ECONOMIC DEVELOPMENT, THE NUTRITION TRAP AND METABOLIC DISEASE

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

May 27, 2019

Abstract

This research provides a unified explanation for: (i) the persistence of malnutrition and (ii) the increased prevalence of metabolic disease (diabetes, hypertension, cardiovascular disease) among normal weight individuals with economic development. Our theory is based on an epigenetically determined set point for BMI or bodyweight, which is adapted to conditions of scarcity in the pre-modern economy, but which subsequently fails to adjust to rapid economic change. During the process of development, some individuals thus remain at their low-BMI set point, despite the increase in their consumption, while others who have escaped the nutrition trap (but are not necessarily overweight) are at increased risk of metabolic disease. The theory is validated with microdata from India, Indonesia, and Ghana and can simultaneously explain inter-regional (Asia-Africa) differences in nutritional status and the prevalence of diabetes.

*Luke: Pennsylvania State University, nkl10@psu.edu. Munshi: University of Cambridge, km619@cam.ac.uk. Oomen: Christian Medical College, anuoommen@cmcvellore.ac.in. Singh: Bank of Lithuania and Kaunas University of Technology, ssingh@lb.lt. We are grateful to Anne Ferguson-Smith, Nita Forouhi, Seema Jayachandran, Venkat Narayan, Nigel Unwin, and numerous seminar participants for their constructive comments. Research support from the National Institutes of Health through grant R01-HD046940, Cambridge-INET, the Keynes Fund and the Newton Trust at the University of Cambridge, and the Agence Nationale de la Recherche (ANR) under the EUR Project ANR-17-EURE-0010 is gratefully acknowledged. We are responsible for any errors that may remain. The views expressed here are those of the authors and do not necessarily reflect the position of the Bank of Lithuania.
Two stylized facts motivate our research: First, the relatively weak relationship between nutritional status and income in developing countries; both across countries (Deaton, 2007) and within country over time (Deaton and Drèze, 2009). Second, the increased prevalence of metabolic disease among normal weight individuals with economic development (Narayan, 2016). Take India, for example, a country which has received much attention in the nutrition and health literatures. India has experienced substantial economic growth and sharp declines in the prevalence of poverty in recent decades. Nevertheless, a surprisingly large fraction of its population remains malnourished, while, simultaneously, the incidence of diabetes and related metabolic disorders (hypertension and cardiovascular disease) has increased dramatically. There is an erroneous belief that the rapid increase in diabetes in countries like India is due to increased obesity; e.g. Diamond (2011). While obesity may well end up being the primary contributor to diabetes, once these countries have developed, we will see below that a relatively small fraction of the Indian population is currently obese and that the risk of diabetes starts to increase at a BMI level that is well within the normal range.

Our unified explanation for why malnutrition stubbornly persists, even as metabolic diseases emerge is inspired by an economics literature on poverty traps (Dasgupta and Ray, 1986; Galor and Zeira, 1993; Banerjee and Newman, 1993) but is based on a biologically determined “nutrition trap”. A growing biomedical literature posits that there exists a predetermined set point for each individual’s body weight or BMI, with metabolic and hormonal adjustments defending the set point against variations in energy intake (food consumption) over the life-course (Prentice et al., 1992; Müller et al., 2010; Farooqi, 2014). In general, the set point is determined by genetics, the environment in early life, and epigenetics. We focus on the epigenetic mechanism, in which genes interact with the environment to create adaptive physical traits (phenotypes) because these traits can persist after the conditions that gave rise to them have ceased to be relevant. As seen below, this combination of initial adaptation and subsequent persistence is a key ingredient in our analysis.

Developing countries were characterized by low and fluctuating food supply for centuries, with economic conditions only improving relatively recently. Given the physiological cost of fluctuating body weight, and given low levels of consumption on average, the set point would have been optimally set at a low BMI or, equivalently, the population would have been characterized by a lean body type (Narayan, 2016). With economic development, consumption will increase, but the individual’s body will defend its inherited BMI set point against these increases in consumption, just as her ancestors’ bodies adjusted to fluctuations in food supply in the pre-modern economy. We posit that once the mismatch between current and ancestral income (consumption) crosses a threshold, the body will no longer be able to defend the set point and the individual’s BMI will start to track more closely with current income. However, because the fine metabolic balance that maintained the set point has now been disrupted, the metabolic load will exceed the metabolic capacity and there will be in tandem an increased risk of metabolic disease (see Wells et al. (2016) and Taylor and Holman (2015) for related arguments).

\[1\] This does not imply that body weight stays constant over an individual’s lifetime. It will adjust with age and fluctuate with variations in food supply. However, these fluctuations will be dampened and the body weight will ultimately return to the age-adjusted set point (unless it deviates past a threshold).
Based on the preceding discussion, the population will be partitioned into two distinct groups during the process of development: Individuals in the first group remain at their BMI set point, despite the increase in their consumption, and are responsible (in part) for the weak observed relationship between nutritional status and current income in developing countries. Individuals in the second group, who have escaped the nutrition trap, but are not necessarily overweight, are the primary contributors to the increased incidence of metabolic disease that accompanies economic development. This partition of the population is only temporary. While our framework has an important feature in common with poverty-trap models – the presence of a threshold – the difference is that in the long-run the pre-modern set point will cease to be relevant.\(^2\) The biological friction that we incorporate is perhaps more closely related to models of institutional adaptation and persistence. For example, Munshi and Rosenzweig (2006) describe how caste-based networks, which emerged in response to labor market imperfections in the pre-modern economy, generated a dynamic inefficiency when they failed to adjust to subsequent structural change in the Indian economy. In the current analysis, the human body adapts to the environment in the pre-modern economy, which was stable for many centuries, but then fails to adjust to rapid economic development, resulting in the persistence of malnutrition and the emergence of metabolic disease.

Our theory has many features in common with Barker’s (1995) influential fetal origins hypothesis.\(^3\) This hypothesis states that the fetus will grow more slowly under conditions of scarcity (Hales and Barker, 1992; Gluckman and Hanson, 2004). However, if subsequent conditions diverge unexpectedly from the initial conditions, then there will be an increased risk of metabolic disease. The robust finding from many studies that have tested this hypothesis is that a combination of low birth weight, generated accidentally by famine or some other one-off adverse event, and high adult BMI puts individuals at greatest risk of metabolic disease. In our framework, the biological adaptation to conditions of scarcity in the pre-modern economy over many generations is measured by the BMI set point. The shock to the stability of the biological system is economic development. Nevertheless, we continue to expect that the risk of disease will be increasing in the mismatch between current BMI and the BMI set point.

It has been previously hypothesized in the biomedical literature that the body can adapt to pre-modern conditions and that the subsequent mismatch between these initial conditions and current conditions can give rise to metabolic disease (Gluckman and Hanson, 2004; Narayan, 2016; Wells et al., 2016). By specifying that the initial condition is measured by an epigenetically determined BMI set point, we are able to rigorously test this hypothesis and to tie together two seemingly unrelated literatures: the well-established literature on the developmental origins of health and disease and the emerging literature on the nutritional status-income relationship in developing countries. The challenge that we face is that the set point is not directly observed. To provide empirical support for our theory, we begin by developing a

\(^2\)Once economic development commences, the bulk of the population will escape its pre-modern set point within a few generations. It has been conjectured that repeated exposure to the same environment over multiple generations increases the stability of epigenetic inheritance (Radford, 2018). Nevertheless, we would expect that epigenetic traits acquired over a period of centuries (unlike genetic adaptation which takes tens of thousands of years) will cease to be inherited relatively quickly; within the space of a few generations, once the conditions that gave rise to them have ceased to be relevant.

\(^3\)In recent years, a rich literature in economics has advanced the fetal origins hypothesis (see Almond and Currie (2011), and Almond et al. (2018), for comprehensive overviews). The economics literature has largely focussed on non-health outcomes, which are outside the scope of this research.
model of nutrition and health in which the existence of a predetermined BMI set point for each individual is taken as given. This set point is determined by the income (consumption) of the individual’s ancestors in the pre-modern economy, which is period 0 in the model. Starting from period 1, which denotes the onset of economic development, each dynasty receives an income shock in each period or generation, which can be positive or negative, but is positive on average. With the accumulation of income shocks over time, dynasties gradually drift away from their initial income level. However, as long as current income remains sufficiently close to ancestral income, a dynasty’s members will continue to remain at their BMI set point. This will only change when the gap between current income and ancestral income crosses a threshold; BMI will now be determined by current income and there will be a discrete increase in nutritional status. Accompanying this escape from the nutrition trap will be an increased risk of metabolic disease.

If data on income, BMI, and metabolic disease were available for each dynasty over many generations, then we could test the structural relationships specified above 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 the threshold) with an accompanying increase in the incidence of metabolic disease. Given that information on ancestral income is unavailable in standard data sets, what we do, instead, is to derive the cross-sectional relationships between current income and both nutritional status and the risk of metabolic disease. This requires us to place additional structure on the distribution of income shocks in each generation; following standard convention, we assume that these shocks are log-normally distributed. Given this distributional assumption, we can prove the following result: (i) Although nutritional status is increasing in current income at all income levels, there is a discontinuous increase in the slope of the relationship at a particular income threshold. Households below the threshold remain at their set point, which is determined by ancestral income. This is why there is a relatively weak relationship between nutritional status and current income for them. (ii) The risk of metabolic disease is constant below the threshold, and increasing in income above the threshold.

We use nationally representative household data from the India Human Development Survey (IHDS) to test the implications of our model. Our main result is that the nutritional status-income relationship (separately for children and adults) and the disease-income relationship are precisely as predicted by the model. The presence of a slope discontinuity, which we detect statistically using Hansen’s (2017) threshold test, is indicative of a set point. The weak relationship between nutritional status and household income below the estimated threshold, which is located close to the median income level in the population, can explain (in part) the first stylized fact. The steep increase in the probability of metabolic disease with income above the same threshold, which corresponds to a BMI that is well within the normal range, helps explain the second stylized fact.

Although our model and the accompanying empirical tests provide an internally consistent and unified explanation for both stylized facts, we must still account for other independent determinants of nutritional status and metabolic disease. The estimating equations include a rich set of covariates, which account for the effect of son preference on nutritional status, as documented by Jayachandran and Pande (2017),

---

4 Nutritional status is measured by height-for-age for children and BMI for adults in the empirical analysis. Alternative measures, based on weight-for-age for children and height for adults, deliver similar results.
as well as spatial variation in food tastes (Atkin, 2013, 2016) and the disease environment (Duh and Spears, 2017; Spears et al., 2013; Dandona et al., 2017). In addition, we use IHDS data to examine the possibility that our results are being driven by variation in two proximate determinants of nutritional status discussed by Deaton (2007) – nutrient intake and children’s illness – with respect to income. In contrast with the discontinuous income effect that we estimate with nutritional status and the probability of metabolic disease as outcomes, there is a positive and continuous relationship between nutrient intake and household income and a negative and continuous relationship between children’s illness and household income. This is confirmed by Hansen’s threshold test, which fails to detect a slope discontinuity with these outcomes.\(^5\)

To provide additional independent support for the presence of a BMI set point, we proceed to directly estimate the structural relationships that underlie the model. Recall that nutritional status is determined by ancestral income, which determines the set point, below the income threshold and by current income above the threshold. We cannot test these relationships with standard data sets, including the IHDS, because they do not provide information on ancestral income. However, this is possible with unique data we have recently collected as part of the South India Community Health Study (SICHS), which covers a population of 1.1 million individuals in rural Tamil Nadu. We estimate the relationship between adult BMI and household income, separately above and below the estimated income threshold, including current income and ancestral income in the estimating equation. The striking result is that ancestral income alone matters below the threshold, whereas current income alone matters above it.

The presence of a set point is evidently not unique to India. To assess the external validity of our theory, we test the model with data from the Indonesia Family Life Survey (IFLS) and the Ghana Socioeconomic Panel Survey (GSPS). The results with the IFLS match exactly with what we obtain with Indian data. In contrast, there is a positive and continuous relationship between household income and nutritional status – for children and adults – with the Ghanaian data (information on metabolic disease is not available in the GSPS). These differences can be explained by the fact that while a set point may be present in other countries, the fraction of the population that has escaped its set point will depend on a country’s stage in the process of development. India and Indonesia are evidently at a stage where a substantial fraction of the population lies on either side of the threshold, resulting in the coexistence of malnutrition and metabolic disease. In contrast, the Ghanaian population appears to be largely at its pre-modern set point, which is why there is no discontinuity. A cross-country comparison of current income and historical income (measured by height in the nineteenth century) provides support for this conjecture: the gap between current and historical income, which determines the fraction of the population that has crossed the threshold, is substantially higher in Asia than in Africa.

The preceding observation leads to the final step of the analysis, where we move from micro-data to cross-regional comparisons. Deaton (2007) observes that adult nutritional status in South Asia is lower than what would be predicted by GDP per capita, whereas the opposite is true for Africa. Moreover, there is an unusually high prevalence of diabetes and related metabolic disorders among South Asians,\(^5\)

As a supplemental check, we examine, and rule out, the possibility that selective child mortality can explain the observed discontinuous relationship between children’s nutritional status and household income.
despite the fact that they have low BMI on average (Narayan, 2017). We show that these seemingly unrelated findings can be easily interpreted through the lens of our model once we take account of the cross-regional income dynamics; i.e. that current income is higher in Asia but historical income (which determines the set point) is higher in Africa.

Although our model is designed to explain nutritional status and the incidence of metabolic disease in developing countries, the preceding arguments can be used to examine the same outcomes for immigrants from those countries residing in advanced economies. Given the enormous income differential between origin and host country, most migrants to advanced economies will escape the nutrition trap in the first generation. This is consistent with the empirical evidence that migrants’ nutritional status converges to the level of the native population very swiftly (Alacevich and Tarozzi, 2017). The BMI set point, however, is heritable and can persist for multiple generations. Given the low set point that the migrants and their descendants are endowed with, these groups will continue to face a high risk of metabolic disease, long after they might have assimilated, culturally and economically.

