MANCHESTER 1824

The University of Manchester

Economics Discussion Paper Series EDP-1411

Nutrition, Activity Intensity and Wage Linkages: Evidence from India

Katsushi S. Imai, Samuel Kobina Annim,
Veena S. Kulkarni, Raghav Gaiha

May 2014

Economics
School of Social Sciences
The University of Manchester
Manchester M13 9PL

Nutrition, Activity Intensity and Wage Linkages: Evidence from India

Katsushi S. Imai *
Economics, School of Social Sciences, University of Manchester, Arthur Lewis Building, Oxford Road, Manchester M13 9PL, UK; Email: Katsushi.Imai@manchester.ac.uk

Samuel Kobina Annim
Department of Economics, University of Cape Coast, Cape Coast, Ghana; Email: skannim@gmail.com

Veena S. Kulkarni
Department of Criminology, Sociology & Geography, Arkansas State University, P.O. Box 2410, AR 72467, USA; Email: vkulkarni@astate.edu

Raghav Gaiha
Faculty of Management Studies, University of Delhi, Kamla Nagar
New Delhi, DL India; Email: gaiha@MIT.EDU

This version 28th May 2014

Abstract

The present study tests the twin hypotheses, namely, (a) the poverty nutrition trap hypothesis that wages affect nutritional status, and (b) the activity hypothesis that activity intensity affects adult nutrition as measured by the Body Mass Index (BMI) in the context of India. The analyses draw upon three rounds of National Family Health Survey (NFHS) data in 1992, 1998 and 2005 and National Council of Applied Economic Research (NCAER) data in 2005. Our results indicate strong support for both the hypotheses in India. Physically intensive activity tends to worsen the nutritional conditions and there is evidence for a poverty nutrition trap associated with labor market participation.

JEL Codes: C21, C23, I14
Key Words: Adult Nutrition, Malnutrition, Poverty Trap, Activity Intensity, Quantile Regressions, Pseudo Panel, India

*Corresponding Author:
Katsushi S. Imai (Dr), Economics, School of Social Sciences, University of Manchester, Arthur Lewis Building, Oxford Road, Manchester M13 9PL, UK; Phone: +44-(0)161-275-4827; Fax: +44-(0)161-275-4928; E-mail: Katsushi.Imai@manchester.ac.uk

Acknowledgements
This study is funded by ESRC research grant entitled 'On the Change of Poverty and Undernutrition in Rural India' (RES-000-22-4028). We appreciate useful comments from the participants in Workshop on Poverty and Inequality in China and India, the University of Manchester in March 2012, the 15th Kansai Development Microeconomics (KDME) Workshop at Osaka, Japan, July 2012, a research seminar at Dosisha University in August 2012, and the conference in Delhi organised by National Council of Applied Economic Research and University of Maryland in June 2013. We thank Md Azam, David Bloom. Subhasish Dey, Per Eklund, Raghbendra Jha, and Takahiro Sato for their valuable advice at various stages of the project. The views expressed here are those of the author and the usual disclaimer applies.

Nutrition, Activity Intensity and Wage Linkages: Evidence from India

1. Introduction

The finding of the study by Deaton and Dreze (2009) that persistent decline in calories and other nutrients in India cannot be attributed to falling per capita incomes and rising food prices, has rekindled an interest to re-examine the long standing discourse on the relationship between wages, efficiency and nutrition. Since the 1950s, the efficiency-wage hypothesis, which was put forward by Leibenstein (1957) and later theorized by a number of researchers (Bliss and Stern, 1978; Dasgupta and Ray, 1986, 1987; Dasgupta, 1993; Swamy, 1997), has been examined in varying contexts (Strauss, 1986; Behrman and Deolalikar, 1989; Weinberger, 2003, Jha et al. 2009). Also, the efficiency wage hypothesis has been expanded to explore the effect of activity intensity and nature on nutrition (Church et al., 2011). Further, an analogous and testable proposition of the efficiency-wage hypothesis is the poverty-nutrition trap (PNT) hypothesis. The PNT hypothesis proposes that relatively high wages would enable the purchase of food which then strengthens adults to continue with their work. Such linkages would create a "virtuous circle" - a cyclical relationship between wages (income), ability to acquire food, nutrition and work efficiency. Conversely, low wages or unemployment would create a "vicious circle" where low wages or unemployment would lead to low purchasing power and nutrition that would make poor workers trapped into poverty through low efficiency. It is therefore not surprising that adult malnutrition has attracted enormous attention in both academic and policy arenas. Following the pioneering contribution of Dasgupta and Ray (1986), the PNT hypothesis was tested by Jha et al. (2009) in the context of rural India.

Premised on the pioneering work on the efficiency-wage hypothesis, the extant literature has largely modeled the relationship around the effect of better nutritional status on wages. However, given that the relationship between wages and nutritional status is complex and mediated through factors, such as sector of employment, gender, interpersonal skills, social interactions, among several others (Jha et al., 2009), empirical research on the subject has failed to reach a consensus. The present study contributes to this literature by examining the role of intensity or nature of activities in addition to wages in predicting nutritional status of adults. The central objective of this study is to estimate the effect of both wages and activity intensity or nature on adult nutrition, as measured by the Body Mass Index (BMI)[1].

The study is organized as follows. Section 2 reviews the background relating to the two central hypotheses on the wage - nutrition link and the activity intensity - nutrition link. Data used for the econometric analysis are explained in Section 3. Section 4 and Section 5 discuss econometric specifications and the results respectively. The final section lays down a few concluding observations.

2. Background

Despite recent high levels of GDP growth and steady poverty, nutritional status has declined, or has improved only slowly, across households with varying income levels in rural areas in India. This has been reported as “the (Indian) empirical puzzle” and discussed by Deaton and Dreze (2009), Palmer-Jones and Sen (2001) and Patnaik (2004, 2007).[2] Drawing upon

[1] It is calculated by dividing weight in kilograms by ‘height in meters’ squared (kg/m^2).

[2] Using the NSS data in 1993, 2004 and 2009, Thorat and Dubey (2012) showed that poverty in terms of household expenditure declined recently - particularly during the period between 2004 and 2009 - across different socio-religious household groups in India. However, the evidence to support the corresponding improvement in nutritional status is still limited. Providing a clue for “the empirical puzzle” is thus important in this compelling setting.

National Sample Survey (NSS) data, Deaton and Dreze (2009) argue that "the calorie Engel curve" that plots per capita total calories or cereal calories and household per capita expenditure, has shifted consistently downwards between 1983 and 2005.[3] Following the efficiency wage hypothesis under which nutritional intakes would affect labor productivity and thus wage rates, the above observation suggests that those undernourished remain trapped in nutrition poverty due to low wage rates or exclusion from the labor market due to low nutritional intakes[4]. Jha et al. (2009) tested the existence of PNT using National Council of Applied Economics Research (NCAER) data in 1994 and showed that the intake of calories or micronutrients (e.g. Iron, Riboflavin, Thiamine) affects agricultural wages for various activities, such as harvesting or sowing and vice versa. The finding by Jha et al. (2009) lends support to the PNT hypothesis. Examining the NCAER data this study found that the elasticities for calories, protein, and five micronutrients (calcium, thiamine, riboflavin, carotene and iron) are positive and significant and therefore an increase in income would increase nutrient intake.

