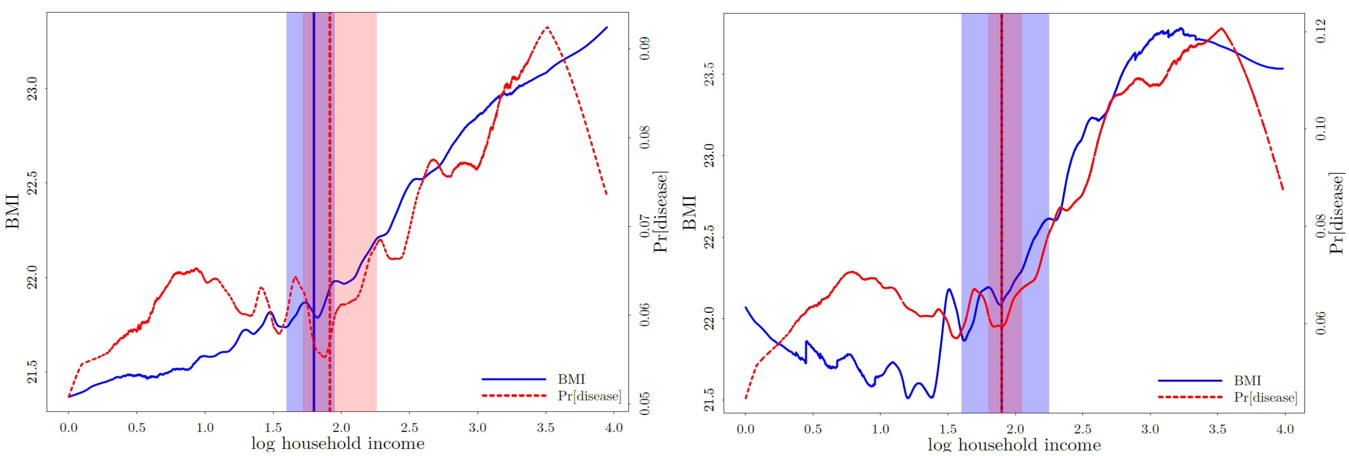
(i) We include measures of household composition as additional covariates in the estimating equation in Figure B1a and Table B1, Columns 1-2.

(ii) We construct a nonparametric shift-share instrument for household income in Figure B1b and Table B1, Columns 3-4. The construction of the instrumental variable and tests of its validity are also reported.

(iii) We separate men and women in Figure B3 and Table B2.

(iv) We separately examine the components of BMI (height, weight) and metabolic disease (diabetes, hypertension, cardiovascular disease) in Figure B4 and Table B3.

Figure B1: Nutritional Status and Metabolic Disease with respect to Household Income (additional covariates and nonparametric shift-share instrument)



(a) Additional covariates

(b) Nonparametric shift-share instrument

Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are included in panels (a) and (b).
For panel (a), 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. For panel (b), additional covariates include land ownership, its interaction with the rural dummy, and the residual (linear, quadratic, and cubic terms) from the first-stage nonparametric regression, as described below. Covariates are partialled out prior to nonparametric estimation.
The vertical lines mark the estimated threshold location and the shaded areas demarcate the corresponding confidence intervals.

Table B1: Piecewise Linear Equation Estimates (additional covariates and nonparametric shift-share instrument)

Robustness exercise:	additional covariates		nonparametric shift-share instrument	
Dependent Variable:	BMI	metabolic disease	BMI	metabolic disease
	(1)	(2)	(3)	(4)
Baseline slope (β_1)	0.281**	0.003	0.192	0.005
	(0.052)	(0.002)	(0.411)	(0.006)
Slope change (β_2)	0.516**	0.014**	1.522*	0.044**
	(0.069)	(0.003)	(0.827)	(0.012)
Threshold location (τ)	1.80	1.95	1.95	1.90
	[1.60, 1.95]	[1.75, 2.30]	[1.60, 2.25]	[1.80, 2.05]
Threshold test $p-$value	0.000	0.000	0.022	0.000
Mean of dependent variable	22.002	0.074	22.275	0.073
N	76,949	148,928	73,708	138,782

Source: India Human Development Survey (IHDS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are included in the estimating equation.
For columns 1-2, 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.
For columns 3-4, additional covariates include land ownership, its interaction with the rural dummy and the residual (linear, quadratic, and cubic terms) from the first-stage nonparametric regression, as described below.
Bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses.
For all columns except for column (3), cluster bootstrapped 95% confidence bands for the threshold location are in brackets. For column (3), 90% confidence bands are provided.
$**$, $*$ significant at 5%, 10%, based on cluster bootstrapped confidence intervals.