Based on the cross-regional differences in historical income, which implies a lower set point in Asian populations, we would expect migrants from those populations to be at elevated risk of metabolic disease. Immigrants from South Asia residing in the U.K. and the U.S. are indeed many times more likely to have metabolic diseases than the native population, despite having lower BMI’s (McKeigue et al., 1991; Oza-Frank and Narayan, 2010; Staimez et al., 2013; Kanaya et al., 2014). Other studies, cited in Gujral et al. (2013), document similar patterns in countries such as Fiji, South Africa, and Singapore to which South Asians moved many generations ago as indentured workers and subsequently became relatively wealthy.\(^6\)

While the model is informative about a variety of health outcomes at the micro and the macro level, it is important, particularly from a policy perspective, to go further and quantify the effect of the set point on malnutrition and the prevalence of metabolic disease. A comparison of counter-factual nutritional status and actual nutritional status (predicted by the estimated model) with IHDS data indicates that stunting among 5-19 year old children would have declined by 30% and that the fraction of underweight adults (with BMI less than 18.5) would have declined by 50% in the absence of a set point. To quantify the contribution of the set point to metabolic disease, we first show, based on the model, that the risk of disease will not respond to variation in BMI below a threshold (where individuals are at their set point), but will be increasing in BMI above the threshold. Estimates with IHDS data locate this threshold at a BMI just under 22 for the country as a whole and below 21 for South India, which is well within the normal range (18.5-25). These findings imply that a substantial fraction of the Indian population is at elevated risk of metabolic disease. The set point is predetermined and, hence, cannot be targeted directly. However, health policies can be designed to take account of the set point, and its consequences.

\(^6\)The elevated risk of metabolic disease will not be permanent. While epigenetic traits acquired in the pre-modern economy may be heritable, they are not as rigid as genetic traits, and in the long-run they will cease to be salient. This is also true for the native population in advanced economies, which presumably went through the same disease-nutrition transition that we describe in this paper, but more than a century ago. We would expect that by now the set point in those populations is independent of the pre-modern economic conditions that drive our analysis, in line with Deaton’s (2007) finding that the income-nutritional status relationship is stronger in advanced economies than in developing economies.
as discussed in the concluding section.

1 Biological Foundations

It has been hypothesized that the fetus will respond to (unexpectedly) low nutrition in the uterine environment by slowing down its growth (Hales and Barker, 1992). If developing germ cells, which will become adult gametes (sperm or eggs) are modified by this response, then the altered phenotype can be passed on to the next generation through what is known as intergenerational epigenetic inheritance (Heard and Martienssen, 2014). In addition, if under-nutrition is sustained over many generations, then it has been conjectured that the fetus will adapt to predicted conditions of scarcity in the extrauterine environment by responding to maternal cues in utero (Gluckman and Hanson, 2006). This adaptation can alter gene expression or epigenes in a way that is both optimal from an evolutionary perspective (Richards, 2006; Jablonka and Raz, 2009) and is transgenerational, extending to descendants more than two generations removed who are not directly exposed to the insult (Sales et al., 2017).

For the purpose of our analysis, all that we require is that the phenotype should have adapted to conditions in the pre-modern economy and that this adaptation should persist for at least a couple of generations after the onset of economic development (Narayan, 2016; Wells et al., 2016). We measure adaptation to pre-modern conditions by the BMI set point, which is determined by ancestral income. Given low food supply and, hence, low levels of consumption on average in the pre-modern economy, the epigenetically determined set point would have been optimally fixed at a low BMI.

Economic development is associated with a substantial increase in income and, with it, consumption. Figure 1, for example, plots GDP per capita (in logs) for India, a country that receives much attention in our analysis, from 1600 to 2016. Income is stable (declining mildly) for the first 350 years, after which it starts to increase steeply. Based on the preceding discussion, the set point in the Indian population would have been determined by economic conditions in the pre-modern economy, prior to 1950. Given the heritability of epigenetic traits, this low-BMI set point would have been transmitted to subsequent generations. This explains, in part, why nutritional status has not improved as expected with economic development. However, we posit that the body can only defend the set point up to a threshold level of consumption, with escape from the nutrition trap being associated with an increased risk of metabolic disease.

Among the metabolic diseases that are positively associated with economic development, Type 2 diabetes has received disproportionate research and policy attention. Type 2 diabetes manifests in two forms (Narayan, 2016): (i) Type 2A diabetes is caused by insulin resistance, particularly among obese individuals. This type of diabetes is most commonly observed in advanced economies, where the epigenetically determined set point associated with economic conditions in the pre-modern economy is no longer relevant. (ii) Type 2B diabetes, which is the focus of our analysis, is caused by poor insulin secretion, and is largely associated with normal weight individuals in developing economies (Narayan, 2017). Individuals who remain at their epigenetically determined set point are not at elevated risk of Type 2B diabetes, even if their consumption has increased with economic development. This is because their metabolism
can adjust to the changes in consumption, ensuring that the body’s energy balance is maintained. It is individuals who have escaped the nutrition trap, but who are not necessarily overweight, who are at elevated risk because the metabolic load now exceeds their metabolic capacity (as also noted by Wells et al. (2016)). This argument is consistent with recent evidence indicating that there is a personal BMI threshold above which normal weight individuals are at elevated risk of developing Type 2 diabetes (Taylor and Holman, 2015).

In our framework, economic development increases the prevalence of diabetes, and related health conditions such as hypertension and cardiovascular disease, through two channels. At the extensive margin, it increases the fraction of the population that has escaped the nutrition trap and is at risk of these diseases. At the intensive margin, it increases the risk of metabolic disease, conditional on the individual having escaped the trap, by increasing the mismatch between ancestral income, which determines the BMI set point, and current income. Studies from across the world that have examined the developmental origins of adult disease provide support for this claim. The ‘thrifty phenotype’ hypothesis (Hales and Barker, 1992) is concerned with environmental shocks, both in developing and developed economies, that make the fetus grow more slowly in response to under-nutrition in utero. This response, which shifts the set point for a single generation or at most two generations (if there is intergenerational epigenetic inheritance, as discussed above) confers immediate advantage. However, a robust finding from the fetal origins literature is that a combination of accidentally low birth weight; i.e. a low set point, and high adult BMI subsequently puts individuals at increased risk of diabetes, hypertension, and cardiovascular disease (Hales et al., 1991; Barker et al., 2002; Bhargava et al., 2004; Li et al., 2016).\(^7\)

\(^7\)Providing support for the mismatch hypothesis, individuals who were subjected to caloric restrictions in utero during the 1944-1945 Dutch famine had a heightened risk of metabolic risk as adults in a subsequently affluent economy (Ravelli et al., 1998). In contrast, fetal survivors of the Leningrad siege did not experience adverse health outcomes during adulthood, presumably because there was little difference between their intrauterine and extrauterine economic environment (Stanner
economies that we consider, the set point is determined by conditions in the pre-modern economy over many generations. The threat to the stability of the biological system is economic development. Either way, it is the mismatch between the set point BMI and the BMI in adulthood (conditional on having escaped the nutrition trap) that determines the risk of metabolic disease.\textsuperscript{8}

As described above, the presence of an epigenetically determined set point can explain both the persistence of malnutrition and the increased incidence of metabolic disease in developing countries. However, while it has been established that environmental cues such as temperature can have transgenerational effects in plants (Heard and Martienssen, 2014) and there is evidence that epigenetic inheritance occurs in small mammals (Radford, 2018), the evidence for epigenetic adaptation and inheritance in humans is sparse. In addition, there is a lack of evidence supporting the presence of a set point in humans (Müller et al., 2010). An important objective of our research will be to fill this gap in the literature. We do this by developing a model that generates predictions for the cross-sectional relationship between current income and nutritional status, as well as metabolic disease, when a BMI set point is present. We will subsequently test these predictions with cross-sectional micro data from multiple developing countries, supplementing the analysis with direct tests of the epigenetic mechanism, going back many generations, that gives rise to the set point.

2 The Model

2.1 Population and Income

The population consists of a large number of infinitely lived dynasties. Each dynasty consists of a single individual in each time period or generation, who is replaced by a single descendant in the period that follows. There is a fixed return on wealth in each period; i.e. an income flow, which is consumed, so that the stock is passed on (without depletion) to the next generation. We will thus use (permanent) income and wealth interchangeably in the discussion that follows. Denote the logarithm of the dynasty’s initial income, in period 0, by \( y_0 \). We normalize so that the distribution of initial income is bounded below at zero. We can think of the initial period as describing the pre-modern economy, while subsequent periods describe the process of development. Permanent income in an economy is well approximated by the log-normal distribution (Battistin et al., 2009). We thus assume that each dynasty receives a permanent, additive and independent income shock \( u_\tau \) in each subsequent period \( \tau \), where \( u_\tau \sim N(\mu, \sigma^2) \). Solving recursively, log-income of a dynasty in period \( t \) is, \( y_t = y_0 + U_t \), 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 \( \mu_t \) and \( t\sigma^2 \) by \( \sigma_t^2 \).

\textsuperscript{8}In our analysis, the risk of diabetes, and metabolic disease more generally, is linked directly to an epigenetically determined set point (adapted to conditions in the pre-modern economy). In the long-run, this set point will cease to be relevant. The set point will then default to the position that is determined by genetics, which is effectively permanent in the time-frame that we consider. Neel’s (1962) ‘thrifty genotype’ hypothesis will now be relevant, and particular individuals and sub-populations that have the misfortune of being endowed with these genes will face a heightened risk of obesity and insulin-resistant diabetes (Martorell, 2005).
2.2 Structural Relationships

In this section we describe the structural relationships between (i) nutritional status, measured by BMI, and income, and (ii) the risk of metabolic disease and income, during the process of economic development.

A dynasty’s set point for its body weight or BMI is determined by its initial income, $y_0$. There is a positive and continuous relationship between consumption and income in each time period. In addition, BMI is increasing continuously in consumption in the initial period; those dynasties that consumed at a higher level in the pre-modern economy will have a higher set point.\(^9\) We thus specify the following relationship between initial BMI, $z_0$, or the set point, and initial income:

$$z_0 = a + by_0.$$  \hspace{1cm} (1)

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

$$z_t = \begin{cases} 
    a + by_0 & \text{if } U_t \leq \alpha \\
    a + by_t & \text{if } U_t > \alpha
\end{cases}$$ \hspace{1cm} (2)

Equation (2) imposes the restriction that the structural 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 $y_t$. We will test this restriction by separately estimating the $b$ parameter, below and above the (estimated) threshold.

Notice that the set point, $z_0$, determined in period 0, is assumed to be fixed across all subsequent generations. Although an epigenetically determined set point may be heritable, it will ultimately cease to be relevant once a changed economic environment has been in place for a sufficient number of generations. Our model thus describes the relationship between nutritional status and income over a finite number of generations during the initial rapid-growth phase of economic development.

Notice also that there is no lower threshold; the implicit assumption is that dynasties do not regress with regard to nutritional status during a period of rapid economic growth. Given historically low levels of food supply in developing countries, the metabolism would have adapted to defend the set point especially vigorously against downward fluctuations in consumption.\(^{10}\) Although mean income is increasing in our model, the distribution of income shocks is unbounded and, hence, a small number of dynasties could, nevertheless, accumulate 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

\(^9\)In practice, epigenetic adaptation occurs over a long period of time. We can thus think of period 0 in the model as spanning multiple generations in the pre-modern era.

\(^{10}\)For example, despite repeated weight cycling in response to seasonal fluctuations in food supply, minimal body weight in a sample of rural Gambian women remained extremely stable (within 1.5 kg.) over a period of 10 years (Prentice et al., 1992). Similarly, while a substantial fraction of the population may remain stunted in a rapidly developing economy, we do not expect dynasties to become systematically shorter over successive generations.
negative outcomes. We thus assume that dynasties always successfully defend the set point in the face of negative income shocks, either biologically or by taking advantage of social safety nets to augment their consumption.\footnote{Given that income shocks are positive on average and their distribution is symmetric, such redistribution is feasible. We are effectively ignoring catastrophic common shocks, such as famines, that can shift set points in an entire generation. Such events have always been rare and are less relevant in the modern economy.}

As long as consumption remains within the threshold associated with the dynasty’s set point, metabolic and hormonal adjustments ensure that the increases in consumption that accompany the increases in income due to economic development do not translate into increases in BMI. Once consumption crosses the threshold, however, the metabolism can no longer maintain the energy balance and BMI starts to track with current income. As discussed in the preceding section, the accompanying mismatch between metabolic capacity and metabolic load simultaneously increases the risk of metabolic diseases. As in the fetal origins literature, this risk is specified to be increasing in the gap between current income, $y_t$, which determines current BMI (conditional on having crossed the threshold) and initial income, $y_0$, which determines the BMI set point. The structural relationship between the probability of metabolic disease, $P(D_t)$, and income can thus be characterized as follows:

$$P(D_t) = \begin{cases} 
\gamma_1 & \text{if } U_t \leq \alpha \\
\gamma_1 + \gamma_2 (y_t - y_0) & \text{if } U_t > \alpha
\end{cases} \quad (3)$$

### 2.3 BMI-Income Relationship

Figure 2 describes the evolution of BMI across multiple generations (time periods) for a single dynasty, based on the structural relationship specified above. For expositional convenience, we assume that the dynasty only receives positive income shocks. Starting from an initial income, $y_0$, the dynasty’s income thus increases monotonically over time. However, it’s members’ BMI will remain at the dynasty’s set point, $z_0 = a + by_0$, until $y_t$ exceeds $y_0 + \alpha$. At that point in time, there will be a discrete increase in BMI, after which BMI will track with current income. If transgenerational data were available for multiple dynasties, then these predictions could be tested directly. However, standard data sets typically provide information on nutritional status and household income at a single (current) point in time. We thus proceed to derive the cross-sectional relationship between BMI and income, as implied by equation (2), when a dynasty-specific set point for body weight is present.