Further, the expectation that engaging in energy sapping activities leads to undernutrition has triggered a lively debate but with little empirical investigation. Theoretically, this is related to the model of health production function (Thomas, 1994) where health or nutritional outcomes as an output is a function of a number of inputs (e.g. nutrient intakes and the quantity and quality of health care, and individual and household characteristics) with a standard utility function of household members under a budget constraint for the household. In this framework, it would be

[3] Deaton and Dreze (2009) offered conjectures to explain the puzzle, such as, improvements in the disease environment and a reduction in work activity intensity. On the other hand, using the NSS data in 1993 and 2004, Gaiha et al. (2013, 2014) have developed an alternative explanation of changes in the consumption of calories, protein and fats and found significant effects of food prices on these nutrients. Also, there has been a considerable increase in eating out in India, which is not captured by the data. However, due to the data limitation, the present study focuses on the PNT and activity hypothesis.

[4] See Jha et al. (2009) for a more detailed review of the PNT hypothesis.

natural to assume that higher activity intensity - which would normally require more calories - results in lower levels of nutritional outcome.

Given the above context, the research questions that the present study proposes to address include; (i) What are the determinants of nutritional status, as measured by BMI, of men and women in rural India?; and (ii) Have those determinants changed over the years?; (iii) Did the poverty trap hypothesis or the "wage-nutrition" hypothesis hold, i.e., were there any individuals who were trapped into nutrition poverty due to their low wages in the labor marker?[5]; and (iv) Did the "activity-nutrition" hypothesis hold, i.e., were there any individuals who were undernourished due to higher intensity of work or activity given that energy expenditure tends to exceed the required energy intake for those people? We will consider the effects of (predicted) wage rates on nutrition to answer the third question. To answer the last question, we will estimate undernutrition measured by BMI with intensity and with type of activity as a proxy for intensity of work. We employ the data provided by India's National Family Health Survey (NFHS henceforth) and National Council of Applied Economic Research (NCAER hereafter).

3. Data

The activity-nutrition and wage-nutrition hypotheses are examined using NFHS and NCAER datasets. The NFHS was initiated in 1992-93 and, since then, two more waves of the data have been collected in 1998-99 and 2005-06. ICF International, via ORC Macro MEASURE DHS (Monitoring and Evaluation to Assess and Use Results of Demographic and Health Surveys), National AIDS Control Organization (NACO) and National AIDS Research

[5] Despite its evident importance, relatively little is known about links between BMI and participation in workfare programs, particularly in India. Using a unique data set for the Indian state of Rajasthan for 2009-10, Jha et al. (2013) attempt to fill this void and examine the association between BMI and participation in NREGS, the latter of which covers not only participation status but also duration and earning. They allow for the mutual endogenity, say, between BMI and earnings.

Institute of the Ministry of Health and Family Welfare (MOHFW) are responsible for the NFHS (International Institute for Population Sciences (IIPS) and Macro International, 2007). The survey is nationally representative and covers fertility, family planning, maternal and child health, gender, HIV/AIDS, nutrition and malaria. Data are collected at the individual level (children, mothers and, lately, fathers), household and at the community level. This study uses data on ever married women[6], aged between 15 to 49 years and resident in rural areas for each of the three rounds of the NFHS survey.

The NCAER data set in 2005, known as the IHDS (Indian Human Development Survey) data, has been employed to explore the effects of both wages and activity intensity on adult BMI. The IHDS data collected through collaborative research between NCAER and the University of Maryland is a multi-topic survey. The 2005 data set is nationally representative covering all the states of India and the thematic areas include education, health, livelihoods, family processes and the social structure within which the households operate. To this end, the IHDS is structured into individual, household and village level datasets.

BMI, our measure of nutrition, is used to examine both thinness and obesity. In this study, both the raw scores and BMI classifications based on the following cut-offs; severely underweight ($BMI<16kg/m^2$); underweight ($16kg/m^2<=BMI<18.5kg/m^2$); normal ($18.5kg/m^2<=BMI<25kg/m^2$); overweight ($25kg/m^2<=BMI<30kg/m^2$) and obese ($BMI>30kg/m^2$) are used.[7]

[6] In addition to the data of women, some data of men were collected in the third round.

[7] (a) Recent literature has questioned the appropriateness of BMI as a proxy for adult nutrition. For instance, WHO expert consultation (2004) addressed the need for applying the population-specific cut-off points for BMI, e.g., as Asian populations differ from European populations in terms of the associations among BMI, percentage of body fat, and health risks. However, WHO agreed that the WHO BMI cut-off points should be retained as international classifications (ibid., 2004). Burkhauser and Cawley (2008) argue that BMI is flawed as it does not distinguish fat from fat-free mass such as muscle and bone. However, as the available

The first objective of investigating the relationship between wages and nutrition is based on log of hourly wages as one of the explanatory variables. The second objective of examining the relationship between activity intensity and nutrition is implemented by using the variables on nature of profession and the classification based on type of employer as an indicator for activity intensity. The former, nature of profession, is premised on the fact that certain types of professions are more manual or sedentary than others. The second case where we use employer type, such as not working, working for a family member, working for someone else, or self-employed, provides another perspective on examining the effect of occupational characteristic on nutrition. Using the NCAER data, nature of profession is grouped into four categories based on the degree of physical activity required.[8]

We will provide below a brief summary of the patterns of malnutrition which our data sets have revealed.[9] We classified the entire sample into three groups, namely "underweight" (BMI<18.5kg/m^2), "normal" (18.5kg/m^2<=BMI<25.0kg/m^2) and "overweight/obese" (BMI>=25.0kg/m^2) for both NFHS and NCAER. We have observed that the proportions of underweight, normal and overweight/obese are 35.3 percent, 52.3 percent and 12.5 percent, respectively, and this pattern is comparable to 35.6 percent, 51.8 percent and 12.6 percent as shown in the NFHS report (International Institute for Population Sciences (IIPS) and Macro International, 2007). The NFHS reports show that the proportion of underweight women declined marginally, from 35.8 percent to 35.6 percent over the period 1999 to 2006. However,

---

data on adult nutrition are limited, we will use BMI as a proxy realising these limitations. (b) Hip circumference and height are more correlated with percentage body fat than anything else, including waist circumference and weight. Taking both into account, an alternative measure, the Body Adiposity Index, was computed. BAI is a good predictor of percentage adiposity, so if the BAI is 30, then the percentage body fat is around 30 per cent. It is reasonably accurate but not terribly accurate (George, 2011). Hence the use of BMI remains widespread.

[8] Ideally, finer categories or more direct measures should be used to capture the activity intensity, but such data are not easily available. We have identified the best measures given the data constraints.

[9] A full set of results will be furnished on request.

concentrating on the rural sample for women, we observed a drop from 43.6 percent to 40.2 percent over the same period.