Instrumental variable estimation:

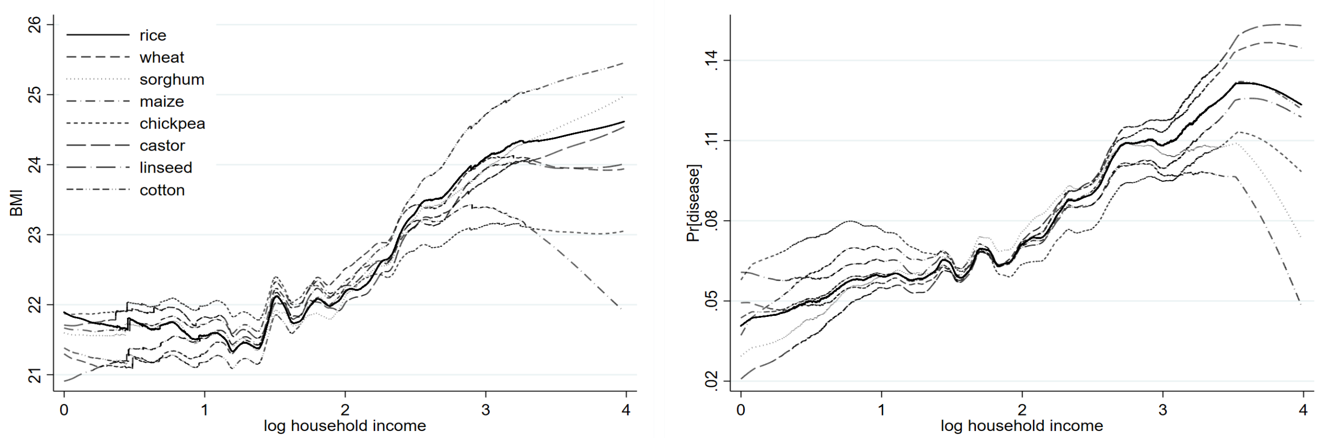
Step 1: We use ICRISAT District Level Data (DLD) for India to construct the growth in output value, at the national level over the 1966-2015 period, for each of the following crops: rice, wheat, sorghum, maize, chickpea, castor, linseed and cotton. We then construct a district-level measure of the growth in value by taking a weighted average of the growth of each crop, where the weight is the acreage allocated to that crop in 1965 divided by total cultivated acreage in that year. District-level growth is interacted with the rural dummy and land owned by the household (obtained from IHDS) to construct the shift-share instrument.

Step 2: We regress household income nonparametrically on the shift-share instrument, after partialling out district effects, the rural dummy, land ownership and the interaction of land ownership with the rural dummy, using the Robinson procedure. The coefficient on the shift-share instrument in a corresponding linear regression has a t-statistic of 3.75 (F=13.67), indicating that the instrument has sufficient statistical power.

Step 3: Following Newey et al. (1999), we include a polynomial (cubic) function of the residuals from the preceding step, land ownership, and its interaction with the rural dummy as additional covariates, which are partialled out together with the standard set of controls, when we nonparametrically estimate the BMI-income and metabolic disease-income relationships.

Step 4: Following Goldsmith-Pinkham et al. (2020), we validate the nonparametric instrumental variable estimates, reported in Figure B1b and Table B1, Columns 3-4 by using acreage shares of individual crops, rather than the growth in value, to construct crop-specific instruments. As Goldsmith-Pinkham et al. note, the estimates will be similar with each crop if the shift-share instrument is valid, and this is indeed what we observe below.

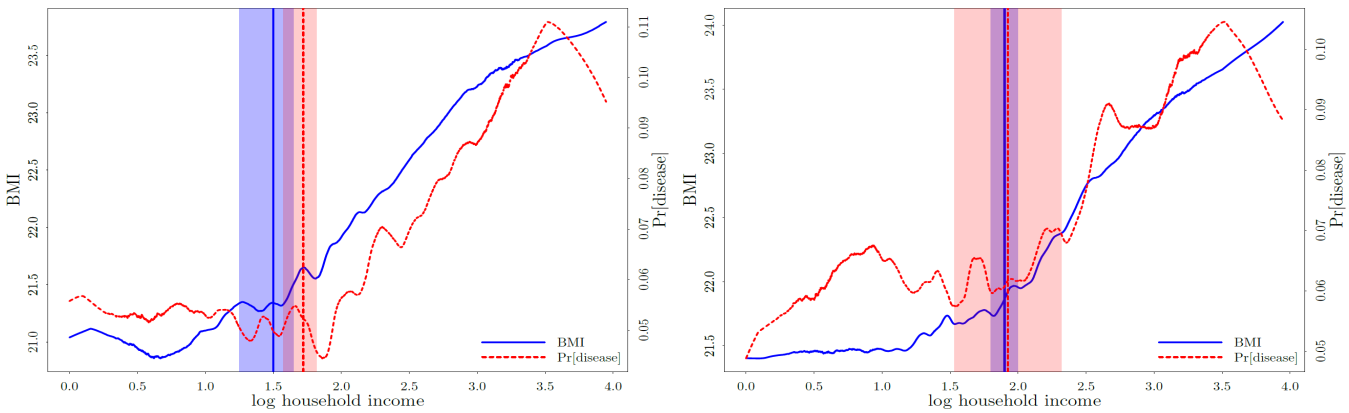
Figure B2: Nutritional Status and Metabolic Disease with respect to Household Income (instrument based on individual crop shares)



(a) BMI (b) Metabolic disease

Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round, together with land ownership, its interaction with the rural dummy, and the residual from the first-stage nonparametric regression (linear, quadratic, and cubic terms) are partialled out prior to nonparametric estimation.