Recall that we normalize so that the initial income distribution is bounded below at zero. We also do not specify a lower threshold for the set point. It follows that all individuals with $y_t \leq \alpha$ must lie within their dynasty’s set point threshold; some of these individuals will belong to dynasties that had initial incomes below $\alpha$ and which subsequently increased their income by relatively little, whereas others will belong to dynasties whose income has drifted down over time. Given the assumed (normal) distribution of income shocks, mean BMI at any given level of income, $y_t$, for $y_t \leq \alpha$ is determined by the following
expression:
\[
\bar{z}(y_t \leq \alpha) = \int_{-\infty}^{y_t} \left[ a + b(y_t - U_t) \right] \frac{\phi(U_t; \mu_t, \sigma^2_t)}{\Phi(y_t; \mu_t, \sigma^2_t)} \, dU_t \\
= a + b(y_t - e^L(y_t)), \quad e^L(y_t) = \frac{1}{\Phi(y_t; \mu_t, \sigma^2_t)} \int_{-\infty}^{y_t} U_t \phi(U_t; \mu_t, \sigma^2_t) \, dU_t 
\]

For individuals with \( y_t > \alpha \), some will have crossed their set point threshold, while others (who started with a higher initial income) will remain within their thresholds. The expression for mean BMI at income level \( y_t \), given that \( y_t > \alpha \), thus includes both types of individuals,
\[
\bar{z}(y_t | y_t > \alpha) = \int_{-\infty}^{\alpha} \left[ a + b(y_t - U_t) \right] \frac{\phi(U_t; \mu_t, \sigma^2_t)}{\Phi(y_t; \mu_t, \sigma^2_t)} \, dU_t + \int_{\alpha}^{y_t} \left[ a + b(y_t - U_t) \right] \frac{\phi(U_t; \mu_t, \sigma^2_t)}{\Phi(y_t; \mu_t, \sigma^2_t)} \, dU_t \\
= a + b(y_t - e^H(y_t)), \quad e^H(y_t) = \frac{1}{\Phi(y_t; \mu_t, \sigma^2_t)} \int_{\alpha}^{y_t} U_t \phi(U_t; \mu_t, \sigma^2_t) \, dU_t 
\]

As shown in the Appendix, closed-form expressions for \( e^L(y_t) \) and \( e^H(y_t) \) can be derived using the properties of the normal and standard normal distributions:
\[
e^L(y_t) = \mu_t - \sigma_t \frac{y_t - \mu_t}{\sigma_t} = \mu_t - \sigma_t \Lambda \left( \frac{y_t - \mu_t}{\sigma_t} \right) 
\]
\[
e^H(y_t) = \mu_t \Phi \left( \frac{\alpha - \mu_t}{\sigma_t}; 0, 1 \right) - \sigma_t \phi \left( \frac{\alpha - \mu_t}{\sigma_t}; 0, 1 \right) / \Phi \left( \frac{\alpha - \mu_t}{\sigma_t}; 0, 1 \right) 
\]
Figure 3: Cross-Sectional Relationships

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

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

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

2.4 Disease-Income Relationship

Taking as given the structural relationship between the probability of metabolic disease, $P(D_t)$, and income, as specified in equation (3) for a single dynasty, the corresponding relationship in the cross-section across dynasties can be derived as follows:

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

The proof (in the Appendix) follows the same steps that we use to derive closed-form expressions for $e^L(y_t)$, $e^H(y_t)$. The $P(D_t) - y_t$ relationship specified by Proposition 2 is described graphically in Figure 3. This relationship is qualitatively the same as the $\bar{z}(y_t) - y_t$ relationship, except that the slope is zero below the threshold. This is because the risk of metabolic disease is constant (and the same) for all
individuals who remain at their set point. Recall that all individuals below the income threshold are at their set point. Above the threshold, in contrast, the risk of metabolic disease is increasing in income. This is due to (i) the greater fraction of individuals who have escaped their set point, and (ii) the increased risk for those who have escaped. Note that the model predicts that the \( \tau(y_t) - y_t \) and \( P(D_t) - y_t \) relationships will exhibit a slope discontinuity at the same income level: \( y_t = \alpha. \)

3 Testing the Model

3.1 Descriptive Statistics

The key variables in the model are income, nutritional status, and the probability of metabolic disease. Although there is a single individual in each generation in our model, multiple individuals will reside in a household. Income will thus be measured at the household level. Nutritional status is measured for each (available) member of the household; by height-for-age for children and BMI for adults. BMI rather than height is used as our benchmark measure of nutritional status for adults because it is directly related to the set point for body weight or BMI that drives the model. The additional advantage of using BMI is that it will respond to nutrient intake into adulthood; this is especially important in a dynamic economy.

The primary tests of the model are conducted with Indian data. This is because the rapidly developing Indian economy is simultaneously characterized by high levels of malnutrition and a high incidence of metabolic disease; the two stylized facts that motivate our research. The core data set that we use for the analysis is the India Human Development Survey (IHDS). This nationally representative household survey, which was conducted in 2004-2005 and 2011-2012, includes detailed information on household income, nutritional status for children and adults residing in the household at the time of the survey, and the incidence of metabolic diseases (diabetes, hypertension, and cardiovascular disease) among adult members of the household. The survey includes, in addition, information on household composition, food intake, short-term morbidity among the children, and detailed geographic locators, which will be used to supplement the analysis.

Figure 4 describes the distribution of household income in the IHDS data, measured as the log of monthly income in thousands of Rupees, averaged over the two survey rounds.\(^\text{14}\) The vertical dashed line in Figure 4 denotes the median income, which is 1.8 in the nationally representative sample of households. Our tests of a slope-change, reported below, will locate an income threshold close to the median income,

\(^{12}\)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 \( y_0 \), in which case the threshold would be located at \( y_t = y_0 + \alpha \). This would change the interpretation of the threshold location, but otherwise leave the analysis unchanged.

\(^{13}\)The Demographic Health Survey (DHS), which is used by Deaton (2007) and Jayachandran and Pande (2017) also contains many of these variables. However, the DHS is not suitable for our purposes because it only collects indicators of asset ownership and access to services, which must then be converted into a crude wealth statistic using principal component analysis. The tests of the model, particularly the statistical tests to locate a slope-change at an income threshold, cannot be implemented without fine-grained income data.

\(^{14}\)Household income includes farm income, non-farm business income, wage income, remittances, and government transfers. To make incomes in the two rounds comparable, we adjust 2004-2005 incomes to 2011-2012 prices. For rural areas, the correction is based on the Consumer Price Index (CPI) for agricultural wage labor and for urban areas it is based on the CPI for industrial workers.
which tells us that it is not just the poorest who remain in the nutrition trap in this economy.

Figure 5a describes the nutritional status of children in the IHDS, separately for children aged 0-59 months and 5-19 years. Nutritional status, measured by the height-for-age, is reported as a z-score, based on child growth standards provided by the WHO. We see that a substantial fraction of Indian children are stunted; with a z-score less than -2. Figure 5b describes the corresponding distribution of adult nutritional status, measured by the BMI. The vertical dotted line in the figure denotes a BMI of 18.5, which is a cutoff conventionally associated with being underweight. We see that a substantial fraction of the Indian population remains below this cutoff, despite the economic progress of the past decades. By international standards, individuals are underweight if their BMI is below 18.5, the normal range is 18.5-25, the overweight range is 25-30, and obesity is defined by a BMI above 30. Based on this convention, most Indians are underweight or normal weight, and only a small fraction are obese. BMI that is too low or too high is physiologically damaging, but the latter is evidently less of a problem in India. We will see below that diabetes and related metabolic disorders, which are commonly associated with obesity in advanced economies, largely affect normal weight individuals in India.

3.2 Cross-Sectional Analysis

Proposition 1 derives the cross-sectional relationship between nutritional status and income when a set point is present: although the relationship is positive at all income levels, there will be a discontinuous shift

---

15Given that the nutritional status measures are age-specific, information from both survey rounds is separately included for those children who appear in both rounds. The growth standard for children aged 0-59 months is based on the Multicentre Growth Reference Study (MGRS), conducted between 1997 and 2003. For children aged 5-19, we use the 2007 WHO Reference, which is a reconstruction of the 1977 National Center for Health Statistics (NCHS) growth standard. Following the recommendation of the WHO, z-scores outside the (-6,6) interval are dropped from the analysis.

16The BMI is defined as the weight in kilograms divided by the square of the height in meters. The BMI was collected for men and women in the 2011-2012 round, but only for a small fraction of men in the 2004-2005 round. As with the children, we include the BMI statistic separately from the two survey rounds when it is available for an adult.
to a steeper slope at a particular income threshold. Proposition 2 derives the corresponding relationship between the risk of metabolic disease and income: while a slope-change at the same income threshold is predicted, the difference is that variation in income is not expected to affect the risk of disease below the threshold.

We test these predictions with nationally representative data from the India Human Development Survey (IHDS) by separately estimating the relationship between income and both nutritional status and the probability of metabolic disease. Household income is measured as the average over the 2004 and 2012 survey rounds. This smooths out noise in the round-specific income measures and, given that the rounds were conducted nearly a decade apart, provides a more accurate estimate of the household’s permanent income. Nutritional status is measured by height-for-age for children and by BMI for adults, with individual information from both survey rounds included in the estimation sample when available. Metabolic disease is constructed as a binary variable that indicates whether an individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.\(^{17}\)

Figure 6a nonparametrically estimates the relationship between the nutritional status of the children and household income. Figure 6b repeats this exercise with nutritional status and the probability of metabolic disease among adult members of the household as outcomes.\(^{18}\) Although our analysis focuses on the income effect, other individual and household characteristics could also determine nutritional status and the risk of metabolic disease. All of the estimating equations in our analysis thus include the following covariates: gender, age (linear, quadratic, and cubic terms), birth order (for the children), caste

\(^{17}\)While BMI can shift up or down from one period to the next, metabolic disease is effectively irreversible; recent findings (Taylor, 2013) indicate that diabetes can be reversed, but this requires weight loss in excess of 15 kg. If an individual is reported to have the disease in the 2004-2005 round, there is thus no additional information content in the 2011-2012 report. Those observations are thus dropped from the estimation sample.

\(^{18}\)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.
group, rural-urban dummy, and district dummies. The effect of gender bias on nutritional status, as documented by Jayachandran and Pande (2017), is captured by the gender and birth order dummies. Geographical variation in food tastes, as emphasized by Atkin (2013, 2016) or in the disease environment, as documented by Spears et al. (2013), Duh and Spears (2017), and Dandona et al. (2017) is captured by the district dummies and the rural-urban dummy. The covariates listed above are partialled out using the Robinson (1988) procedure prior to the nonparametric estimation reported in Figures 6a and 6b.

It is evident from both figures, and all four outcomes, that the income effect is weaker at lower income levels, with a slope-change at an income threshold between 1 and 2. To test formally for a slope-change and to place statistical bounds on the location of the threshold (where relevant) we implement a procedure developed by Hansen (2017). The procedure involves sequential estimation of the following piecewise linear equation:

$$z_i = \beta_0 + \beta_1 y_i + \beta_2 (y_i - \gamma) + x_i \lambda + \epsilon_i,$$

(8)

where $z_i$ is an outcome of interest; e.g. nutritional status, $y_i$ is household $i$’s income, $\gamma$ is the location of the income threshold (which must be estimated), $\beta_1$, $\beta_2$ are slope parameters, and $x_i$ is a vector of additional covariates. This equation is estimated at different assumed income thresholds (values of $\gamma$), starting at a very low income level and then covering the entire income range in small increments. An F-type statistic is computed at each assumed threshold, based on a comparison of the sum of squared residuals at that assumed threshold and the minimized value across all assumed thresholds. This statistic will have a minimum value of zero by construction, and the assumed income threshold corresponding to that value will be our best estimate of the true threshold. If there is indeed a slope-change, then

---

19Age is measured in years, except for the analysis with 0-59 month children where it is measured in months. The birth order is top coded at 3.
the F-type statistic will increase steeply as the assumed threshold moves away (on either side) from the income level at which it is minimized.

Figures 7a and 7b plot the F-type statistic across the range of assumed thresholds for children’s nutritional status and the adult outcomes, respectively. Bootstrapped, outcome-specific 5% critical values for the F-type statistic are also reported in the figures, allowing us to locate the threshold with the requisite degree of statistical confidence. The F-type statistic increases steeply as the assumed threshold moves away from the income level at which it is minimized, which implies, in turn, that the location of the threshold can be bounded with a relatively high degree of statistical precision. Our best estimate of the threshold location matches closely for the 0-59 month and the 5-19 year old children. Nutritional status is measured by height-for-age for the children and BMI for the adults. Despite the fact that we are using different measures, the estimated threshold for the adults in Figure 6b, with BMI as the dependent variable in the estimating equation, is very close to what we obtain for the children in Figure 6a. The estimated threshold location with the probability of metabolic disease as the outcome shifts to a slightly higher income level, but we will see below that the 95% confidence intervals for the threshold location overlap across all outcomes.

The same (wild) bootstrap procedure that is used to compute the critical values and, hence, the 95% confidence interval for the threshold location in Figures 7a and 7b can also be used to compute standard errors for the slope parameters in a piecewise linear equation estimated at the threshold we have located. Moreover, a similar bootstrap procedure can be used to test our statistical model with a slope change at an income threshold, as described in equation (8), 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
Table 1: Piecewise Linear Equation Estimates - nutritional status and metabolic disease

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>HFA 0-59 (1)</th>
<th>HFA 5-19 (2)</th>
<th>BMI (3)</th>
<th>Disease (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline slope ($\beta_1$)</td>
<td>-0.049 (0.071)</td>
<td>0.024 (0.031)</td>
<td>0.239** (0.046)</td>
<td>0.001 (0.002)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.365** (0.068)</td>
<td>0.206** (0.031)</td>
<td>0.940** (0.054)</td>
<td>0.025** (0.003)</td>
</tr>
<tr>
<td>Threshold location ($\gamma$)</td>
<td>1.40 [1.20, 2.00]</td>
<td>1.50 [1.35, 1.90]</td>
<td>1.65 [1.55, 1.75]</td>
<td>1.90 [1.80, 2.05]</td>
</tr>
<tr>
<td>Threshold test $p$-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>-1.991</td>
<td>-1.649</td>
<td>22.002</td>
<td>0.670</td>
</tr>
<tr>
<td>N</td>
<td>21634</td>
<td>48845</td>
<td>76949</td>
<td>147729</td>
</tr>
</tbody>
</table>

Source: India Human Development Survey (IHDS)

Disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.