We compute the proportions of these three groups according to individual and household characteristics. First, education is closely associated with the nutritional status. While a majority of underweight women had no education, two-fifths of the overweight/obese women in rural areas have secondary education. Second, in the context of the activity-nutrition hypothesis, we observe that classification of employer type (working for a family member, working for someone else, and self-employed) by nutrition shows some differences within each of the three nutrition categories. For example, the proportion of respondents "not working" is higher for the overweight/obese group than for the underweight. In contrast, the proportion of those "working for someone else" is lower for the obese group than the underweight and normal groups. Disaggregation by profession (professional/managers, clerical/sales/services, famers/fishermen and laborers/production workers) shows that there were relatively more "farmers/fishermen" found in the underweight group than the obese group while the share of "professional/managers and clerical/sales/services" was higher in the obese group than in the underweight group. These observations call for further examinations of the activity intensity-nutrition relationship.

4. Econometric Specifications

The main econometric techniques that we employ to examine the twin hypotheses of wage-nutrition and activity-nutrition are a) quantile, b) pseudo panel, c) instrumental variable (IV) and Heckman sample selection regression models. In each of the different specifications, the least squares technique is used to estimate the structural model. The following motivates the choice of different estimation techniques.

First, previous research indicates that it is important to take into consideration different effects of the correlates of malnutrition across different classifications of BMI. For instance, Strauss and Thomas (1998) find that the Vietnamese war had varied height effects on different sub-groups in the sample with different effects across short, average and tall people across the south and north. Similarly, the effect of activity intensity on malnutrition is expected to be more evident among the obese than the severely underweight. That is, the obese are more likely to respond to weight changes than the severely underweight in consuming extra calories. For this reason, quantile regression is appropriate for examining the effect of the correlates at different points of conditional distribution of a dependent variable. Second, to take account of unobserved effects for age and state cohort, we construct a pseudo panel that helps in addressing this potential bias. These econometric models are used in the case of NFHS data only as the NCAER data are limited in the coverage of the anthropometric data for adults.

Third, in an attempt to investigate causality between wages and malnutrition, on one hand, and between activity intensity and malnutrition, on the other, the IV model and Heckman sample selection regression model are applied in the case of the 2005 NCAER data. We use the IV model to investigate the wage-nutrition relationship given that there may exist a two-way causal relationship between the nutritional status and the wage rate in our hypothesis testing. The nutritional status is premised on "how much better (or healthier) we eat given a little more income" and the wage rate is based on "how healthier (or more productive) we become by eating a little bit more" (Strauss, 1986; Behrman and Deolalikar, 1989).[10] To address this from an

[10] Swamy (1997) focused more on the effect of nutrition on worker's productivity in the context of rural India. According to him, the nutrition-based efficiency wage theory predicts that wages should be rigid because lowering them would reduce worker's productivity and would increase the cost per efficiency unit, but he found that a wage cut lowers the cost per efficiency unit of labor. While this aspect is important, we focus more on the effect of wages (which are determined by proxies for activity intensities) on nutrition to make the estimation tractable under the data constraint.

econometric perspective, we use the availability of trade-union in a community and the state-level consumption inequality as instruments for wages in the BMI equation. The rationale for the former is that the presence of a trade union enhances the bargaining power of workers to negotiate for better wages for vulnerable people (such as the malnourished). The availability of a trade union is therefore more likely to be correlated with wages than BMI.[11] The validity of the instrument is supported by the specification tests. Also, the consumption inequality as an instrument for wages is based on the traditional income-consumption relationship. Although a study for the United States (Krueger and Perri, 2006) suggests that the increase in income inequality has not been accompanied by substantial increase in consumption inequality, the association between consumption inequality and income (wages) is far from fully refuted.

Finally, to carefully explore the effect of wages and activity intensity on nutrition, we compare the OLS and IV estimates for the case in which we eliminate the effect of sample selection in the wage equation. In this case, we use predicted wages from the corrected wage equation in the nutrition equation. Although this may be subject to the criticism that the wage values are only the estimates, and not real values, it offers an opportunity to identify the effect of wages on nutrition albeit with possible errors (sample selection bias). In the Heckman sample-selection model, the probit model is estimated to exclude the respondents who were unemployed in the first step, and then the wage equation is estimated in the second step which corrects the sample selection bias through the inverse Mills ratio derived by the probit model in the first step. This implies that the estimated wage effect will not capture the characteristics of

[11] While having a trade union in the community is not very usual in rural India, the rural trade union, such as, AIAWU (All India Agricultural Workers Union) has played an important role in some areas (e.g. Haryana) in recent years (Byres et al., 2013). One could suspect that the presence of a trade union may be correlated with the community unobservables which may affect BMI, but the statistical correlation between the availability of trade union and BMI is weak and the specification tests validate our instruments (Table 4).

the unemployed. In this study, we use the infant dependency ratio as an exclusion restriction variable for employment. The choice of infant dependency ratio is based on the hypothesized positive relationship between higher economic dependency and labor market participation. That is, a household with higher dependency burden has greater impetus to search for a job and participate in the labor market than households with lower dependency burden. However, this hypothesized positive relationship in the case of female labor market participation is likely to be rejected, at least in the short-run, due to health constraints around delivery period. This is a topic of long standing academic discourse (see, for example, Bilsborrow, 1977).

Model 1 – Ordinary Least Squares Regression

Equation 1 below represents the ordinary least squares (OLS) estimation of the functional relationship between BMI and its correlates.

$$BMI_i = \alpha_0 + \alpha_1 IND_i + \alpha_2 HH_h + \alpha_3 ST_\rho + \mu_i \qquad (1)$$

where *i* stands for individual (or the $i^{\text{-th}}$ individual), *h* for household, and ρ for state.[12] *BMI* is the adult's body mass index. *IND* is a vector of individual factors, specifically, the age and its square (to account for non-linearity between age and BMI), education, measured as a categorical variable ("no education", "attempted or completed primary", "attempted or completed secondary" and "attempted or completed any higher than secondary education"), working status, measured as a categorical variable ("not working", "working for a family member", "working for someone else" and "self-employed") and marital status ("currently married", "formerly married"

[12]Ideally, variance should be clustered at household level, but we do not take account of the clustering effects as the commands for the 'robust' estimator or quantile regression do not allow us to incorporate the clustering effects. However, the regression results of OLS with clustering effects and those without provide us with very similar results, suggesting that the clustering effects would not significantly affect the results in our case.

and “never married”). *HH* is a vector of household level characteristics, that is, wealth, measured as a continuous variable[13], religious affiliation of the household head measured as a categorical variable (Hindu, Muslim, Christian, Sikh, Buddhist/Neo-Buddhist, Jain and Other), social group of the household head measured as a categorical variable (scheduled caste, scheduled tribe, other backward group and non-backward group), household size, measured as a non-categorical variable, distance to water measured as a continuous variables, agricultural land size owned by the household and its square). *ST* is a vector of state level indicators- specifically, characteristics of different locations classified into BIMARU (Rajasthan, Uttar Pradesh, Bihar and Madhya Pradesh), North (Jammu & Kashmir, Himachal Pradesh, Punjab, Uttaranchal and Haryana), South (Maharashtra, Gujarat, Goa, Andhra Pradesh, Karnataka, Kerala and Tamil Nadu), East (Sikkim, Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, Meghalaya, Assam, West Bengal, Jharkhand, Orissa and Chhattishgarh) and Delhi and state level prices of commodities (sugar, eggs and cereals)[14]. μ is the error term which is independently and identically distributed (*i.i.d).*