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



(a) Men (b) Women

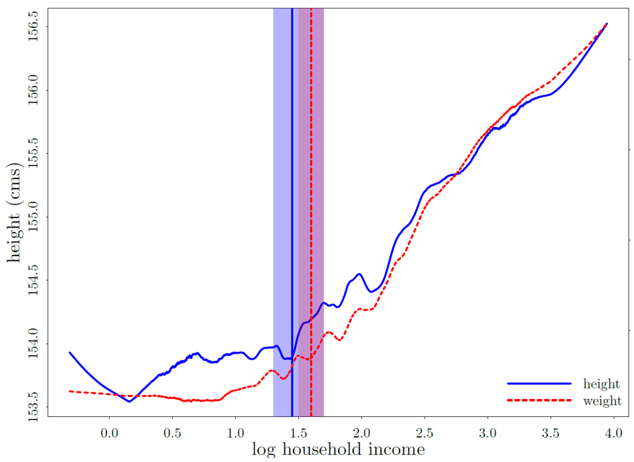
Source: India Human Development Survey (IHDS)
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for caste group, rural area, district, and survey-round are partialled out prior to nonparametric estimation.
The vertical lines mark the estimated threshold location and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

Table B2: Piecewise Linear Equation Estimates (separately for men and women)

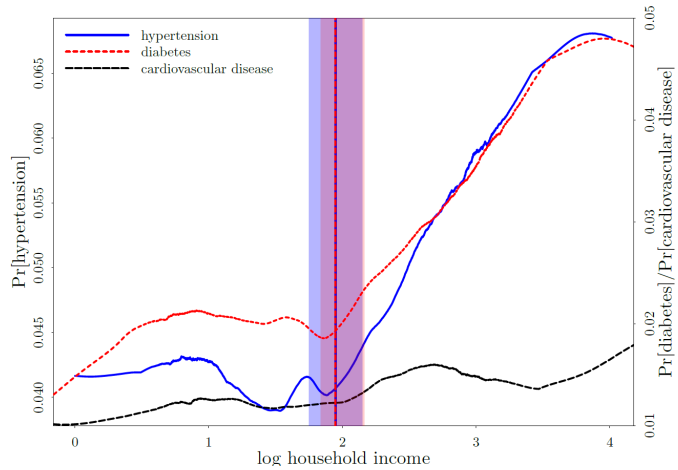
Dependent variable:	BMI		metabolic disease	
Sample:	men (1)	women (2)	men (3)	women (4)
Baseline slope (β_1)	0.342**	0.225**	−0.001	0.005
	(0.104)	(0.062)	(0.003)	(0.003)
Slope change (β_2)	0.877**	0.980**	0.038**	0.018**
	(0.112)	(0.079)	(0.004)	(0.005)
Threshold location (τ)	1.50	1.75	1.90	1.95
	[1.25,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

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

Figure B4: Alternative Nutritional Status Measures and Metabolic Diseases (separately) with respect to Household Income



(a) Height and weight



(b) Hypertension,diabetes, and cardiovascular disease

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

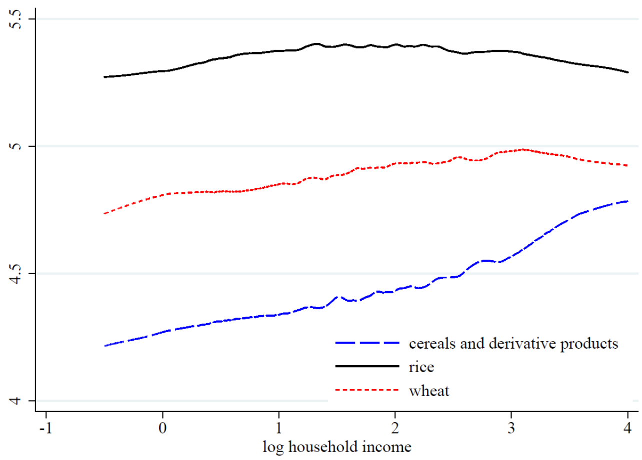
Table B3: Piecewise Linear Equation Estimates (alternative nutritional status measures, hypertension, diabetes, and cardiovascular disease)