Logarithm of household income is the independent variable.

Covariates listed in the text are included in the estimating equation.

Bootstrapped standard errors are in parentheses.

Bootstrapped 95% confidence bands for the threshold location are in brackets.

* significant at 10%, ** at 5% and *** at 1%

each outcome. The reported point estimates of the baseline slope parameter ($\beta_1$) and the slope-change parameter ($\beta_2$) are obtained at our best estimate of the true threshold, $\gamma$, for each outcome. As predicted by our model with a set point, the slope increases to the right of the threshold with each outcome (the slope-change coefficient is positive and significant). Proposition 1 indicates, in addition, that the slope to the left of the threshold should be positive with nutritional status as the outcome. This result is obtained for adults (Column 3) but not children (Columns 1-2), perhaps because sample sizes are smaller for the children or because the income effect strengthens over the life-course. In line with Proposition 2, there is no relationship between the probability of metabolic disease and household income below the threshold in Column 4, in contrast with the strong positive relationship above the threshold.

The estimated threshold location ranges from 1.4 to 1.9 for the four outcomes, with some amount of overlap in the confidence intervals between any pair of outcomes. Recall that the median income in our nationally representative sample of households is 1.8. Based on our model, all households with income to the left of the threshold remain in the nutrition trap, as do some households to the right of the threshold. This implies that a substantial fraction of the Indian population remains in the nutrition trap at this stage of economic development, with this group being partly responsible for the weak relationship between nutritional status and income that has been documented in the literature. Among the households to the right of the threshold, those that have escaped the nutrition trap are at elevated risk of metabolic

---

20The only exception is adult disease and nutritional status, where the confidence intervals are extremely narrow with both outcomes. As reported below, confidence intervals with these two outcomes overlap with IHDS South India data as well as with Indonesian data from the IFLS.
disease. The micro-evidence we have provided can thus explain the co-existence of malnutrition and a high incidence of diabetes and other metabolic disorders at this stage in India’s economic development, as a consequence of an underlying predetermined set point in the population.

We complete this section by verifying the robustness of this evidence in a number of ways: First, we show in Appendix Table A1 that the results are robust to including period-specific income in place of average income (over the two survey rounds). Second, we show in Appendix Table A2 that the results continue to be obtained when the outcomes are restricted to the 2011-2012 survey round. Third, we include average education among adult women and adult men in the household, as well as household composition, measured by the number of children, the number of teens, the number of adults, and the number of adults engaged in physical labor as additional covariates in the estimating equation in Appendix Table A3. While these variables could independently determine feeding practices, health seeking behavior, and other decisions that determine nutritional status and health outcomes, we see that the results are robust to their inclusion. Fourth, we show in Appendix Table A4 that the results continue to be obtained with alternative measures of nutritional status; weight-for-age for the children and height for adults.

3.3 Alternative Explanations

The additional covariates that we include in the estimating equations account for two independent determinants of nutritional status in India: gender bias and a culturally determined preference for particular foods. The district dummies and the rural-urban dummy will also subsume spatial variation in the disease environment and the availability of health services. However, we must account for the possibility that the proximate determinants of nutritional status emphasized by Deaton (2007) – nutrient intake and children’s illness – vary with household income in a way that independently generates our results. Our model assumes a positive and continuous relationship between nutrient intake (consumption) and income. It is the biologically determined set point that breaks the smooth relationship between nutritional status and consumption and, by extension, income. Suppose, instead, that the nutrient intake-income relationship strengthens discontinuously above an income threshold. Alternatively, suppose that there is a discontinuous change in the children’s illness-income relationship. Either way, the nonlinear nutritional status-income relationship that we estimate could be obtained without a set point.

To assess the validity of these alternative explanations, we nonparametrically estimate the nutrient intake-household income relationship in Figure 8a and the children’s illness-household income relationship in Figure 8b using IHDS data. Nutrient intake is measured by the consumption of calories and fat (in grams) at the household level. Children’s illness is measured by whether the child (aged 0-19) is reported

---

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

22 Deaton (2007) also considers energy expenditure (physical activity) as a determinant of nutritional status. Ng and Popkin (2012) decompose total energy expenditures into types of activity: work, active leisure, travel, and domestic tasks. The work category accounted for over 80% of the total energy expenditure in 2000 and 2005 in India. We will thus incorporate the type of work activity in the analysis that follows.

23 Social norms determine feeding practices, health seeking behavior, sanitary practices, and other behaviors that contribute to nutrient intake and the disease environment. These norms can change discontinuously when income in the relevant social group, consisting of multiple dynasties, crosses a threshold level, providing an alternative explanation for our results.

---
to have had diarrhea and cough in the past month. The usual set of covariates, plus household composition and the number of adults engaged in physical labor, are partialled out prior to estimation using Robinson’s procedure. We see that there is a positive and continuous relationship between the intake of calories and fat and household income in Figure 8a.\textsuperscript{24} In contrast, there is a negative and continuous relationship

\textsuperscript{24}Our finding that nutrient intake is increasing continuously in household income does not contradict Deaton and Drèze (2009) who document a decline in real food consumption, even as income increased over time in India, using National Sample
Table 2: Piecewise Linear Equation Estimates - nutrient intake and children’s illness

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>nutrient intake</th>
<th>children’s illness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>log calories</td>
<td>log fat</td>
</tr>
<tr>
<td>Baseline slope ($\beta_1$)</td>
<td>0.058***</td>
<td>0.121***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.002</td>
<td>−0.005</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Imposed threshold ($\gamma$)</td>
<td>1.65</td>
<td>1.65</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>12.511</td>
<td>8.514</td>
</tr>
<tr>
<td>N</td>
<td>74662</td>
<td>74662</td>
</tr>
</tbody>
</table>

Source: India Human Development Survey (IHDS)
Logarithm of household income is the independent variable.
Covariates listed in the text are included in the estimating equation.
Standard errors are reported in parentheses.
* significant at 10%, ** at 5% and *** at 1%

between the incidence of both diarrhea and cough with household income in Figure 8b.

The dotted vertical line in Figure 8a marks the spot at which we located the income threshold with adult BMI as the outcome. The vertical line in Figure 8b marks the threshold location with height-for-age for 5-19 year olds as the outcome. In neither figure do we observe a discontinuous slope-change at the imposed threshold. Indeed, Hansen’s threshold test fails to locate a slope-change at any assumed threshold. Figure 9a tests for a slope-change in the nutrient intake- household income relationship and Figure 9b applies the test to the children’s illness- household income relationship. In contrast with the V-shaped pattern for the F-type statistic that we documented with nutritional status and the risk of metabolic disease as outcomes, the F-type statistic never even exceeds the critical value with three of the four outcomes in Figure 8. For the one outcome – fat intake – where it does, the F-type statistic only exceeds the critical value on one side (to the right) of the assumed threshold at which the statistic is minimized. We cannot place bounds on the threshold location and, hence, we cannot locate a threshold at conventional levels of statistical confidence with any outcome in Figure 8.

Table 2 reports piecewise linear equation estimates, with household income as the independent variable and nutrient intake and children’s illness as outcomes. Nutrient intake is measured for calories, fat, and

---

Survey (NSS) data. Deaton and Drèze (2009) posit that declining levels of physical activity and improvements in the disease environment with economic development could have generated this decline. Providing empirical support for this hypothesis, Duh and Spears (2017) exploit variation within districts over time (with NSS data) and across households in the cross-section (with IHDS data) to establish that an improvement in the disease environment, specifically associated with a reduction in diarrheal disease, does indeed reduce caloric consumption. A rich set of covariates are included in our estimating equation. Among the covariates are caste category, a rural-urban dummy, district dummies, the number of children, teenagers, and adults in the household, and the number of household members engaged in physical labor. These covariates will capture variation in both physical activity and the disease environment across households. Once these confounding factors are accounted for, nutrient intake will increase with income, which is what we observe.
sugar in Columns 1-3 and children’s illness is measured by diarrhea, fever, and cough in the last month in Columns 4-6. Since we cannot locate a threshold with these outcomes, we impose the slope-change in Columns 1-3 at the income level where the slope was located with adult BMI as the outcome and in Columns 4-6 at the income level where the slope was located with height-for-age for 5-19 year olds as the outcome. In contrast with the results obtained with nutritional status and the probability of metabolic disease as outcomes in Table 1, the baseline slope coefficient in Table 2 is large in magnitude, relative to the slope change coefficient, and statistically significant with each outcome. The slope change coefficient is statistically insignificant with each outcome, except for sugar intake where the point estimate is negative. Our results for nutritional status cannot be explained by either a discontinuous relationship between nutrient intake and household income or children’s illness and household income. Similarly, our results for metabolic disease, diabetes in particular, cannot be explained by a discontinuous increase in sugar consumption above an income threshold.

Although the proximate determinants of nutritional status do not vary discontinuously with household income, could the observed nonlinearity be generated by selective child mortality? Suppose that there is a positive and continuous relationship between mean nutritional status and household income, with a fixed dispersion in nutritional status at each level of income, as in Figure 10. If children can only survive above a subsistence nutrition level, and this constraint only binds at lower income levels, then as observed in the figure there will be a discontinuous relationship between mean nutritional status and income. Although the nutritional status-income relationship now precisely matches the prediction of our model, notice that it is driven entirely by households at the lower end of the nutritional status distribution, at each income level. Child mortality is concentrated in the first five years and, hence, if the nutritional status-income relationship is distorted by child mortality, this will show up most clearly among the 5-19 year olds. Figures 11a and 11b report quantile regression estimates of the baseline slope coefficient ($\beta_1$) and the slope-change coefficient ($\beta_2$) in a piecewise linear equation with child (aged 5-19) height-for-age as the outcome.

$$\beta_1$$

Standard errors are not bootstrapped on Table 2 because the threshold location is exogenously imposed.
Figure 11: Conditional Mean and Conditional Quantile Coefficients (child nutritional status with respect to income)

(a) Baseline slope

(b) Slope change

Source: India Human Development Survey (IHDS)

dependent variable. Coefficient estimates for the same equation, evaluated at the mean of the dependent variable rather than at each quantile, were reported earlier in Table 1, Column 2. It is evident from Figures 11a and 11b that those results were not driven by a small fraction of households at the bottom of the nutritional status distribution, as the alternative explanation based on selective child mortality would imply. We cannot statistically reject the hypothesis that the estimated coefficients at each quantile are equal to the corresponding conditional mean coefficient.26

Although we are unable to come up with an alternative explanation for the results that are obtained, some caveats are in order. First, we use coarse measures of nutrient intake – calories, fat, sugar – measured at the household level in our analysis. Food intake at the individual level is difficult to measure and recent evidence (Forouhi et al., 2014) indicates that nutrient-types must be measured at an extremely fine level to accurately predict the risk of diabetes. Second, while our nutritional status measures, based on weight and height, are accurately measured, metabolic diseases (although diagnosed) are self-reported. It is possible that wealthier households are more likely to visit the doctor and, hence, to be diagnosed with these conditions. However, for such differential reporting, mis-measurement of food intake, or any other omitted variable to explain all of our results, it must explain the discontinuity in the relationship between household income and both nutritional status and the risk of metabolic disease, as well as the fact that the threshold is located at the same income level for both outcomes. There is no obvious reason why this would be the case. Nevertheless, to provide independent support for our model, we proceed to explicitly test the underlying mechanism, based on a predetermined BMI set point.

26Deaton (2007) considers the possibility that variation in child survival with income could explain the weak nutritional status-income relationship that he documents across countries. However, evidence from numerous studies, cited in Alderman et al. (2011) indicates that selective mortality would have a negligible effect on the nutritional status-income relationship in most contexts. This appears to hold true in contemporary India as well.
3.4 The Mechanism

Given an epigenetically determined set point for body weight, equation (2) describes the structural relationship between nutritional status and income as follows: For individuals who remain at their set point, BMI is determined by ancestral income, $y_0$. For individuals who have escaped the nutrition trap, BMI is determined by current income, $y_t$. We cannot test these relationships with standard data sets such as the IHDS because $y_0$ is unobserved. This is why we derive and test the corresponding cross-sectional BMI-income ($z_t - y_t$) relationship. However, unique data from the South India Community Health Study (SICHS), together with particular features of the marriage institution in India, can be used to directly test the structural relationship implied by equation (2). The analysis that follows builds on recent research by Borker et al. (2018), which uses SICHS data to examine the functioning of marriage markets in India.

The SICHS covers a rural population of 1.1 million individuals residing in Vellore district in the South Indian state of Tamil Nadu. Borker et al. (2018) provide a detailed description of the study area, documenting that it is representative of rural Tamil Nadu and rural South India with respect to socioeconomic and demographic characteristics; e.g. age distribution, marriage patterns, literacy rates, and labor force participation. Two components of the SICHS are relevant for our analysis: a census of all 298,000 households residing in the study area, completed in 2014, and a detailed survey of 5,000 representative households, completed in 2016. The SICHS census collected each household’s income in the preceding year. The SICHS survey collected information on the marriage of the household head and his wife, as well as their parents. These data are supplemented with historical records, obtained from the British Library in London, on the agricultural revenue tax per acre of cultivated land that was collected from each village in the Northern Tamil Nadu region (extending beyond the study area) in 1871. As shown below, current household income from the SICHS census, information on marriages over two generations from the SICHS survey, and historical village tax revenue data, taken together, can be used to construct measures of ancestral income, $y_0$, for each household.