Model 2 – Quantile Regression

As stated above, we augment the OLS regression with a quantile regression. Koenker and Basset (1978) prove that for any distribution the median is a better measure of location. The regression median[15] is more efficient compared to OLS which is underpinned by the Gaussian assumptions. In contrast to OLS, quantile regression sorts the data and identifies a threshold (τ) to estimate the

[13]The wealth score in the NFHS data is computed based on the asset index, a composite index weighting a number of different proxies for household assets, such as type of flooring, number of bedrooms, radio, TV, radio or telephone. See Rutstein and Johnson (2004) for details of the wealth score.

[14] For a similar specification but a different method of estimation, see Jha et al. (2013).

[15] The proof of the median regression can be easily replicated for other percentiles (quantiles).

coefficient (β) that minimizes the sum of absolute residuals. The general set-up of quantile regression, Equation (2), is solved from an optimization perspective using linear programming:

$$\hat{\beta}_{(t)} = \underset{\beta_\tau \in \Re^K}{\arg\min} \sum_{i=1}^{n} \rho_\tau \left(y_i - x'_i \beta \right) \qquad (2)$$

where estimated β(τ) called 'tauth' (τ^{th}) regression quantile estimates the coefficient at a specified threshold (τ). τ is the sample quantile and takes on any value between 0 and 1. The expression $\rho_\tau \left(y_i - x'_i \beta \right)$, the absolute value function, weights the absolute difference between y_i and $x'_i \beta$ with τ and by (1 – τ) for all observations below the estimated hyperplane. Koenker and Basset (1978) estimate conditional quantiles using the minimization procedure synonymous with least squares.

In contrast to the previous research (Bassole, 2007; Aturupane et al., 2008) that use traditional thresholds (for example, 10^{th}, 25^{th} and 50^{th}), the present study identifies respective thresholds that characterize the following group of respondents; severely underweight (BMI< 16.0), underweight (16.0<=BMI<18.5), normal (18.5<=BMI<25.0), overweight (25.0<=BMI<30.0) and obese (BMI >= 30.0). The rationale is to help place the observations in a context for policy targeting.

Model 3 – Pseudo Panel Regression

The econometric analysis further makes a case to incorporate unobserved unit-specific characteristics. The unobserved unit-specific (individual, household and community) also affects adult nutritional status. The use of pseudo panel is useful in our study of India where real panel data on nutritional issues at the national level is rare. In this study, we generate a pseudo panel based on mother's age cohorts for each state over the three waves of NFHS survey. Deaton

(1986) puts forward a pseudo panel when more than one cross-sectional data has a common variable, for example, age, education and location. The use of such variables is premised on the assumption that the classifications rarely change overtime and they are exogenously determined outside the model.

We specify the functional form relationship (Equation (3)), followed by multiple regression of the pseudo panel (Equation (3a)). The specification of Equation (3a) below is the fixed effects (FE) model applied to pseudo panel data. We do not specify the random effects (RE) model specification of Equation (3). The rationale for choosing FE is a counter test on the assumption that the unit-specific effects are constant over time and as such arbitrary correlation (clumsy construction) is assumed between the unobserved heterogeneity term and the explanatory variables (see Wooldridge, 2009).

The functional form of the pseudo panel is specified as:

$$BMI_{i\,gt} = f\left(IND_{i\,gt},\, HH_{i\,gt}\right) \qquad (3)$$

where subscript *g* is the cohort captured by age cohorts (classified into seven categories) for 18 states[16] and *t* stands for year. This yields a sample size of 378 (7 x 18 x 3). All the notations remain the same as in Equation (1).

The estimable form of Equation (3) which takes the form of a real panel is specified as follows:

$$\overline{BMI}_{gt} = \sum_{r=1}^{q} \delta \overline{X^{r}}_{gt_r} + \varphi D_t + \bar{a}_g + \bar{\mu}_{gt} \qquad (3a)$$

[16] Over time states in India have been reclassified and that results in difference in the states across the three waves of the NFHS. The states used for the pseudo panel are; Arunachal Pradesh, Assam, Bihar, Gujarat, Haryana, Jammu and Kashmir, Karnataka, Kerala, Maharashtra, Manipur, Meghalaya, Mizoram, Nagaland, Orissa, Punjab, Rajasthan, Tripura and Utter Pradesh.

where the first term on the right hand side is a simple aggregation of the set of all the explanatory variables as specified in Equation (1) above.[17] We denote the subscript of this composite term by *r* to represent the individual and household explanatory variables. Subscripts *g* indicates the cohort constructed by mother's age and state and *t* stands for each round of the NFHS survey. The term *D* captures the time effect and the last two variables $\bar{a}_g$ and $\bar{\mu}_{gt}$ are made up of the (time-invariant) unobserved heterogeneity term (or fixed effects term) and the idiosyncratic error term. The daunting issue is to examine the extent to which the pseudo panel approximates a real panel. With the absence of real panel data it is virtually impossible to address this issue, hence we rely on the argument by Verbeek and Nijman (1992) and Verbeek (1996) that the estimator is consistent if the number of observations in cohort *g* tends to infinity, $\bar{\mu}_{gt} \rightarrow \mu_g^*$. Since the number of households within each cohort is large in our case, we are confident that the estimator is likely to be almost consistent.

Model 4 – Instrumental Variable Regression

The theoretical underpinning of the PNT suggests that wages and nutrition are endogenous in the respective equations of the other (see Jha et al., 2009). This implies that while in the estimation of BMI wages are endogenous, BMI is endogenous in the prediction of wages. Ideally, this will require the use of an instrument each for BMI and wages to estimate systems equation. In this study, however, we attempt to resolve the endogeneity inherent in the BMI equation. As discussed previously, the instruments used for wages are: availability of trade union in a village and district level consumption inequality. Equation (4) below shows the econometric specifications for the reduced form equations.

[17] Subscripts, *i*, *h* and ρ in Equation (1) are omitted in Equation (3a).

$$Lnwages_i = \alpha_0 + \alpha_1 IND_i + \alpha_2 HH_h + \alpha_3 ST_h + \alpha_4 TU + \alpha_5 Consumption\_Gini + \mu_i \quad (4)$$

where $Lnwages_i$ is the log of wage at the individual level and *TU* and *Consumption Gini* are the two instruments used in correcting for endogeneity of wages in the BMI equation. The vector notations remain the same as specified in Equation (1). However, when we use the NCAER data, it is feasible to examine occupation-based physical activity disaggregated into 'Professionals/managers', 'Clerical/Sales/Services', 'Famers/Fishermen' and 'Laborer/ Production workers'.