Measures:	alternative nutrition measure		metabolic disease		
Dependent variable:	height (1)	weight (2)	hypertension (3)	diabetes (4)	cardiovascular disease (5)
Baseline slope (β_1)	0.191	0.656**	0.001	0.001	0.001**
	(0.135)	(0.150)	(0.002)	(0.001)	(0.0005)
Slope change (β_2)	0.836**	2.863**	0.018**	0.017**	
	(0.144)	(0.174)	(0.003)	(0.002)	
Threshold location (τ)	1.45	1.60	1.95	1.95	
	[1.30,1.70]	[1.50,1.70]	[1.75, 2.15]	[1.85, 2.15]	
Threshold test $p-$value	0.000	0.000	0.000	0.000	
Mean of dependent variable	154.483	52.578	0.049	0.027	0.014
N	77,000	77,143	147,858	147,684	147,626

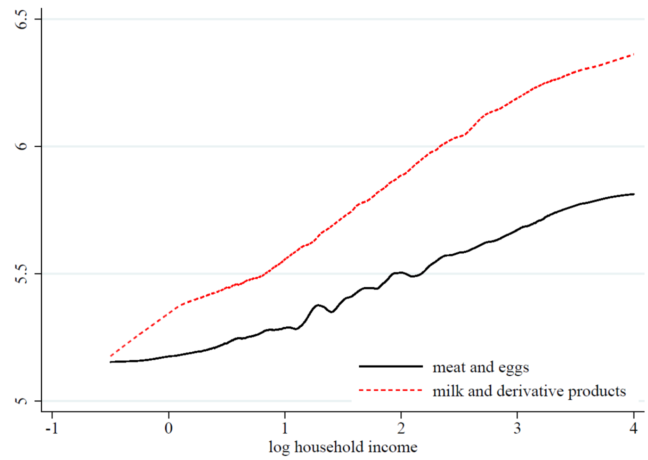
Source: India Human Development Survey (IHDS)
Logarithm of household income is the independent variable.
The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round are included in the estimating equation.
For columns (1)-(4), bootstrapped standard errors, clustered at the level of the primary sampling unit, are in parentheses. For column (5), standard errors are clustered at the primary sampling unit level.
Cluster bootstrapped 95% confidence bands for the threshold location are in brackets.
$**$ significant at 5%, based on cluster bootstrapped confidence intervals.

B.3 Nutrient Intake

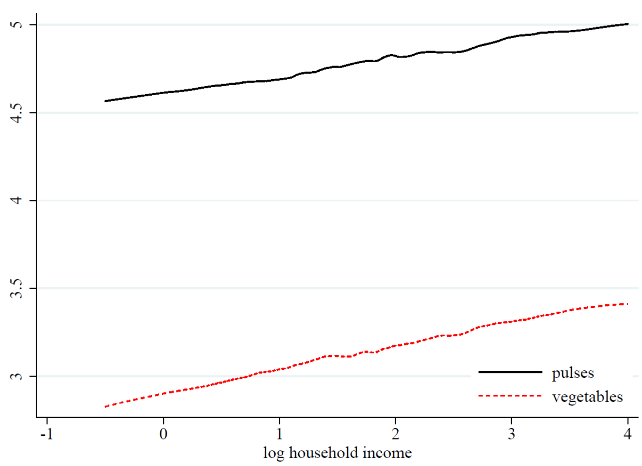
Figure B5: Expenditure on different food categories with respect to household income