Each dynasty consists of a single individual in each generation in our model. We now consider an extension to the model in which there are two individuals – a man and a woman – in each generation. They are succeeded by two children – a male and a female – preserving the gender balance in the population over time. What we refer to as (permanent) income in the model can be equivalently interpreted as wealth, with the couple consuming an amount that is equal to the return on their wealth in each generation. Their wealth is thus passed on (equally divided) to the next generation. Once we introduce males and females in the model, we must specify how they match. Our model incorporates a particular feature of the marriage institution in India, which is that matches are arranged by the parents and relatives of the groom and bride, with families matching assortatively on wealth. Each individual in the model thus

---

27. The SICHS study area was purposefully selected to be representative of rural South India, defined by the following states: Tamil Nadu, Andhra Pradesh, Karnataka, and Maharashtra. Munshi and Rosenzweig (2016) define the South Indian region by the same set of states. Kerala is excluded from the list because it is an outlier on many socioeconomic characteristics.

28. 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.

29. Borker et al. (2018) use data on marriages from the SICHS survey to document that approximately 85% of marriages...
matches with a partner who inherits the same amount of wealth. The total wealth inherited by the couple is augmented by a wealth (or permanent income) shock to determine the total amount of wealth (or income) that is available to them for consumption. Figure 12 describes the matching and wealth process, as described above, for a single dynasty over three generations. Linking our model to the data, household heads in the SICHS survey are aged 25-60. Their grandparents would have been working 60-80 years ago; i.e. in the first half of the twentieth century, which is when the Indian economy began to develop after centuries of stagnation, as observed in Figure 1. We thus assume that the current generation of adults in the SICHS data is period \( t = 3 \) in the model (the structural estimates will provide additional support for this assumption). Figure 12 thus describes the income process over the first three generations of Indian economic development for an adult, \( I = \{M, W\} \), from the current generation, where \( M \) denotes the household head and \( W \) denotes his wife.\(^{30}\)

Once the dynasty consists of a man and a woman in each generation, ancestral income can be measured by either \( y_{0M}I/2 \) or \( y_{0W}/2 \); i.e. the wealth inherited by the paternal or maternal grandparents, respectively. Note that we do not need to go back further because the income shocks only commence in the grandparents’ generation. All the ancestors of the paternal (maternal) grandparents would thus have inherited \( y_{0M}/2 \) (\( y_{0W}/2 \)). Which ancestral wealth measure is appropriate depends on whether epigenetic traits are transmitted through the mother or the father. Trans-generational epigenetic inheritance was traditionally assumed to occur exclusively through the maternal line (Lind and Spagopoulou, 2018). However, recent research indicates that paternal traits can be transmitted epigenetically (Jablonka and Raz, 2009) and thus we allow for both possibilities in our analysis. If epigenetic inheritance occurs through the female line, then the initial wealth that determines the set point in period 0 will be the mother’s mother’s inherited wealth; i.e. \( y_{0W}/2 \). If epigenetic inheritance operates through the male line, then the set point

\[^{30}\text{Each paternal (maternal) grandparent inherited wealth } y_{0M}/2 \text{ (} y_{0W}/2 \). This inheritance was augmented by wealth shocks } u_{1M}^{M}, u_{1W}^{W} \text{, respectively, so that the grandparents on both sides ended up with wealth } y_{1}. \text{ The parents of individual } I \text{ both inherited } y_{1}/2 \text{, and this initial endowment was augmented by a wealth shock, } u_{2}, \text{ so that they ended up with wealth } y_{2}. \text{ Both the husband } (I = M) \text{ and the wife } (I = W) \text{ in the current generation thus inherit } y_{2}/2 \text{ and so their family ends up with wealth, } y_{3} = y_{2} + u_{3}.\]
to determine the father’s father’s inherited wealth; i.e. $y_{0IM}^M/2$.

To construct measures of $y_{0IM}^W/2$ and $y_{0IM}^M/2$, we make use of the 1871 revenue tax data. The revenue tax per acre of cultivated land was based on detailed measures of soil quality, irrigation, and other growing conditions in the village. It thus would have been highly correlated with agricultural productivity and, by extension, (permanent) household income in the village in 1871. Suppose that log household income in 1871, which is long before period 1 in our model, is increasing linearly in the per acre revenue tax. To begin with, assume that there was no income heterogeneity within the village in 1871. Then the dynamic income process specified by our model implies that $y_{2014} = \lambda R_{1871} + U_{2014}$, where $y_{2014}$ is log household income in 2014, obtained from the census, $R_{1871}$ is per acre revenue tax in the village in 1871, and $U_{2014}$ represents the subsequent accumulation of income shocks from period 1 onwards, which we place around the middle of the twentieth century. Note that $\lambda R_{1871}$ is equivalent to $y_0$ in our model.\footnote{The implicit assumption underlying the relationship between historical and current income is that households or dynasties, in particular, the men in those dynasties, have remained in the same village for many generations. This assumption is supported by recent evidence that permanent male migration from rural to urban areas is extremely low in India (Munshi and Rosenzweig, 2016). Providing additional support for the low spatial mobility in India, the correlation between the caste composition of each village in the study area in 1871, based on the colonial population census, and the same statistic in 2014, based on the SICHS census, is as high as 0.85.} In practice, the relationship between historical income and the 1871 revenue tax at the village level would have varied across castes or jatis, whose members were historically engaged in different occupations. We account for this by including caste fixed effects and the interaction of the fixed effects with the 1871 revenue tax in the estimating equation. The historical revenue tax strongly predicts current household income, with the F-statistic measuring joint significance of the revenue tax variable and the revenue tax-caste interactions equal to 20.4. We thus use that part of the variation in current income that is explained by the historical revenue tax variable, caste, and the revenue tax-caste interaction to measure $y_0$.\footnote{Caste will determine historical income and current outcomes. Hence, while caste fixed effects are included in the equation that predicts $y_0$, they are also included in the equation that estimates nutritional status below. Note that the estimated coefficients can be used to predict $y_0$ even outside the study area because the historical tax revenue data are available for the entire northern Tamil Nadu region.}

To separately measure ancestral income along the male and female line, we take advantage of the fact that marriage in India is patrilocal, with women moving into their husbands’ homes, which are often outside their natal village. Based on data from the SICHS survey, over 80% of women in the study area move outside their natal village when they marry. Given that men do not move when they marry, predicted current income based on the historical revenue tax in an individual’s natal village determines ancestral income along the male line, going back all the way to 1871. Given the assumption that income shocks commence in the grandparents’ generation, this will also be the father’s father’s inherited wealth, $y_{0IM}^M/2$. In contrast, predicted current income based on the historical tax revenue in the mother’s natal village determines ancestral income along the female line; i.e. the mother’s mother’s inherited wealth, $y_{0IW}^M/2$.\footnote{Historical tax revenue in the mother’s natal village determines historical income on the maternal grandfather’s male line, going back to 1871. Given that the income shocks only commence in the grandparents’ generation, this will be the wealth that he inherited. Given positive assortative matching on wealth, this will also be the wealth that the maternal grandmother inherited.}

Having constructed measures of ancestral income that are specific to the household head and his
Figure 13: Nutritional Status and Metabolic Disease with respect to Income (South India)

(a) Height-for-age (0-59 months)
(b) Height-for-age (5-19 years)
(c) BMI
(d) Metabolic disease

Source: India Human Development Survey (IHDS) and South India Community Health Study (SICHS)
Covariates listed in the text are partialled out prior to nonparametric estimation.

wife, along the male and the female line, the next step is to locate the current income threshold at which SICHS households escape the nutrition trap. We do this by implementing the same procedure that was used to locate a threshold earlier, with adult BMI as the outcome, using IHDS data. Prior to that, as a test of internal validity, we verify that the nutritional status-household income and metabolic disease-household income relationships obtained with IHDS data are also obtained with SICHS data in Figures 13a-13d. The same set of covariates that were included in the estimating equation and partialled out prior to nonparametric estimation with the IHDS data are included here as well. To smooth out transitory shocks, we take the average of the household income reported in the SICHS census and the SICHS survey as our measure of household income. As a basis for comparison, we also include the corresponding nonparametric plot obtained with IHDS data for the South Indian states in each figure. The first observation from Figures 13a-13d is that the estimated relationships between each outcome and
The second observation is that the estimated relationships with SICHS and the IHDS South India data match very closely across the income distribution in each figure.

We next proceed to locate an income threshold, with adult BMI as the outcome, using SICHS data. Figure 14 reports the result of Hansen’s threshold test with SICHS data and, for comparison, with IHDS South India data. The F-type statistic used for the test increases steeply as the assumed income threshold moves away (on either side) from the income level at which it is minimized; the location of the threshold can thus be bounded relatively tightly. Notice that the threshold is located at precisely the same point with SICHS and IHDS South India data. Table 3 uses this result to separately estimate the adult BMI - household income relationship above and below the estimated threshold. Columns 1-2 report the estimation results with IHDS South India data; as with the all-India data in Table 1, the relationship is positive and statistically significant above and below the threshold, although it is substantially larger above. Columns 3-4 repeat this exercise with the SICHS data; the results are qualitatively the same, except that the BMI-income relationship below the threshold is no longer statistically significant. Columns 5-6 add ancestral income to the estimating equation. Epigenetic inheritance has traditionally been assumed

---

[34] The SICHS data set is not large enough to locate an income threshold with precision, except for adult BMI as the outcome (reported below). However, a threshold can be located for each outcome using the IHDS data for South India. The threshold location, the baseline slope coefficient, and the slope change coefficient, with bootstrapped standard errors, are reported for each outcome in Appendix Table A5. Reassuringly, the estimated coefficients are very similar to what we obtained above with the all-India data.

[35] Nutritional status is systematically higher with SICHS data relative to IHDS South India data (this can be observed by comparing the range of the Y-axes with each data set in Figures 13a-13c). 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). Once we control for the level, however, the SICHS and the IHDS South India data track very closely with household income for each outcome.
to occur along the female line and, hence, we include our measure of ancestral income based on the 1871 revenue tax in the mother’s natal village as an additional regressor in the estimating equation. The coefficient on this variable is positive and significant below, but not above, the threshold.\textsuperscript{36} The coefficient on current income, in contrast, continues to be positive and statistically significant above (but not below) the threshold.\textsuperscript{37} Recall that ancestral wealth on the female line is based on the 1871 revenue tax in the mother’s natal village, whereas the corresponding statistic based on the individual’s own village measures ancestral income on the male line. Including our measures of ancestral income on both the male and the female line in Table 3, Columns 7-8 we see that it is only ancestral income on the female line that contributes to current nutritional status. Consistent with the traditional view, our analysis indicates that epigenetic inheritance (at least with respect to nutritional status) occurs along the female line.

The results reported above provide support for an epigenetically determined set point, and for epigenetic adaptation and inheritance. As noted, there is little evidence of either in humans. A widely cited study in the epigenetics literature, Lumey (1992), finds that individual who were \textit{in utero} in the 1944-1945 Dutch famine were more likely to have low birth weight children. Although this correlation is consistent with intergenerational epigenetic inheritance, other explanations are available; for example, it is well known that mothers with small frames can have lower birth weight offspring because their reproductive tracts are narrower. Our analysis is also subject to this potential concern because it is based on the mother’s ancestral wealth, which will be correlated with the wealth in her natal home. However, there is an additional dimension to our analysis, which is that ancestral income must only matter below the independently estimated income threshold. This striking result provides the strongest evidence to date that we are aware of in support of epigenetic adaptation – to historically stable local (village-level) economic conditions – and of epigenetic inheritance over multiple generations.

Having validated the structural BMI-income relationship specified by equation (2), we now proceed to equation (3). This equation specifies that the probability of metabolic disease will be constant for individuals who remain at their set point and increasing in the difference between current income, \( y_t \), and ancestral income, \( y_0 \), for individuals who have escaped their set point. Although we have an appropriate measure of \( y_0 \) with the SICHS data, we cannot directly verify the relationship implied by equation (3) because the SICHS sample is too small to accurately locate an income threshold (which separates the two types of individuals) with the risk of metabolic disease as the outcome. What we do instead is to derive

\textsuperscript{36}Based on our model, ancestral income will also determine BMI for some individuals above the estimated income threshold (who remain at their set point). However, their numbers will depend on the initial distribution of income and the evolution of income over time.

\textsuperscript{37}Notice that the sample size declines when we include ancestral income in the estimating equation. This is because respondents in the SICHS survey sometimes did not know the administrative block in which their mother’s natal village was located (this information is needed because many villages have the same name). In addition, some current village names could not be matched with the names in the historical records. Reassuringly, the sample attrition does not appear to be systematic. Appendix Figure A1a reports the nonparametric relationship between adult BMI and household income using the full SICHS sample and the reduced sample (consisting of individuals for whom ancestral income on the female side is available). The relationships with the two samples are almost identical. Appendix Figure A1b reports the corresponding threshold tests. While a threshold can also be located with the reduced sample, it shifts to the left, which explains the decline in the number of observations below the threshold, going from the full sample to the reduced sample in Table 3. Nevertheless, and in line with the view that the reduced sample is not systematically selected, the coefficient on current household income is almost identical in Table 3, Columns 4 and 6.
Table 3: BMI - Income Relationship (below and above the threshold)

<table>
<thead>
<tr>
<th>Dep. Var.:</th>
<th>IHDS South India</th>
<th>BMI</th>
<th>SICHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample relative to threshold:</td>
<td>below</td>
<td>above</td>
<td>below</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Current income</td>
<td>0.200**</td>
<td>0.961***</td>
<td>0.116</td>
</tr>
<tr>
<td>(female line)</td>
<td>(0.084)</td>
<td>(0.073)</td>
<td>(0.251)</td>
</tr>
<tr>
<td>Ancestral income</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(male line)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ancestral income</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(male line)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma )</td>
<td>1.70</td>
<td>1.69</td>
<td>1.55</td>
</tr>
<tr>
<td>[Threshold location]</td>
<td>[1.55, 1.90]</td>
<td>[1.31, 2.05]</td>
<td>[1.00, 2.24]</td>
</tr>
<tr>
<td>( N )</td>
<td>10194</td>
<td>12122</td>
<td>2426</td>
</tr>
</tbody>
</table>

Source: India Human Development Survey (IHDS) and South India Community Health Study (SICHS)
Ancestral income constructed with SICHS data and 1871 revenue tax data.
Covariates listed in the text are included in the estimating equation.
* significant at 10%, ** at 5% and *** at 1%

and estimate the biological relationship between the risk of metabolic disease and BMI that is implied by equation (3), in combination with Proposition 1.