Following the standard approach for IV, Equation (4) is estimated in the first stage and the predicted values are jointly plugged into the structural equation, Equation (1). In estimating IV, efficiency is compromised due to large standard errors as a result of the two- stage estimations. Hence IV is preferred to OLS if only the former yields consistent results. This largely depends on the validity of choice of instruments. In this study, the validity (strength and relevance) and identification of the instruments are examined by using the Kleibergen-Paap rk LM statistic (underidentification), Cragg-Donald Wald F statistic (weak identification) and Hansen J statistic (overidentification). The first stage results and the Hausman test are used for choosing between OLS and IV estimations.

Model 5 – Heckman Sample Selection Regression

The Heckman sample selection model is premised on the argument that the estimation of wage equation should not be only based on truncated data since the unemployed are purposively (or non-randomly) excluded from the model. The model based on the non-randomly selected sample yields biases estimators. To address this, we estimate the probit model for both employed and unemployed respondents (employment probability equation) that are randomly selected in the first stage. The main requirement for the Heckman sample selection is the exclusion restriction condition. The inclusion of the exclusion restriction entails that the employment probability equation (also termed as the selection equation) should incorporate a variable that explains the probability of getting employed but not wages. In the present analysis, we opt for the infant dependency ratio as the exclusion restriction. Like the IV, the two stage Heckman model compromises on efficiency of the coefficients, hence the model should only be chosen over OLS if it yields consistent estimates, which would require a large sample. Otherwise the OLS based on the sub-sample should be chosen over the Heckman model.

Equations (5) and (5a) below present the selection and outcome equations for the Heckman sample selection estimation. Equation (5) is a binary outcome model (probit) and Equation (5a) is an OLS regression. The first stage is the probit model estimation, based on which the inverse Mills ratio (ratio of the probability density function to the cumulative distribution function of a distribution) is estimated and plugged into the OLS equation.

1st stage: selection equation (probit)

$$Emp^{*}_{i} = \alpha_0 + \alpha_1 PID_i + \alpha_2 PID'_h + \alpha_3 INF_h + \mu_i \qquad (5)$$

2nd stage: wage equation (OLS)

$$Lnwages_i = \alpha_0 + \alpha_1 PID_i + \alpha_2 PID'_h + \varepsilon_i \qquad (5a)$$

$$E\left(Lnwages_i \middle| Lnwages_i \text{ is observed}\right) = E\left(Lnwages_i \middle| Emp^{*}_{i} > 0\right) \qquad (5b)$$

where Emp^{*}_{i} is a binary response (= 1 if respondent is employed and 0 otherwise). PID_i stands for variables at the individual level, that is, age and its square, sex and education (measured as categorical variables). The variables at the household level (denoted by PID'_h) include household size and its square, social group of household head, presence of trade union in a village, district level consumption inequality and location dummies. INF_h represents the 'exclusion restriction' variable, that is, infant dependency ratio. In Equation (5a), $Lnwages_i$ represents the log of hourly wages and all other variables have the same notations as in Equation (5). ε_i and μ_i in both Equations (5) and (5a) are error terms with the following properties: (μ, ε) ~ N(0, 0, σ^2_u, σ^2_ε, $\rho_{\varepsilon u}$). The first two terms in the description of the distribution of the error terms are the zero mean condition for the two equations, the third and fourth terms, σ^2_u, σ^2_ε, represent constant variance for the respective equations and the last term, $\rho_{\varepsilon u}$, is the correlation between the error terms of the two equations. Estimations of Equations (5) and (5a) give the expected value of log of wage conditional on the probability that log of wage is observed as in Equation (5b).

5. Results

The econometric results are given in Tables 1-4. We report (i) the results for rural women for NFHS data 1, 2, and 3 in Table 1; (ii) those for rural men for NFHS 3 in Table 2[18]; (iii) those for selected variables for pooled OLS as well as pseudo panel model for rural women for NFHS 1, 2, and 3 in Table 3; and (iv) those for all the variables for NCAER data. Column 1 of Tables 1–2 gives the OLS results and columns 2–5 give the quantile regression results. A brief summary of the results is given below.

(i) The Results based on NFHS data (Tables 1-3)

Among various explanatory variables on household, individual, and state characteristics, we highlight age, household wealth, education, working status - our proxies for physical intensity,

[18] BMI of men are unavailable in the first and the second surveys.

and food price to save the space. For simplicity, we denote 1992-3 (NFHS-1) as 1992, 1998-9 (NFHS-2) as 1998 and 2005-6 (NFHS-3) as 2005.[19]

[Place Table 1 here]

Age

In Tables 1 and 2, fairly consistent and expected results are observed for BMI among rural women across 1992, 1998 and 2005 (Table 1) and for rural men in 2005 (Table 2). For instance, in the case of OLS, the coefficient estimate of age is positive and significant for women in 1998 and 2005 and for men in 2005, while it is not significant for women in 1992. The presence of a non-linear quadratic term (square of age) shows that the marginal effect of age on BMI gets smaller, that is, the observed positive effect weakened at older ages for women in 1998 and 2005 and for men in 2005. This evidence supports a finding by Jha et al. (2013). However, inspecting the quantile regression results for rural women, the age effect is not necessarily significant for all the groups. For women positive and significant effects of age tend to be observed for the (relatively) overweight, not for the underweight in 1998 and 2005, whilst for men age is positive and significant for all the groups with a greater coefficient estimate as they shift from 'underweight' to 'obese' in 2005. Comparing the results in 1992 with those in 1998, we conjecture that older and relatively overweight women became more overweight or obese in the 1990s possibly due to their better nutritional intakes. In Table 3 where pooled model or pseudo-panel model (or fixed effects model - which is chosen over random effects model guided

[19] The findings based on NFHS data but not presented in the tables include: (i) respondents with heads belonging to scheduled castes have lower BMI compared to their counterparts in 1998; (ii) in some instances, household heads who are Hindus, Muslims, Sikhs and Jains have lower BMI than the rest (e.g. Christian); (iii) use of a flush toilet is associated with higher BMI; and (iv) size of household is negatively associated with BMI of adults in the case of OLS for 1998 and 2005.

by Hausman test) is applied for three rounds, age is positive and significant with a negative coefficient for age squared, which is by and large consistent with the results of Tables 1 and 2.

[Place Table 2 here]

Wealth

In all the cases of Tables 1 - 3, the coefficient estimate of wealth score is positive and highly significant, suggesting that BMI of both men and women increases with household wealth. This is presumably because wealthier households are able to afford more food which will improve their BMI. In Tables 1 - 2, the results of quantile regression indicate that the coefficient estimate for wealth scores is higher for more overweight or obese groups than underweight groups, or, the marginal effect of wealth on BMI (or our proxy for nutritional poverty) is non-linear. Wealth score is positive and significant in Table 3 (pooled or pseudo panel model).