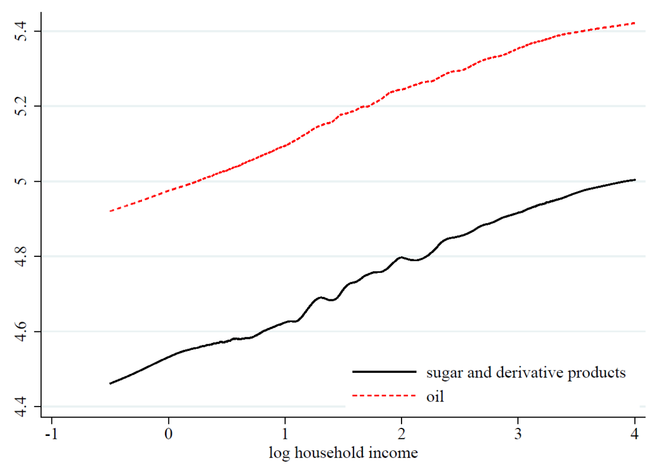
(a) Cereals, rice and wheat



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



(c) Pulses and vegetables



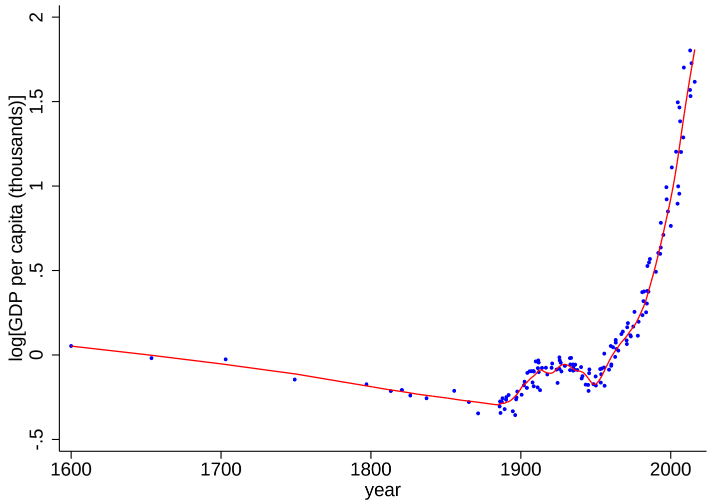
(d) Oil and sugar and derivative products

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

C Validating the Model

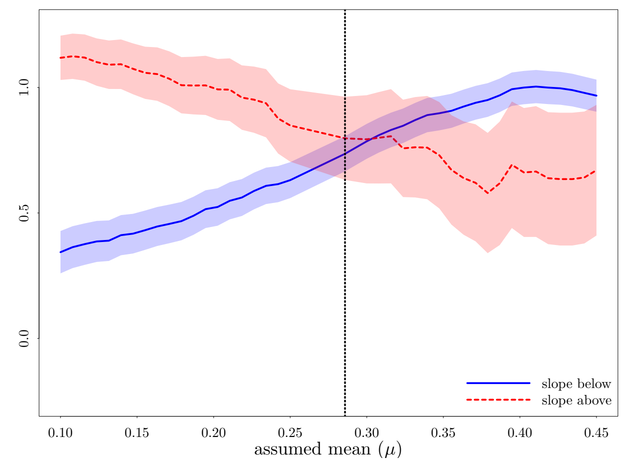
C.1 Internal Validity

Figure C.1: Evolution of Income in India

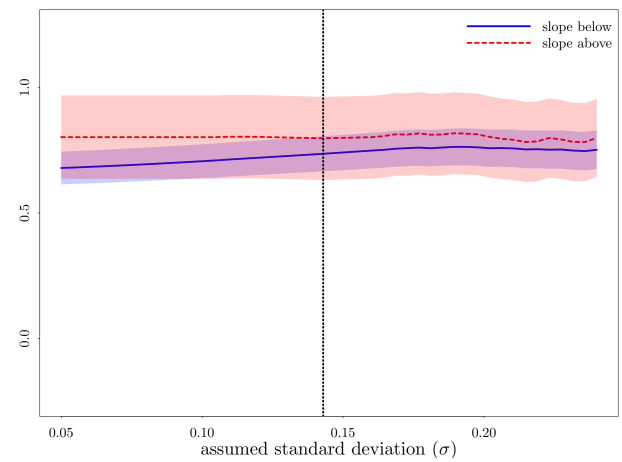


Source: Maddison Project Database (2018)
GDP per capita is measured in 2011 US dollars.

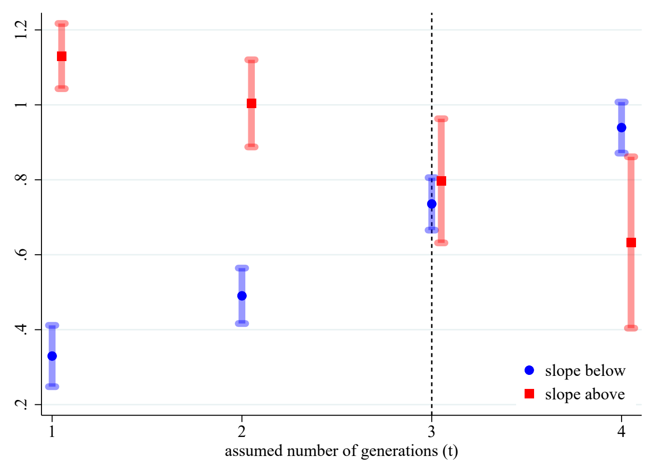
Figure C.2: Sensitivity of Slope Coefficients with respect to Parameter Values