Proposition 1 indicates that there is a positive relationship between adult BMI and household income both below and above the income threshold. Equation (3) specifies that there is no relationship between the risk of metabolic disease and household income below the same threshold and a positive relationship above it. If heterogeneity in household income is the source of forcing variation, then this implies that there will be no relationship between disease and BMI below a BMI threshold (which corresponds to an underlying income threshold) and a positive relationship above the threshold. Figure 15a tests the preceding prediction by nonparametrically estimating the relationship between the probability of metabolic disease and BMI with IHDS all-India and IHDS South India data. The usual set of covariates are included in the estimating equation and partialled out prior to nonparametric estimation using Robinson’s procedure. There appears to be no relationship between the probability of disease and BMI up to a BMI threshold and a positive relationship above the threshold. This is confirmed by Hansen’s threshold test, reported in Figure 15b, where a threshold is located, with tight bounds on the 95% confidence interval, with both all-India and South India data.

The coefficients from piecewise linear equations, estimated with a slope-change at the thresholds located above, and with bootstrapped standard errors in parentheses, are reported in Table 4. The
baseline slope coefficient, measuring the disease-BMI relationship below the estimated threshold, is small in magnitude, and we cannot reject the hypothesis that it is equal to zero with South Indian data. The slope-change coefficient, measuring the change in the disease-BMI relationship above the estimated threshold, is an order of magnitude larger than the baseline slope coefficient and is precisely estimated with both samples. The threshold BMI at which there is a discontinuous slope change is estimated at 21.8 with all-India data and 20.6 with South India data. Although the threshold locations are precisely estimated, there is still substantial overlap in the confidence intervals and thus we cannot statistically reject the hypothesis that the threshold is the same with the two samples. Even if we take the (higher) threshold location, obtained with all-India data, as the benchmark, a BMI of 21.8 is well within the normal range (18.5-25). The risk of metabolic disease increases discontinuously at an extremely low BMI in the Indian population, and we will return to this observation in the concluding section of the paper.38

3.5 Structural Estimation and Quantification

Having tested and validated the model, we proceed to estimate the key structural parameter, $b$, in the model. Recall from equation (2) that there is a linear structural relationship, with slope $b$, between BMI, $z_t$, and income, both below and above the threshold, with the relevant income measure switching from $y_0$ to $y_t$. Based on the specified distribution of income shocks, the resulting cross-sectional $z_t - y_t$ relationship

38Our model does not precisely characterize the relationship between the risk of metabolic disease and BMI in the absence of a set point. Based on the observed relationship in developed economies, we expect that it will be positive and continuous, starting at a low level of risk, which only increases steeply above a relatively high BMI (around 25). Thus, while individuals who remain at their set point are protected, it would appear that those individuals would be at low risk even in the absence of a set point. At the same time, the set point substantially increases the risk of metabolic disease for normal weight individuals, with BMI in the 22-25 range, who have escaped the nutrition trap.
### Table 4: Piecewise Linear Equation Estimates: Metabolic Disease - BMI

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Pr [metabolic disease]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline slope (β₁)</strong></td>
<td>IHDS All India</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>0.003**</td>
<td>0.001</td>
</tr>
<tr>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td><strong>Slope change (β₂)</strong></td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>0.006**</td>
<td>0.007**</td>
</tr>
<tr>
<td>(0.001)</td>
<td>(0.002)</td>
</tr>
<tr>
<td><strong>Threshold location (γ)</strong></td>
<td></td>
</tr>
<tr>
<td>21.80</td>
<td>20.60</td>
</tr>
<tr>
<td>[20.20, 22.80]</td>
<td>[18.80, 22.20]</td>
</tr>
<tr>
<td><strong>Threshold test p–value</strong></td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Mean of dependent variable</strong></td>
<td>0.066</td>
</tr>
<tr>
<td>N</td>
<td>76103</td>
</tr>
</tbody>
</table>

Source: India Human Development Survey (IHDS)

BMI is the independent variable.

Covariates listed in the text are included in the estimating equation.

* significant at 10%, ** at 5% and *** at 1%

---

is derived in equations (4) and (5) as follows:

\[ \pi(y_t | y_t \leq \alpha) = a + b(y_t - e^L(y_t)) \]

\[ \pi(y_t | y_t > \alpha) = a + b(y_t - e^H(y_t)) \]

Closed-form expressions for \(e^L(y_t), e^H(y_t)\), as functions of \(y_t, \mu_t, \sigma_t^2\), and \(\alpha\) are derived in equations (6) and (7). The \(\alpha\) parameter can be estimated from the location of the threshold. Based on the discussion that followed Figure 12, we assume that \(t = 3\) in the current generation; i.e. that the income shocks that accompany economic development commenced in the grandparents’ generation. Recall that \(\mu_t = t\mu\) and \(\sigma_t^2 = t\sigma^2\); it then follows that if the parameters of the distribution of income shocks, \(u_t \sim N(\mu, \sigma^2)\) can be estimated, then \(e^L(y_t), e^H(y_t)\) can be computed for any level of current income, \(y_t\). Once these adjustment terms are included in the estimating equation, the structural slope parameter, \(b\), can be independently estimated, below and above the income threshold.

To estimate the parameters of the distribution of income shocks, 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, we use the (real) income distribution in 1951, 1981, and 2011 and, in particular, the change in these distributions, to estimate the \(\mu\) and \(\sigma\) parameters.\(^{39}\)

\(^{39}\)The World Inequality Database provides the 99 fractiles of the income distribution; \(p_0p_{91} \ldots, p_{80}p_{99}\), 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.
Table 5: Structural Parameter Estimates

<table>
<thead>
<tr>
<th>Sample: IHDS All India</th>
<th>HFA 5-19</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep. variable:</td>
<td>without correction</td>
<td>with correction</td>
</tr>
<tr>
<td>Specification:</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Slope below threshold ( (\beta_L) )</td>
<td>0.011</td>
<td>0.132***</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Slope above threshold ( (\beta_H) )</td>
<td>0.221***</td>
<td>0.166***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>( F )-statistic ( (\beta_L = \beta_H) )</td>
<td>44.69</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>[0.000]</td>
<td>[0.374]</td>
</tr>
<tr>
<td>Imposed threshold</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>N</td>
<td>48846</td>
<td>48846</td>
</tr>
</tbody>
</table>

Source: India Human Development Survey (IHDS)

Logarithm of household income (with and without adjustment term) is the independent variable. Covariates listed in the text are included in the estimating equation.

* significant at 10%, ** at 5% and *** at 1%

Table 5 reports coefficient estimates from a piecewise linear equation, using IHDS all-India data, with child (aged 5-19) height-for-age in Columns 1-2 and adult BMI in Columns 3-4 as outcomes. In addition to household income, the usual covariates are included in each estimating equation. The slope-change in the estimating equation is imposed at the income level where the threshold was previously located, separately for each outcome. Columns 1 and 3 report benchmark estimates without including the \( e_L(y_t), e_H(y_t) \) adjustment terms. This specification is essentially the same as what we estimated earlier in Table 1, except that we now report the slopes below and above the threshold (rather than the slope-change). Columns 2 and 4 report estimates with the adjustment terms included in the estimating equation. Although we can easily reject the null hypothesis that the slopes below and above the threshold are equal in Columns 1 and 3, without the adjustment, we cannot reject the null once the adjustment terms are included.

When \( e_L(y_t), e_H(y_t) \) are included in the estimating equation, the slope coefficients can be interpreted as the structural, \( b \), parameter in the model. We impose the restriction in the model that the structural 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 \( \mu \) and \( \sigma^2 \) 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 \( \mu \) and \( \sigma^2 \) for which the simulated income distribution in 1981 and 2011 matches most closely with the actual distribution.

40 Appendix Table A6 repeats this exercise under alternative assumptions about the onset of economic development; i.e. by setting \( t = 2, 3, 4, \) or 5. Providing support for the maintained assumption that \( t = 3 \), we see that it is only with this specification that the estimated slopes are the same below and above the threshold, with both child and adult nutrition status as outcomes. Moreover, for those outcomes with \( t = 4 \) or 5 where the estimated slope is the same below and above the threshold, the point estimates are very similar to what we obtain with \( t = 3 \).
nutritional status - income relationship is the same below and above the threshold. The estimates reported in Columns 2 and 4 indicate that the restriction we have imposed is supported by the data. Moreover, as implied by Proposition 1, the slope without the adjustment term is less than (greater than) $b$, below (above) the threshold.

One benefit of the structural estimation is that it allows us to test restrictions that are imposed on the data by the model. An additional benefit is that it allows us to quantify the consequences of the nutrition trap. If the set point is irrelevant, there will be a linear relationship between household income and nutritional status: $\tau = a + by_t$. The estimated $b$ parameter can thus be used to predict what nutritional status would have been in the absence of the nutrition trap. Figure 16a reports actual height-for-age, predicted height-for-age (based on the model), and the counter-factual height-for-age (in the absence of the nutrition trap) for children aged 5-19. Figure 16b reports the corresponding relationships with adult BMI as the outcome. The usual set of covariates are partialled out, and the dotted vertical line in each figure marks the location of the income threshold. Based on these estimates, the fraction of stunted children (with a z-score below -2) would decline by 30% and the fraction of underweight adults (with a BMI below 18.5) would decline by 50% if the set point were absent.41 The dampening of the nutritional status-income relationship below the threshold, which we attribute to a predetermined set point, has important consequences for child and adult nutritional status in India, and we will return to this point in the concluding section of the paper.

41These statistics are based on a comparison of predicted and counter-factual malnutrition, taking account of the independent impact of the covariates.
4 External Validity

4.1 Micro Evidence Across Countries

The presence of a set point is evidently not unique to India. The next step in the analysis is thus to assess the applicability of the model to other developing countries. To test the model, the following data are required: (i) Household income, preferably at multiple points in time. (ii) Nutritional status of adults and children. (iii) Indicators of metabolic disease. (iv) Household composition and detailed geographical indicators. The additional constraint is that a large sample is needed to locate a slope-change with precision. India is unusual in that two independent data sets are available that satisfy this requirement. A search of publicly available data sets from other countries recovered just two data sets that are suitable to test our model: the Indonesia Family Life Survey (IFLS) and the Ghana Socioeconomic Panel Survey (GSPS), although the GSPS does not contain information on metabolic disease.\(^{42}\) We thus proceed to test the model with these two data sets, just as we did with the IHDS and SICHS for India.

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

At what stage in the development process are Indonesia and Ghana or, equivalently, how does current income in those countries compare with historical income? Although income data from the Madison Project Database for African countries only go back to 1950, adult height is available for many developing countries as far back as the nineteenth century. It is standard practice to use adult height as a proxy for income, and the standard of living, in historical research (Steckel, 1995).\(^{43}\) Figure 17 thus plots the relationship between current per capita GDP and adult height in 1900 for a number of developing countries, including India, Indonesia, and Ghana.\(^{44}\) The first point to take away from the figure is that there has been a reversal of fortunes over the past century, reflected by the negative relationship between current income and our proxy for historical income. The second point to take away from the figure is that the mismatch between current income and historical income is greater in Asia than in Africa.

\(^{42}\)Other well known data sets that we considered, but were determined to be unsuitable, include the Demographic Health Survey (DHS), the Living Standards Measurement Study (LSMS), Young Lives, and the China Health and Nutrition Survey (CHNS).
\(^{43}\)As noted by Steckel (1995) and Deaton (2007), genes are important determinants of individual height (and nutritional status more generally) but cannot explain variation across populations.
\(^{44}\)We include all countries in South and South East Asia and Sub-Saharan Africa that satisfy the following requirement: their GDP per capita must be less than $12,000, which roughly corresponds to the upper bound for lower-middle income countries set by the World Bank. The same criterion is applied in the cross-country figures that follow.
Figure 17: Current and Historical Income Across Countries

![Graph showing current and historical income across countries.]

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

two regions that we will focus on in the macro analysis.\textsuperscript{45} Based on these aggregate statistics, a larger fraction of the population is likely to have escaped the nutrition trap in Asian countries relative to African countries. Focusing on specific countries, we would expect the tests of the model to generate similar results for India and Indonesia. In contrast, we would expect a larger fraction of the population in Ghana to have remained at its set point. If we do locate a threshold in that country, it will be relatively far to the right in the income distribution.

Figure 18a nonparametrically estimates the relationship between children’s nutritional status and household income, separately for children aged 0-59 months and 5-19 years, using Indonesia Family Life Survey (IFLS) data. Figure 18b reports the corresponding nonparametric relationships between adult BMI, the probability of metabolic disease, and household income. The same set of covariates that were included in the estimating equation with Indian data are included here as well and, as usual, are partialled out 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. Visual inspection of Figures 18a and 18b indicates that the relationship with Indonesian data look very similar to what we obtained with Indian data; there is a weak or absent relationship between household income and each outcome at low levels of household income and a steeper relationship at higher income levels.