[Place Table 3 here]

Education

The role of education in BMI for women changed dramatically from 1992 to 1998 and 2005. The results of OLS in Table 1 show that education is not statistically significant in 1992, while all the categories, "primary school", "secondary school" and "higher than secondary" are positive and significant in 1998, and "primary" and "secondary" are positive and significant in 2005. On the results of quantile regression for 1992, we find a positive and significant role of secondary school education in increasing BMI for the severely undernourished with the marginal effect

0.25 (i.e. attempting or completing secondary school would improve BMI by 0.25 *ceteris paribus*), but in the meantime secondary education would significantly reduce BMI for overweight or obese women. The coefficient estimates of education are statistically insignificant in many cases of quantile regression for 1992.

However, in case of quantile regression for 1998, education becomes more important in improving women's BMI, but a positive effect was observed for the relatively well-nourished women and not so for the undernourished. Neither primary school nor secondary school education turned out to be significant for "severely underweight" or "underweight", while higher education significantly increases BMI of underweight women. For 2005, however, education appears to have a more important role in improving BMI for the relatively underweight women. The coefficient estimate of "primary school" is positive and significant for both "severely underweight" and "underweight" women and "secondary school" is positive and significant for "underweight" women. We observe in Table 2 that the pattern of the results of education in quantile regression, in year 2005, for rural men is similar to that for rural women. In Table 3, all categories of education are positive and significant in Column (1) wherein all the samples of rural women across three rounds are pooled. In pseudo model (fixed effects model, Column (2)), the category, "no education", tends to reduce BMI of rural women significantly, as expected.[20]

Working Status

As discussed in the previous section, we test whether more physically intensive work reduces BMI. In 1992, "working for someone else" positively and significantly affects women's BMI (OLS, the first panel of Table 1). In case of the quantile regression, the relationship between the

[20] Categorical variables of education are not used for pseudo panel as they are correlated once aggregated at the cohort level.

category, "working for someone else", and BMI is observed to be positive and significant for "severely underweight" women. Further, as per the quantile regression estimates, while the coefficient of "working for family member" is positive and significant only for "severely underweight" women, it is negative and statistically insignificant for "overweight" or "obese" women. This implies that those severely underweight tended to be poor and could easily find jobs and so finding the job itself was likely to improve their nutritional status even if they work for family member or somebody else.

From the second and third panels of Table 1, we observe that working for someone else or a family member is associated with lower BMI in 1998 and 2005. In both years, this relationship is observed in the case of the OLS and for underweight, normal, overweight and obese in quantile regression. The effect of working for someone else in the case of severely underweight women weakens and turns positive and significant again in Column 2 ("severely underweight") for 2005.

Column (2) of Table 3 (fixed effects for the pseudo panel) shows that employer type, a proxy for occupation based physical activity affects women's BMI significantly. That is, BMI of women "working for someone else" or "being self-employed" is lower than that of women 'not working'. This is consistent with the previous findings based on the cross-sectional estimations. Cohorts with a higher proportion of wealthy people are associated with higher BMI. From an education perspective, greater proportion of non-educated individuals in a cohort is related to lower BMI scores. In a nutshell, both occupation-based physical activity intensity and wealth (proxy for wages) are related to BMI of adults.

Food prices

On the effect of food prices on BMI, we observed that price of higher price of eggs is associated with lower BMI for women in 1998 and 2005 as well as for men in 2005, but such an effect was not observed in 1992 (Tables 1 and 2). Price of eggs is negative and significant in case where pseudo panel model is applied (Table 3). On the other hand, higher price of cereals is associated with higher BMI in most of the cases (Tables 1, 2 and 3)[21]. This may sound counter-intuitive, but it could be the case that higher prices would increase farmers' income, which would improve the nutritional status due to the increased purchasing power of the household.

(ii) **Results based on NCAER data (Table 4)**

In Table 4, the NCAER data are used to further explore the relationship between occupation-based physical activity intensity and nutrition, on one hand, and investigate the nutrition-wage relationship, on the other. As indicated earlier, three variants of econometric analysis, namely; OLS, IV and Heckman sample-selection models are used. To examine the effects of the two hypotheses separately the OLS equation is estimated twice. The first equation estimates a restricted model, that is, without occupation-based physical activity variables, while the second OLS equation estimates an unrestricted model. The results based on the OLS equations (Columns (1) and (2) of Table 4) show that wages in both equations are highly significant and positive in explaining the BMI of adults. On the other hand, the results in Column (2) of Table 4 indicate that adults engaged in occupations that are physically (manually) strenuous (farmers, fishermen, laborers and production workers) have lower BMI compared to adults engaged in more or less sedentary work (professionals, managers, clerical and sales personnel).

[21] It is positive and significant only for men with "normal" range of BMI in Table 2.

[Place Table 4 here]

The IV and OLS (based on Heckman wage regression) results are given in Columns (3) and (4) of Table 4. In both estimations the positive effect of wages on nutrition is confirmed. However, the effect of occupation-based physical activity intensity on nutrition is observed only in the OLS results which are based on predicted wages obtained from Heckman wage estimation model. The OLS results (Column (4) of Table 4) support the hypothesis that adults engaged in occupation-based *physically* intensive job have lower BMI. Indeed, the result is more revealing as we observe a 10 percent statistical significance on the dummy for clerical/sales/services personnel. That is, compared to professionals, clerical and sales personnel have lower BMI, though both are engaged in seemingly sedentary form of work. A possible reason is that when comparing managers to clerks, the latter are more likely to be active.

Interpretations of both the IV and the OLS (based on Heckman wage model) are supported by the first stage regressions shown in Appendix and the post estimation tests in the last six rows of Table 4. In Appendix, the first stage regression for the IV shows that both instruments (availability of trade union in a village and district level consumption inequality) are significant at the 1 percent level.[22] To verify validity of our instruments, the null hypotheses of

[22] The Appendix reports the econometric results of the labor market participation equation and the wage equation where we used the dummy variable on whether a worker is male (taking the value 1) or not (0) to capture the gender effect, which has turned out to be positive and statistically significant. Underlying this are the active debates on female labor force participation (FLP) in India. FLP varies considerably across different regions reflecting, e.g., the social and cultural norms, opportunities for education for women and child care facilities. For instance, a woman not in the labor force may have chosen to do so due to the socio-cultural reason and this may not necessarily imply low nutritional status. However, as we do not have enough data to disentangle these complex causal relationships, we have estimated the simpler version of wage and labor market participation equations where the gender dummy is used, while we realise the limitations of our approach.

underidentification and weak identification are rejected (Table 4). Also, from the last row of Table 4, we fail to reject the null hypothesis that the instruments are jointly valid. Finally, to make a choice between the OLS and IV, the Hausman test rejects the null hypothesis that both OLS and IV are consistent, but OLS is efficient. That is, rejecting the null hypothesis means that it is only the IV that yields consistent estimates.

In the case of the Heckman-wage regression, infant dependency ratio is significant in the employment probability equation at one percent level and the rho (correlation between the error terms in the participation and outcome equations) is significant. The latter suggests a rejection of the null hypothesis that there is no correlation between the employment probability and wage equations. This supports the use of the Heckman sample-selection in a wage equation.