(a) Mean of the income shock



(b) Standard deviation of the income shock



(c) Number of generations

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

C.2 Diabetes-BMI Association

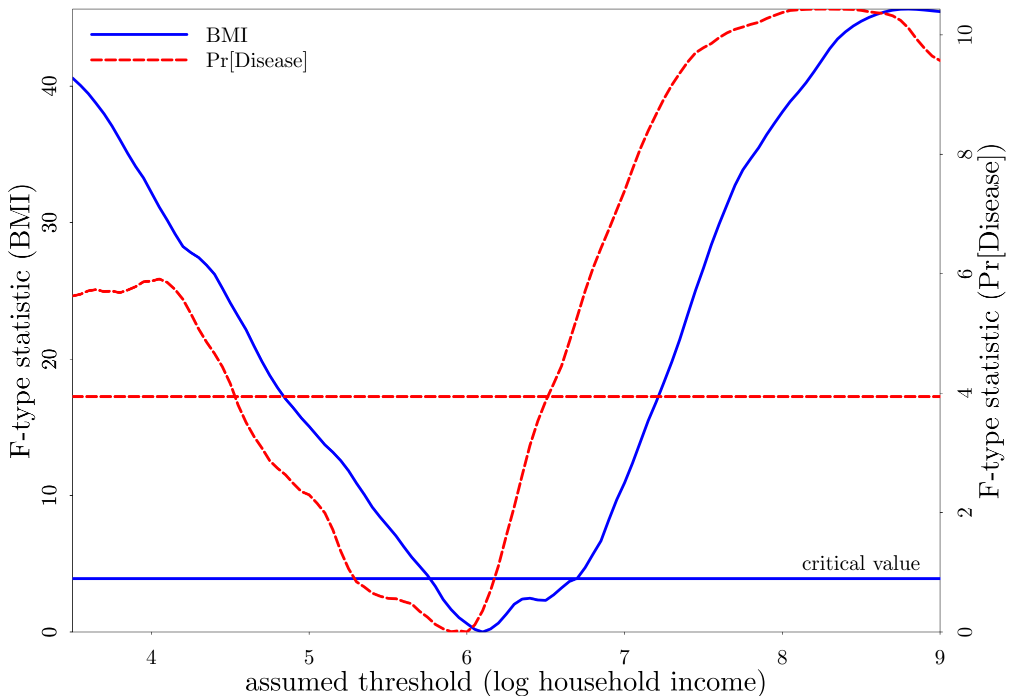
Table C.1: Piecewise Linear Equation Estimates (reported metabolic disease and measured diabetes)

Dependent variable:	metabolic disease (IHDS) (1)	diabetes (DHS) (2)
Baseline slope (β_1)	0.003**	0.0002
	(0.001)	(0.0002)
Slope change (β_2)	0.006**	0.010**
	(0.001)	(0.0002)
Threshold location (τ)	21.80	22.60
	[20.20, 22.80]	[22.40, 22.60]
Threshold test $p-$value	0.000	0.000
Mean of dependent variable	0.066	0.057
N	76,103	777,533

Source: India Human Development Survey (IHDS), Demographic Health Survey (DHS) 2015-16
BMI is the independent variable. The standard set of covariates: age (linear, quadratic, and cubic terms) and dummies for gender, caste group, rural area, district, and survey-round for IHDS 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.

C.3 External Validity

Figure C.3: Nutritional Status and Metabolic Disease with respect to Income (Indonesia)



Source: Indonesia Family Life Survey (IFLS)
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, ethnicity, rural area, regency, and survey-round are included in the estimating equation at each assumed threshold for the threshold test. Cluster bootstrapped 5% critical values are used to bound the threshold location.

Table C.2: Piecewise Linear Equation Estimates (Indonesia)

Dependent variable:	BMI (1)	metabolic disease (2)
Slope below (β_L)	0.067	-0.001
	(0.065)	(0.010)
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
N	30,812	24,788

Source: Indonesia Family Life Survey (IFLS)
Metabolic disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.
Logarithm of household income is the independent variable.
The following covariates: age (linear, quadratic, and cubic terms) and dummies for gender, ethnicity, rural area, regency, and survey-round are included in the estimating equation.
Bootstrapped standard errors, clustered at the sub-regency level 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.

Comparing per capita incomes in India and Indonesia: We use height for the 1900 birth cohort to measure historical income and GDP per capita in 1960 (the earliest available year) and 2010 to measure subsequent changes with economic development. As Deaton Deaton (2007) notes, genes are important determinants of individual height (and nutritional status more generally) but cannot explain variation across populations. This measure of historical income is also not inconsistent with our model; recall that nutritional status, which we specify with respect to BMI but which also includes stature, is assumed to be increasing continuously with contemporaneous income in the pre-modern economy. This relationship is only associated with a discontinuity in subsequent periods (generations) with economic development on account of the persistent set point. Based on these measures of income, historical per capita incomes were lower in Indonesia than in India, but these cross-country differences have now reversed.