Figure 18c reports the result of Hansen’s threshold test with children’s nutritional status as outcomes, while Figure 18d reports the results of the test with nutritional status and the risk of metabolic disease for

\textsuperscript{45}Deaton (2007) and Deaton and Drèze (2009) note that real per capita incomes were historically lower in South Asia than in Africa. However, incomes in South Asia have been growing relatively rapidly since the 1980’s and now exceed those in Africa.
adults as outcomes. Once again, the results are very similar to what we obtained earlier with Indian data. A threshold is located with a high degree of statistical precision with each outcome. Our best estimate of the threshold location matches almost exactly for the younger and older children and is very close for the two adult outcomes. Table 6 reports coefficient estimates, with bootstrapped standard errors, from a piecewise linear regression with a slope-change at the estimated threshold. As with the IHDS data, we can easily reject the null hypothesis that there is no slope-change at an income threshold with each outcome. In addition, the confidence intervals for the threshold location overlap with each pair of outcomes. The baseline slope coefficients and the slope-change coefficients are broadly the same as what we estimated with Indian data, except that the baseline slope coefficient with adult BMI as the dependent variable and the slope-change coefficient with the risk of metabolic disease as the dependent variable are no longer significant at conventional levels.
Table 6: Piecewise Linear Equation Estimates: nutritional status and metabolic disease (Indonesia)

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>HFA 0-59 (1)</th>
<th>HFA 5-19 (2)</th>
<th>BMI (3)</th>
<th>Disease (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline slope ($\beta_1$)</td>
<td>-0.021 (0.031)</td>
<td>0.027 (0.018)</td>
<td>0.041 (0.054)</td>
<td>-0.002 (0.009)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.127** (0.045)</td>
<td>0.081** (0.019)</td>
<td>0.494** (0.059)</td>
<td>0.019 (0.009)</td>
</tr>
<tr>
<td>Threshold test $p$-value</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.020</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>-1.377</td>
<td>-1.420</td>
<td>23.502</td>
<td>0.163</td>
</tr>
<tr>
<td>N</td>
<td>7771</td>
<td>18375</td>
<td>33219</td>
<td>25422</td>
</tr>
</tbody>
</table>

Source: Indonesia Family Life Survey (IFLS)

Disease indicates whether the individual has been diagnosed with diabetes, hypertension, or cardiovascular disease.

Logarithm of household income is the independent variable.

Covariates listed in the text are included in the estimating equation.

* significant at 10%, ** at 5% and *** at 1%

Figures 19a and 19b report the nonparametric relationship between nutritional status (separately for children and adults) and household income, using data from the Ghana Socioeconomic Panel Survey (GSPS). As noted, the GSPS does not collect data on metabolic disease. However, the full set of covariates that were used in the Indian and Indonesian analyses are available and can be partialled out prior to nonparametric estimation. The GSPS was conducted in two waves; 2009-2010 and 2013. As usual, the outcomes are measured in both waves, while household income is averaged over the two waves. In contrast with the nonlinear income effects that we estimated with each outcome using Indian and Indonesian data, nutritional status appears to be increasing continuously in Figure 19a and 19b. Formal statistical support for this observation is provided in Figures 19c and 19d, where the Hansen test is unable to detect an income threshold with any outcome. As reported in Table 7, there is a positive and statistically significant relationship between household income and nutritional status, both for children and adults. Where the Ghana data differ from the Indian and Indonesian data is that there is no slope change. Our interpretation of this finding, which is in line with the fact that current and historical incomes are relatively close in Africa, is that the bulk of the Ghanaian population remains at its pre-modern set point. Consistent with this interpretation, the income coefficient in Table 7, Column 3, with adult BMI as the outcome, is similar in magnitude to the income coefficient to the left of the threshold with Indian data, reported in Table 1, Column 3.46

46We focus on this comparison because the income coefficient to the left of the threshold is less precisely estimated for all other nutritional status outcomes, with Indian and Indonesian data.
The nutrition-income puzzle that Deaton (2007) uncovered is that nutritional status in South Asia is lower than what would be predicted by GDP per capita, whereas the reverse is true for Africa. To explain this stylized fact through the lens of our set point model, consider a variant of the model that is adapted to a cross-country setting with aggregate data. We make the following assumptions: (i) A fixed fraction of the population, \( \pi \), remains within its set point in each country, \( j \), in the current time.

Deaton considers a number of reasons for this stylized fact: First, he considers the possibility that there may be genetic differences across populations. He rules this out by noting that both South Asians and Africans who migrate to advanced economies quickly converge to the nutritional level of the native population (within a couple of generations). Second, he considers the possibility that the disease environment, associated with diarrheal disease, is especially unfavorable in South Asia. Using child and infant mortality as proxies for diarrheal disease, Deaton finds no evidence in support of this mechanism. Finally, Deaton considers, and rules out, the possibility that Africans have higher caloric intake than South Asians, despite having lower per capita income.
period, \( t \). \(^{48}\) (ii) Log income, \( y_t^j \sim N(\mu_t^j, \sigma_t^2) \). (iii) Each dynasty in country \( j \) has the same income, \( y_0^j \), in the initial period, 0. Given these assumptions, and taking advantage of the properties of the normal distribution, average BMI in country \( j \) in the current period, \( z_t^j \), can be expressed as a weighted average of initial income, \( y_0^j \), and average current income, \( \mu_t^j \): \(^{49}\)

\[
\begin{align*}
    z_t^j &= a + b \left[ \pi y_0^j + (1 - \pi) \left( \mu_t^j + \sigma_t \phi \left( \Phi^{-1}(\pi; 0, 1); 0, 1 \right) \right) \right] \\
    &= \mu_t^j + \pi \phi \left( \Phi^{-1}(\pi; 0, 1); 0, 1 \right) + (1 - \pi) \mu_t^j - \pi \phi \left( \Phi^{-1}(\pi; 0, 1); 0, 1 \right)
\end{align*}
\]

(9)

Taking expectations conditional on \( \mu_t^j \), \( E(z_t^j | \mu_t^j) \) is increasing in \( E(y_0^j | \mu_t^j) \). Looking back at Figure 17, if we drew a horizontal line through the figure at any level of current (average) income, it is evident that historical heights (which proxy for historical incomes) would be higher for African countries. \( E(y_0^j | \mu_t^j) \) is higher in Africa, which implies that \( E(z_t^j | \mu_t^j) \) is higher in Africa from equation (9). \(^{50}\) Figure 20 tests this hypothesis by plotting average BMI against current GDP per capita. Drawing a vertical line through the figure at any level of current income, BMI is higher in African countries than in Asian countries. The same result (not reported) would be obtained if we replaced adult BMI with the fraction of children that are (not) stunted or with adult height (the measure used by Deaton). Our model, based on a biological friction, is able to explain the well documented differences in nutritional status, conditional on income, 

\(^{48}\) This simplifying assumption is evidently at odds with the preceding discussion and we will discuss the consequences of relaxing it below.  

\(^{49}\) Let \( y^*_t \) denote the income threshold above which households escape their set point. \( \pi = Pr[y_t^j \leq y^*_t] = \Phi(y^*_t; \mu_t^j, \sigma_t^2) \). By the property of the normal distribution, \( y^*_t = \Phi^{-1}(\pi; \mu_t^j, \sigma_t^2) = \mu_t^j + \sigma_t \Phi^{-1}(\pi; 0, 1) \). By the property of the normal distribution, once again, and substituting the expression for \( y^*_t \) derived above, average income above the threshold can be expressed as:

\[
E[y_t^j | y_t^j \leq y^*_t] = \mu_t^j + \pi \phi \left( \Phi^{-1}(\pi; 0, 1); 0, 1 \right)
\]

\(^{50}\) Although our simple macro model assumes that \( \pi \) is the same in all countries, the fraction of the population that remains at the pre-modern set point will in general be larger in African countries because \( \mu_t^j - y_0^j \) will be smaller on average for them. The condition that now needs to be satisfied, from equation (9), is that \( \mu_t^j - \pi (\mu_t^j - y_0^j) \) should be larger for an African country relative to an Asian country with the same \( \mu_t^j \). Nevertheless, we still expect \( E(z_t^j | \mu_t^j) \) to be higher in Africa as long as the direct effect of \( \mu_t^j - y_0^j \) dominates the indirect effect, through \( \pi \).
between South Asia and Africa. Indeed, it can explain the wider difference between Africa and Asia, not just South Asia, as observed in Figures 17 and 20.

Although other mechanisms have been proposed to explain Deaton’s puzzle, an appealing feature of our mechanism, based on a biologically determined set point is that it also has implications for the emergence of metabolic diseases during the process of economic development. The micro evidence, presented above, indicates that the risk of these diseases increases when (normal weight) individuals escape the nutrition trap. While we expect to observe this phenomenon in any developing economy, the prevalence of metabolic disease at a particular point in time will depend on the fraction of the population that has escaped the nutrition trap, together with the mismatch between current income and historical income for those who have escaped. We would naturally expect these conditions to vary across populations, and the literature has indeed identified large differences in the prevalence of diabetes and related metabolic conditions. As with the nutrition literature, South Asians have received disproportionate attention. While diabetes was virtually nonexistent in South Asia until a few decades ago, rapid economic growth in India in particular has been accompanied by a substantial increase in the prevalence of the disease among normal weight adults (Ramachandran, 2005; Narayan, 2017).

Making the same assumptions as above, the aggregate version of the disease-income relationship specified in equation (3) can be expressed as:

$$D_j^t = \Gamma_1 + \Gamma_2 (1 - \pi) \left[ \mu_j + \sigma_t \phi \left[ \Phi^{-1}(\pi; 0, 1); 0, 1 \right] - y_0^j \right],$$  \hspace{1cm} (10)$$

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

Taking expectations conditional on average BMI, \( z_j^t \), in equation (10), \( E(D_j^t | z_j^t) \) is increasing in \( E(\mu_j^t - y_{j0}^t | z_j^t) \). Recall from Figure 17 that for any level of average current income, \( \mu_j^t \), average historical income, \( y_{j0}^t \), is higher in African countries than in Asian countries. We know from equation (9) that \( z_j^t \) is a weighted average of \( \mu_j^t \) and \( y_{j0}^t \). Thus, if an African and Asian country have the same average BMI, then the Asian country must have higher \( \mu_j^t \) and lower \( y_{j0}^t \). Based on this argument, \( E(\mu_j^t - y_{j0}^t | z_j^t) \) is higher in Asia than in Africa and, hence, \( E(D_j^t | z_j^t) \) must be higher as well.\(^{51}\)

Figure 21 tests this prediction by plotting diabetes prevalence against average BMI. Drawing a vertical line through the figure at any BMI level, diabetes is higher in Asian countries than in African countries. Notice that while India is somewhat of an outlier in the figure, other Asian countries are even bigger outliers and not all of them are South Asian. Although the diabetes literature has tended to focus on South Asians as a particularly vulnerable group, our analysis, as with the analysis of the nutritional status - income relationship, indicates that inter-regional differences in diabetes prevalence extend to the Asian continent as a whole.\(^{52}\)

\(^{51}\)Although we assume for simplicity that the fraction of the population that has escaped the nutrition trap is the same in all countries, we noted above that \( \pi \) would be higher in Asia (because \( \mu_j^t - y_{j0}^t \) is higher in that region). This adjustment would increase the fraction of the population at risk of metabolic disease in Asian countries and reinforce the prediction that \( E(D_j^t | z_j^t) \) will be higher in Asia than in Africa.

\(^{52}\)It is well known that the Asian phenotype is characterized by relatively high central adiposity (abdominal fat), which in turn is associated with increased risk of diabetes; e.g. Gujral et al. (2013). This may reflect the epigenetic structure of that population, associated with a relatively low BMI set point, rather than a particular genotype. As with differences in nutritional status across populations, only 5-10% of Type-2 diabetes risk can be attributed to genetic factors (Voight et al., 2010).
Conclusion

This paper provides a unified explanation for two stylized facts: (i) the relatively weak relationship between nutritional status and income in developing countries, and (ii) the increased prevalence of metabolic disease (diabetes, hypertension, cardiovascular disease) among normal weight individuals with economic development. Our explanation is based on a set point for body weight or BMI, which is adapted to economic conditions in the pre-modern economy, but which fails to subsequently adjust to rapid economic change. This implies that during the process of development, the population will be divided into two distinct groups: Individuals in the first group remain at their set point BMI, despite the increase in their consumption, and are (partly) responsible for the weak relationship between nutritional status and current income. Individuals in the second group, who have escaped the nutrition trap, but are not necessarily overweight, are the primary contributors to the increased risk of metabolic disease.

To test this theory, we develop a model of nutrition and health in which the presence of an epigenetically determined set point is taken as given. The implications of this model are validated with micro-data from multiple countries; India, Indonesia, and Ghana. In addition, we use unique data, recently made available from the South India Community Health Study (SICHS), to verify the structural relationships underlying the model; in particular, we document that adult BMI is determined by ancestral income on the female line alone for households below an estimated income threshold (who are at their set point), whereas adult BMI is determined exclusively by current income for households above the threshold (who have escaped the nutrition trap). These findings provide the strongest evidence to date of epigenetic adaptation (to local economic conditions in the pre-modern economy) and inheritance (over multiple generations). To complete the analysis, the model is adapted to aggregate data, allowing us to simultaneously explain why nutritional status in Africa (Asia) is higher (lower) than what would be predicted by current GDP per capita, as well as why there is higher prevalence of diabetes, for given BMI, in Asian versus African countries.

Our structural estimates and accompanying counter-factual simulations for India, a country where both stylized facts have been well documented, indicate that stunting among 5-19 year olds would have declined by 30% and the fraction of underweight adults (with BMI below 18.5) in the population would have declined by 50% in the absence of a set point. Malnutrition is associated with physical and cognitive under-development among children and physiological and psychological impairment among adults (Dasgupta and Ray, 1986; Dasgupta, 2013). While nutrition programs are an obvious solution to this problem, it has been observed that such programs are often ineffective (Duh and Spears, 2017). The insight from our analysis is that nutrition programs in developing countries will only be successful if they are intense enough and sustained enough to move individuals out of their set point.