6. Conclusion

The present study, drawing upon three rounds of NFHS data in 1992, 1998 and 2005 and NCAER data in 2005, tests two hypotheses; (a) the poverty nutrition trap hypothesis predicting wage affects nutritional status and (b) the activity hypothesis postulating that activity intensity affects adult nutrition in terms of Body Mass Index (BMI). We employ the following three econometric models. First, we apply quantile regressions to each round of cross-sectional data to take account of different behavioral response among different nutritional groups. Second, we construct a pseudo panel model to see any common pattern over the years. Finally, instrumental variable (IV) and Heckman sample-selection regression models are employed to test the poverty nutrition trap hypothesis, taking account of the sample selection bias associated with the labor market participation and the endogeneity of wages in the BMI equation.

Our results strongly support both hypotheses. That is, there exists a poverty nutrition trap associated with the labor market participation. After taking account of the sample selection bias associated with the labor market participation and the endogeneity of wages, we find that those who are left out from the labor market or experience lower wages tend to have lower levels of nutrition in terms of BMI. Further, our estimates show that those who are doing manual labor or more physically intensive and demanding activities (e.g. farmers, fishermen, laborers and production workers) are more likely to be undernourished than those who are doing less intensive activities (e.g. professionals, managers).

Both results would explain why the improvement of BMI has been relatively slow at all the ranges of nutritional groups. At the low end of the income distribution, people would need to enter into the labor market and earn wages to escape poverty. Additionally, even when they manage to find jobs, low wages and/or physically demanding jobs would prevent them from improving nutritional conditions. Only if they are able to earn higher wages, enter into the jobs which would require less physically demanding work and/or are self- employed possibly with higher education, they would be able to improve their nutritional status. However, this opportunity appears to be still relatively limited. In terms of policy implications, facilitating diversifications of activities of the poor, for example, through providing employment in non-farm or service sectors would be effective as a poverty alleviation strategy in reducing the prevalence of malnutrition.

Although we have not analysed explicitly the factors underlying the continued reduction of nutritional intakes, our results suggest that it is too optimistic to relate this reduction to the fact that more and more people are now doing physically less demanding work as a result of economic growth. It is more likely that a substantial number of rural people have found it difficult to escape from the poverty nutrition trap, or to shift to physically less demanding activities. Hence poverty alleviation programmes aimed at directly and/or indirectly addressing the problem of nutritional deprivations should continue to serve as an important role in rural India.

References

Aturupane, H., A. D. Deolalikar and D. Gunewardena, "The Determinants of Child Weight and Height in Sri Lanka: A Quantile Regression Approach," UNU WIDER Research Paper No. 2008/53 United Nations University, 2008.

Bassole, L., "Child malnutrition in Senegal: Does access to public infrastructure really matter? A quantile regression analysis," Paper presented at African Economic Conference 2007: Opportunities and Challenges of Development for Africa in the Global Arena, Addis Ababa, November 15-17, 2007.

Behrman, J. R. and A. D. Deolalikar, "Agriculture wages in India: The role of health, nutrition and seasonality," In *Seasonal variability in Third World agriculture*, David E. Sahn (ed.), Johns Hopkins University Press, Baltimore, 1989.

Bilsborrow, R. E., "Effects of Economic Dependency on Labour Force Participation Rates in Less Developed Countries," *Oxford Economic Papers*, 29(1), 61-83, 1977.

Bliss, C. and N. Stern, "Productivity, wages and nutrition: Part I: The theory," *Journal of Development Economics*, 5(4), 331-362, 1978.

Burkhauser, R. V. and J. Cawley, "Beyond BMI: The value of more accurate measures of fatness and obesity in social science research", *Journal of Health Economics*, 27, 519–29, 2008.

Byres, T. J., K. Kapadia and J. Lerche (eds.), *Rural Labour Relations in India*, Routledge, Abingdon and New York, 2013.

Chhabra P. and S. Chhabra, "Distribution and Determinants of Body Mass Index of Non-Smoking Adults in Delhi, India," *Journal of Health Population and Nutrition*, 25(3), 294-301, 2007.

Church, T. S., T. M. Thomas, C. Tudor-Locke, P. Y. Katzmarzyk, C. P. Earnest, R. Q. Rodarte, C. K. Martin, S. N. Blair, and C. Bouchard, "Trends over 5 Decades in U.S. Occupation-Related Physical Activity and Their Associations with Obesity," *PLoS ONE* 6(5): e19657. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0019657 (accessed 26th May 7, 2014), 2011.

Dasgupta, P., *An inquiry into well-being and destitution*, Oxford University Press, Oxford, 1993.

Dasgupta, P. and D. Ray, "Inequality as a determinant of malnutrition and unemployment: Theory," *Economic Journal*, 96, 1011–1034, 1986.

Dasgupta, P. and D. Ray, "Inequality as a determinant of malnutrition and unemployment: Policy," *Economic Journal*, 97, 177–188, 1987.

Deaton, A., "Panel Data from a Time Series of Cross Sections," *Journal of Econometrics*, 30, 109-126, 1986,

Deaton, A. and J. Drèze, "Food and Nutrition in India: Facts and Interpretations," *Economic and Political Weekly*, 44(7), 42-65, 2009.

Dubey, A. and S. Thorat, "Has Growth Been Socially Inclusive during 1993-94 - 2009-10?", *Economic and Political Weekly*, 47(10), 43-53, 2012.

Gaiha, R., R., Jha, and V. Kulkarni, "Demand for Nutrients in India, 1993-2004," *Applied Economics*, 45(14), 1869-1886, 2013.

Gaiha, R., R., Jha, and V. Kulkarni, Diets, Malnutrition and Disease-the Indian Experience, New Delhi: Oxford University Press, forthcoming, 2014.

George, A. “Obesity expert: A better fat measure than BMI ?”, An Interview with R. Bergman, New Scientist, 17 March, 2011.

International Institute for Population Sciences (IIPS) and Macro International, *National Family Health Survey (NFHS-3), 2005–06: India: Volume I.* IIPS, Mumbai, 2007.

Patnaik, U., “The Republic of Hunger,” *Social Scientist*, 32 (9-10), 9-35, 2004.

Patnaik, U., “Neoliberalism and Rural Poverty in India,” *Economic & Political Weekly*, 42(30), 3132-50, 2007.

Jha, R., R. Gaiha, S. Anurag, “Calorie and Micronutrient Deprivation and Poverty Nutrition Traps in Rural India,” *World Development*, 37(5), 982-991, 2009.

Jha, R., R. Gaiha, and M. K. Pandey, “Body Mass Index, Participation, Duration of Work and Earnings, under NREGS Evidence from Rajasthan,” *Journal of Asian Economics*, 26, 14–30, 2013.

Koenker, R. and G. Basset, Jr., “Regression Quantiles,” Econometrica, 46 (1), 33-50, 1978.

Krueger, D. and F. Perri, “Does Income Inequality lead to Consumption Inequality: Evidence and Theory,” *Review of Economic Studies*, 73(1), 163-193, 2006.

Leibenstein, H., *Economic Backwardness of Economic Growth*, Wiley, New York, 1957.