Table C.3: Historical height and income, and current income for India and Indonesia

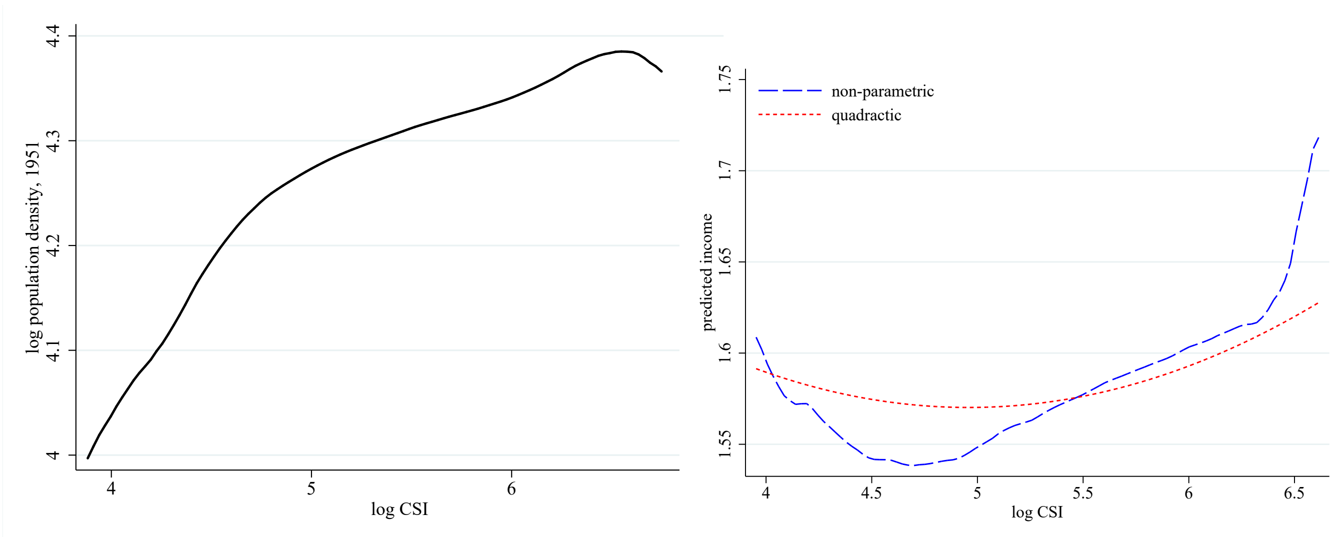
	Height 1900 (cms) (1)	GDPPC (1960) (2)	GDPPC (2010) (3)
India	155	1044	4386
Indonesia	150	1298	7394

Source: NCD-RisC for the height of the 1900 birth cohort and Penn World Table 9.0 for GDP per capita (GDPPC) in 1960 and 2010

D The Mechanism: Adaptation and Mismatch

D.1 Historical Income

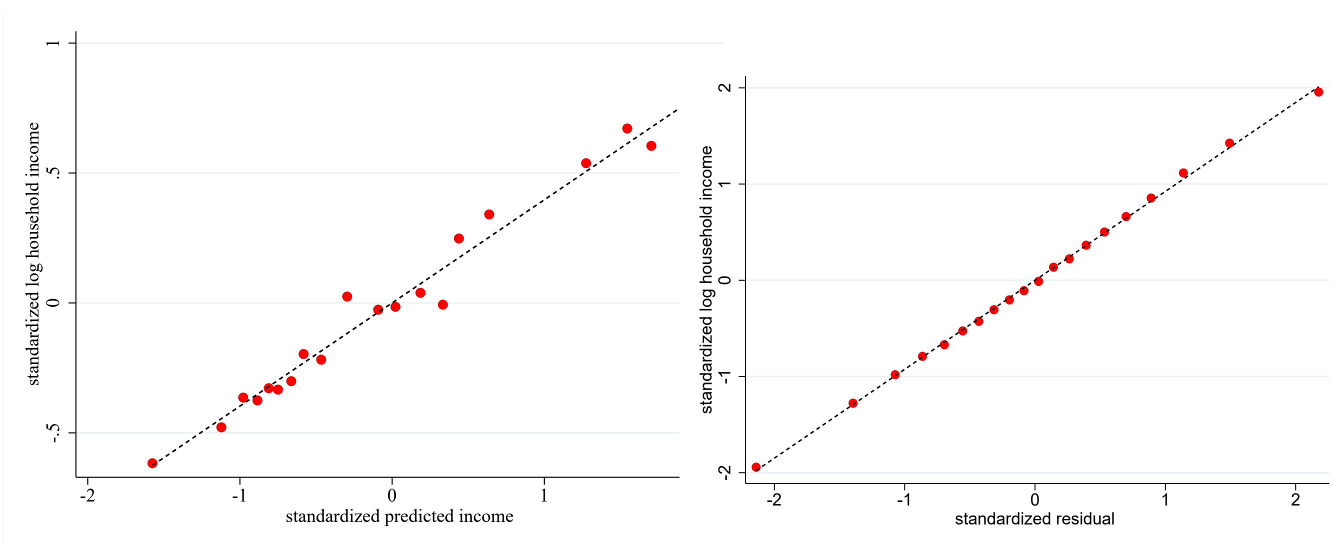
Figure D.1: Population Density and Predicted Household Income with respect to Caloric Suitability Index (CSI)