Nutritional status will inevitably improve with economic development and an increasing fraction of the population will escape the nutrition trap. Our analysis indicates that this nutritional transition will be accompanied by an increased incidence of metabolic disease. Although the set point is predetermined, the nutrition transition can be smoothed if the population compensates for the increased risk of metabolic disease by adopting a healthier lifestyle. It is thus imperative that governments in developing countries,
which are likely to face an epidemic of metabolic disease in the coming decades, take adequate steps to improve the prevention and treatment of these conditions. Screening will be an important component of these programs, and successful screening requires the at-risk population to be accurately identified. It has been recommended that the lower bound for the overweight range in Asian populations be reduced from 25 to 23, to account for the fact that these populations are at elevated risk of metabolic disease at lower BMI (Deurenberg-Yap et al., 2002; Pan et al., 2004). Our analysis, based on rigorous statistical tests and conducted with representative Indian data, indicates that there is a discontinuous increase in the risk of metabolic disease at a BMI below 22. The estimated threshold is even lower for South India, at a BMI below 21. The public health challenge faced by countries like India, which will need to successfully navigate the nutrition-disease tradeoff over the next couple of generations, may be even greater than what is currently envisaged.

References


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

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>HFA 0-59 (1)</th>
<th>HFA 5-19 (2)</th>
<th>BMI (3)</th>
<th>Disease (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline slope</td>
<td>-0.032</td>
<td>-0.007</td>
<td>0.183**</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.030)</td>
<td>(0.031)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.334**</td>
<td>0.178**</td>
<td>0.856**</td>
<td>0.030**</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.030)</td>
<td>(0.048)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Threshold location ($\gamma$)</td>
<td>1.30</td>
<td>1.40</td>
<td>1.80</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>[1.05, 1.80]</td>
<td>[1.25, 1.60]</td>
<td>[1.70, 1.85]</td>
<td>[2.05, 2.35]</td>
</tr>
<tr>
<td>Threshold test $p$-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>-1.991</td>
<td>-1.652</td>
<td>21.996</td>
<td>0.067</td>
</tr>
<tr>
<td>N</td>
<td>21534</td>
<td>46545</td>
<td>76189</td>
<td>146287</td>
</tr>
</tbody>
</table>

*Source: India Human Development Survey (IHDS)*

Covariates listed in the text are included in the estimating equation.

Bootstrapped standard errors are in parentheses.

Bootstrapped 95% confidence bands for the threshold location are in brackets.

* significant at 10%, ** at 5% and *** at 1%

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

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>HFA 0-59 (1)</th>
<th>HFA 5-19 (2)</th>
<th>BMI (3)</th>
<th>Disease (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline slope ($\beta_1$)</td>
<td>-0.071</td>
<td>0.045</td>
<td>0.294**</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.116)</td>
<td>(0.034)</td>
<td>(0.062)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.345**</td>
<td>0.188**</td>
<td>0.861**</td>
<td>0.036**</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.035)</td>
<td>(0.074)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Threshold location ($\gamma$)</td>
<td>1.30</td>
<td>1.55</td>
<td>1.60</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>[0.75, 1.85]</td>
<td>[1.25, 2.05]</td>
<td>[1.50, 1.75]</td>
<td>[1.70, 2.05]</td>
</tr>
<tr>
<td>Threshold test $p$-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>-1.900</td>
<td>-1.578</td>
<td>22.189</td>
<td>0.098</td>
</tr>
<tr>
<td>N</td>
<td>10363</td>
<td>35764</td>
<td>53005</td>
<td>74166</td>
</tr>
</tbody>
</table>

*Source: India Human Development Survey (IHDS)*

Covariates listed in the text are included in the estimating equation.

Bootstrapped standard errors are in parentheses.

Bootstrapped 95% confidence bands for the threshold location are in brackets.

* significant at 10%, ** at 5% and *** at 1%
Table A3: Piecewise Linear Equation Estimates (adult education and household composition included as additional covariates)

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>HFA 0-59 (1)</th>
<th>HFA 5-19 (2)</th>
<th>BMI (3)</th>
<th>Disease (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline slope ($\beta_1$)</td>
<td>-0.024 (0.076)</td>
<td>0.044 (0.040)</td>
<td>0.286** (0.049)</td>
<td>0.002 (0.002)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.275** (0.074)</td>
<td>0.134** (0.038)</td>
<td>0.398** (0.059)</td>
<td>0.011** (0.003)</td>
</tr>
<tr>
<td>Threshold location ($\gamma$)</td>
<td>1.40 [1.00, 2.20]</td>
<td>1.50 [1.05, 1.95]</td>
<td>1.65 [1.40, 1.90]</td>
<td>1.95 [1.65, 2.40]</td>
</tr>
<tr>
<td>Threshold test $p$-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>-1.991</td>
<td>-1.649</td>
<td>22.002</td>
<td>0.067</td>
</tr>
<tr>
<td>N</td>
<td>21634</td>
<td>48846</td>
<td>76949</td>
<td>147729</td>
</tr>
</tbody>
</table>

Source: India Human Development Survey (IHDS)
Covariates listed in the text are included in the estimating equation. Additional covariates include average adult male and female education, as well as household composition (number of children, teens, adults, and the number of adults engaged in physical labor). Bootstrapped standard errors are in parentheses. Bootstrapped 95% confidence bands for the threshold location are in brackets. * significant at 10%, ** at 5% and *** at 1%

Table A4: Piecewise Linear Equation Estimates (alternative nutritional status measures)

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>WFA 0-59 (1)</th>
<th>WFA 5-19 (2)</th>
<th>Height (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline slope ($\beta_1$)</td>
<td>-0.052 (0.055)</td>
<td>-0.004 (0.029)</td>
<td>0.191 (0.111)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.287** (0.055)</td>
<td>0.331** (0.036)</td>
<td>0.836** (0.121)</td>
</tr>
<tr>
<td>Threshold location ($\gamma$)</td>
<td>1.30 [1.05, 2.20]</td>
<td>1.75 [1.55, 1.95]</td>
<td>1.45 [1.30, 1.65]</td>
</tr>
<tr>
<td>Threshold test $p$-value</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>-1.512</td>
<td>-1.634</td>
<td>22.002</td>
</tr>
<tr>
<td>N</td>
<td>24843</td>
<td>23030</td>
<td>76949</td>
</tr>
</tbody>
</table>

Source: India Human Development Survey (IHDS)
Nutritional status for children is measured by weight-for-age (WFA) and for adults by height. Covariates listed in the text are included in the estimating equation. Bootstrapped standard errors are in parentheses. Bootstrapped 95% confidence bands for the threshold location are in brackets. * significant at 10%, ** at 5% and *** at 1%
### Table A5: Piecewise Linear Equation Estimates (IHDS South India)

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>HFA 0-59</th>
<th>HFA 5-19</th>
<th>BMI</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td></td>
</tr>
<tr>
<td>Baseline slope ($\beta_1$)</td>
<td>0.014</td>
<td>0.025</td>
<td>0.200**</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.026)</td>
<td>(0.095)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Slope change ($\beta_2$)</td>
<td>0.564</td>
<td>0.464*</td>
<td>0.803**</td>
<td>0.025**</td>
</tr>
<tr>
<td></td>
<td>(0.442)</td>
<td>(0.031)</td>
<td>(0.116)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Threshold location ($\gamma$)</td>
<td>2.60</td>
<td>2.90</td>
<td>1.70</td>
<td>2.00</td>
</tr>
<tr>
<td></td>
<td>[1.30, 3.10]</td>
<td>[2.65, 3.20]</td>
<td>[1.50, 1.90]</td>
<td>[1.70, 2.30]</td>
</tr>
<tr>
<td>Threshold test $p$-value</td>
<td>0.012</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Mean of dependent variable</td>
<td>-1.792</td>
<td>-1.655</td>
<td>22.186</td>
<td>0.066</td>
</tr>
<tr>
<td>N</td>
<td>4291</td>
<td>9934</td>
<td>22316</td>
<td>40869</td>
</tr>
</tbody>
</table>

**Source:** India Human Development Survey (IHDS)

Covariates listed in the text are included in the estimating equation.

Bootstrapped standard errors are in parentheses.

Bootstrapped 95% confidence bands for the threshold location are in brackets.

* significant at 10%, ** at 5% and *** at 1%

### Table A6: Structural Parameter Estimates (varying the onset of economic development)

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>HFA 5-19 Years</th>
<th>BMI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current generation:</td>
<td>t = 2</td>
<td>t = 3</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Slope below ($\beta_L$)</td>
<td>0.071***</td>
<td>0.132***</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Slope above ($\beta_H$)</td>
<td>0.214***</td>
<td>0.166***</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>$F$-statistic ($\beta_L = \beta_H$)</td>
<td>20.477</td>
<td>0.797</td>
</tr>
<tr>
<td></td>
<td>[0.000]</td>
<td>[0.372]</td>
</tr>
<tr>
<td>Imposed threshold</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>N</td>
<td>48846</td>
<td>48846</td>
</tr>
</tbody>
</table>

**Source:** India Human Development Survey (IHDS)

Standard errors are in parentheses and $p$-values are in square brackets.

* significant at 10%, ** at 5% and *** at 1%
Figure A1: BMI - Income Relationship

(a) Nonparametric relationships BMI - Income

(b) Threshold tests

Source: South India Community Health Study (SICHS).
Covariates listed in the text are partialled out prior to nonparametric estimation.
B Mathematical Appendix

Proof of Proposition 1: We first derive closed-form expressions for \( e^L(y_t) \), \( e^H(y_t) \). Focusing on the numerator of the \( e^L(y_t) \) expression in (4), we can write

\[
\int_{-\infty}^{y_t} U_t \phi(U_t; \mu_t, \sigma_t^2) \, dU_t = \int_{-\infty}^{y_t} U_t \frac{1}{\sqrt{2\pi} \sigma_t} \exp \left[ -\frac{1}{2} \left( \frac{U_t - \mu_t}{\sigma_t} \right)^2 \right] \, dU_t \\
= \int_{-\infty}^{y_t - \mu_t/\sigma_t} (\sigma_t x_t + \mu_t) \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \frac{x_t^2}{\sigma_t^2} \right] \, dx_t
\]

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 \( \frac{d \phi(x_t; 0, 1)}{dx_t} = -x_t \phi(x_t; 0, 1) \). A similar transformation of \( \Phi(y_t; \mu_t, \sigma_t^2) \) in the denominator of the \( e^L(y_t) \) expression in (4) gives us the closed-form expression for \( e^L(y_t) \) in equation (6). The corresponding expression for \( e^H(y_t) \) in equation (7) is derived by replacing \( y_t \) with \( \alpha \) 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 (6) in equation (4) and differentiate with respect to \( y_t \). Given the properties of the inverse Mill's ratio,

\[
\frac{d \bar{z}(y_t|y_t < \alpha)}{d y_t} = b \left[ 1 + \Phi \left( \frac{y_t - \mu_t}{\sigma_t} \right) \right] \in (0, b)
\]

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 (7), that the numerator is independent of \( y_t \) and the denominator is increasing in \( y_t \). Hence, \( \frac{d e^H(y_t)}{d y_t} < 0 \), which implies \( \frac{d \bar{z}(y_t|y_t > \alpha)}{d y_t} > b \).

Note, from equations (6) and (7), that \( e^L(y_t) = e^H(y_t) \) at \( y_t = \alpha \), and thus, from equations (4) and (5), 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

\[
\lim_{y_t \uparrow \alpha} \frac{d \bar{z}(y_t|y_t \leq \alpha)}{d y_t} \neq \lim_{y_t \downarrow \alpha} \frac{d \bar{z}(y_t|y_t > \alpha)}{d y_t}
\]

From equations (4) and (5), a necessary and sufficient condition for the preceding inequality to be satisfied is that \( \frac{d e^L(y_t)}{d y_t} \neq \frac{d e^H(y_t)}{d y_t} \) at \( y_t = \alpha \). Using equations (6) and (7), 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 (6), \( e^L(y_t) = \frac{\mathcal{L}(v_t)}{\Phi(v_t; 0, 1)} \), where \( \mathcal{L}(v_t) = \mu_t \Phi(v_t; 0, 1) - \sigma_t \phi(v_t; 0, 1) \). From equation (7), \( e^H(y_t) = \frac{\mathcal{L}(v_t)}{\Phi(v_t; 0, 1)} \) 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 \( \frac{d e^L(y_t)}{d y_t} \neq \frac{d e^H(y_t)}{d y_t} \) is that \( \frac{d \mathcal{L}(v_t)}{d y_t} \neq \frac{d \mathcal{L}(v_t)}{d y_t} \) at \( y_t = \alpha \). \( \frac{d \mathcal{L}(v_t)}{d y_t} = 0 \). From the property of the standard normal distribution, \( \phi'(v_t; 0, 1) = -v_t \phi(v_t; 0, 1) \), and, hence, \( \frac{d \mathcal{L}(v_t)}{d y_t} \bigg|_{y_t=\alpha} = \frac{\alpha}{\sigma_t} \phi(\overline{v}; 0, 1) > 0 \).
Proof of Proposition 2: From equation (3),

\[
P(D_t | y_t \leq \alpha) = \int_{-\infty}^{y_t} \gamma_1 \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, dU_t = \gamma_1
\]  

(11)

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

\[
= \gamma_1 + \gamma_2 \int_{\alpha}^{y_t} U_t \frac{\phi(U_t; \mu_t, \sigma_t^2)}{\Phi(y_t; \mu_t, \sigma_t^2)} \, dU_t
\]

Following the same steps that we used to derive the expression for \(e^L(y_t)\) in (6), we can write

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

(12)

From equation (11), \(\frac{dP(D_t | y_t \leq \alpha)}{dy_t} = 0\) and from equation (12), \(\frac{dP(D_t | y_t > \alpha)}{dy_t} > 0\) because \(\Lambda'(\cdot) < 0\) and \(\Phi \left( \frac{y_t - \mu_t}{\sigma_t}; 0, 1 \right)\) is increasing in \(y_t\). This also establishes that there is a slope discontinuity at \(y_t = \alpha\). Further, substituting \(y_t = \alpha\) in equation (12) eliminates the term inside square brackets, implying that there is no level discontinuity at \(y_t = \alpha\).