(a) Population density (b) Predicted household income

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

Figure D.2: Household Income with respect to Predicted Income and the Residual

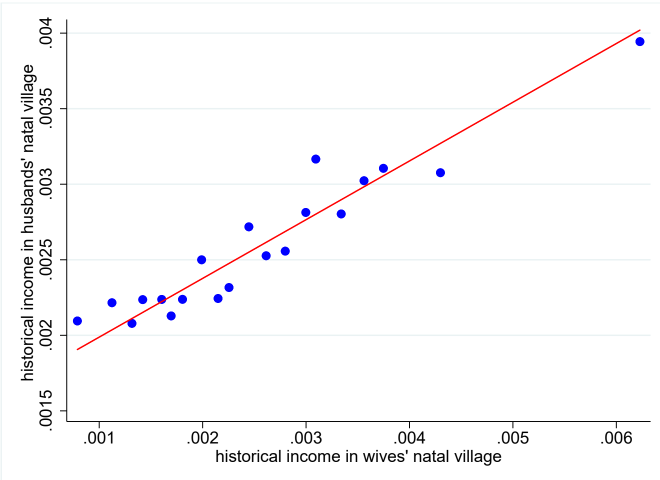


(a) Predicted income (b) Residual

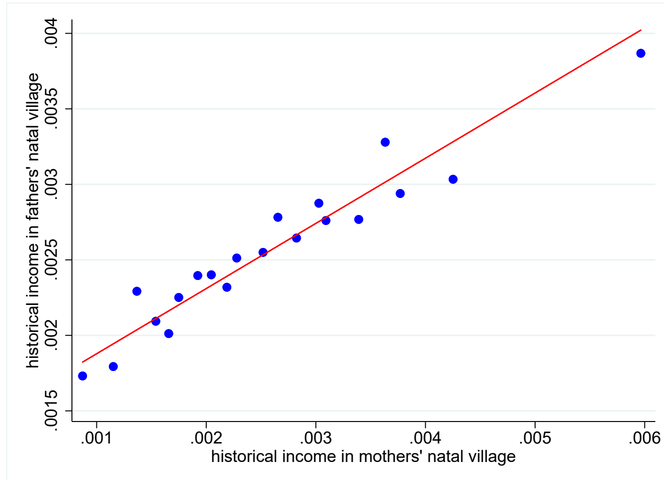
Source: FAO-GAEZ dataset, India Human Development Survey (IHDS)
This figure reports binned scatter plots describing the relationship between current household income, y_t, and (i) predicted income, which is our measure of y_0, and (ii) the residual from the estimating equation, which is our measure of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

D.2 Ancestral Income

Figure D.3: Assortative Matching on Historical Income



(a) Current generation

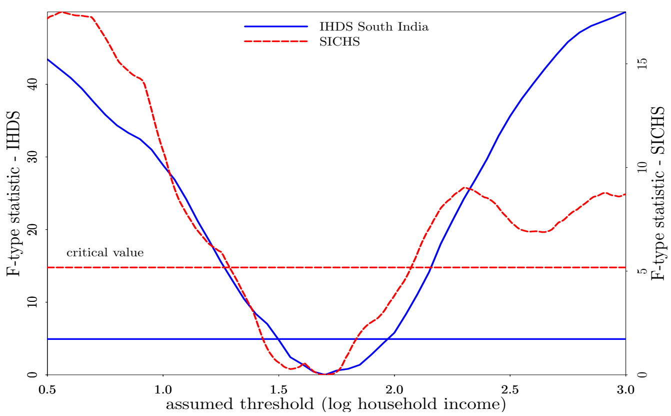


(b) Parents' generation

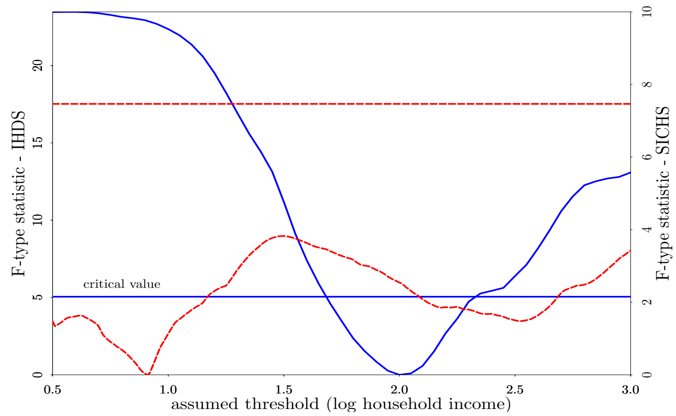
Source: South India Community Health Study (SICHS)
Historical income is measured by tax revenue per acre of cultivated land in 1871 in the individual's natal village. The number of bins in the binned scatter plot is set equal to 20.

D.3 Locating a Threshold

Figure D.4: Threshold Tests - Nutritional Status and Metabolic Disease (IHDS and SICHS)



(a) BMI



(b) metabolic disease

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

Table D.1: Piecewise Linear Equation Estimates – Nutritional Status and Metabolic Disease (South India)

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

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