Palmer-Jones, R. and K. Sen, “On Indian Poverty Puzzles and Statistics of Poverty,” *Economic & Political Weekly*, 36(3), 211-217, 2001.

Rutstein, S. O. and K. Johnson, *The DHS Wealth Index. DHS Comparative Reports No. 6. Calverton*, ORC Macro, Maryland, 2004.

Strauss, J., "Does Better Nutrition Raise Farm Productivity?" *Journal of Political Economy*, 94(2), 297-320, 1986.

Strauss, J. and D. Thomas, "Health Nutrition and Economic Development," *Journal of Economic Literature*, 36(2), 766-817, 1998.

Swamy, A. V., "A simple test of the nutrition-based efficiency wage model," *Journal of Development Economics*, 53(1), 85-98, 1997.

Verbeek, M., "Pseudo Panel Data," in *The Econometrics of Panel data: A Handbook of the Theory with Applications, Second Edition, Advanced Studies in Theoretical and Applied Econometrics 33*, Matyas, L. and P. Sevestreyas (eds.), Kluwer Academic, Boston and London, pp.280-92, 1996.

Verbeek, M. and T. E. Nijman, "Can Cohort Data Be Treated As Genuine Panel Data?" *Empirical Economics*, 17, 9-23, 1992.

Weinberger, K., "The Impact of Micronutrients on Labor Productivity: Evidence from Rural India," in Proceedings of the 25th International Conference of Agricultural Economists. Durban, South Africa, 2003.

WHO expert consultation, "Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies", *The Lancet*, 363(9403), 157-163, 2004.

Wooldridge, J. M., *Introductory Econometric: A Modern Approach. 4th Edition*, South Western Educational Publishing, Mason, Ohio, 2009.

TABLE 1

Summary of Econometric Results for Rural Women: NFHS 1, 2 & 3 (1992-3, 1998-9, & 2005-6)

Explanatory Variables	(1)	(2)	(3)	(4)	(5)	(6)
		Quantile Regression$^{\$}$				
	OLS	Severely underweight	Under weight	Normal	Over weight	Obese
NFHS 1 (1992-3) Women						
Age	-0.03	-0.02	-0.02	-0.14	-0.27	0.31
	[-0.80]	[-0.92]	[-0.44]	[-1.06]	[-0.64]	[0.46]
Age Squared	0.00	0.00	0.00	0.00	0.00	-0.01
	[0.70]	[1.11]	[0.35]	[0.69]	[0.36]	[-0.50]
Wealth Scores	**0.44**	**0.22**	**0.32**	**1.41**	**2.20**	**3.20**
	[5.71]**	**[4.15]****	**[3.92]****	**[5.32]****	**[2.73]****	**[2.44]***
Education (Attempted or	0.02	0.06	0.01	0.03	-0.63	-1.41
Completed Primary)[1]	[0.25]	[0.85]	[0.10]	[0.10]	[-0.53]	[-0.62]
Education (Attempted or	0.04	**0.25**	0.09	-0.45	**-2.57**	**-4.54**
Completed Secondary)	[0.38]	**[4.07]****	[0.88]	[-1.42]	**[-2.62]****	**[-2.80]****
Education (Attempted or Comp-	0.02	0.30	0.11	**-1.69**	**-4.72**	-3.67
pleted Higher than Secondary)	[0.07]	[1.13]	[0.23]	**[-1.83]+**	**[-2.62]****	[-1.08]
Working for family member	0.10	**0.16**	0.02	-0.03	-1.13	-0.14
	[1.14]	**[2.95]****	[0.19]	[-0.13]	[-1.21]	[-0.05]
Working for someone else	**0.17**	**0.18**	0.07	0.20	-1.18	-2.05
	[1.95]+	**[2.91]****	[0.81]	[0.59]	[-1.26]	[-1.21]
Self employed	**-0.56**	-0.20	-0.35	**-0.88**	**-3.96**	**-7.30**
	[-3.60]**	[-1.24]	[-1.57]	**[-2.14]***	**[-3.99]****	**[-4.44]****
Price of Sugar (State level)	**-0.41**	**-0.15**	**-0.53**	**-1.10**	**-1.82**	**-4.40**
	[-6.07]**	**[-3.13]****	**[-6.41]****	**[-5.79]****	**[-3.26]****	**[-3.03]****
Price of eggs (State level)	0.02	0.01	**0.04**	0.07	0.22	0.82
	[1.05]	[0.54]	**[2.14]***	[0.99]	[1.05]	[2.35]*
Price of Cereals (State level)	**0.02**	**0.03**	0.00	-0.02	0.09	-0.21
	[2.05]*	**[3.33]****	[0.16]	[-0.35]	[0.72]	[-1.00]
N	10336	10336	10336	10336	10336	10336
	Adj. R^2	0.034; F-statistics	9.43; Log-likelihood -2.6e+04			
NFHS 2 (1998-9) Women						
Age	**0.05**	-0.01	-0.01	0.02	0.05	0.27
	[4.08]**	[-0.24]	[-0.98]	[0.83]	[1.69]+	[2.72]**
Age Squared	-0.00	-0.00	0.00	0.00	0.00	-0.00
	[-0.75]	[-0.65]	[1.07]	[1.51]	[2.30]*	[-1.18]
Wealth Scores	**1.03**	**0.35**	**0.60**	**1.02**	**1.80**	**2.27**
	[26.32]**	**[7.87]****	**[12.95]****	**[21.96]****	**[13.69]****	**[5.69]****
Education (Attempted or	**0.21**	-0.03	-0.01	**0.25**	**0.46**	**1.30**
Completed Primary)[1]	**[4.55]****	[-0.54]	[-0.14]	**[6.93]****	**[3.25]****	**[2.10]***
Education (Attempted or	**0.19**	-0.09	-0.02	**0.28**	**0.45**	0.35
Completed Secondary)	**[3.63]****	[-1.10]	[-0.37]	**[2.64]****	**[2.53]***	[0.60]
Education (Attempted or Comp-	**0.43**	-0.02	**0.30**	**0.63**	**0.95**	1.39
leted Higher than Secondary)	**[3.40]****	[-0.09]	**[2.26]***	**[2.70]****	**[2.36]***	[1.31]
Working for family member	**-0.11**	0.12	**0.08**	**-0.09**	**-0.38**	**-1.22**
	[-2.84]**	[1.64]	**[2.54]***	**[-2.37]***	**[-2.41]***	**[-3.74]****
Working for someone else	**-0.25**	0.03	**-0.09**	**-0.23**	**-0.63**	**-1.21**
	[-5.99]**	[0.60]	**[-1.82]+**	**[-4.67]****	**[-4.63]****	**[-2.48]***
Self employed	-0.02	0.18	**0.14**	-0.08	0.03	-0.76
	[-0.33]	[1.29]	**[2.13]***	[-0.69]	[0.12]	[-1.04]
Price of Sugar (State level)	0.03	**-0.08**	-0.00	0.05	**0.17**	0.04
	[1.16]	**[-1.90]+**	[-0.09]	[1.61]	**[1.90]+**	[0.13]
Price of eggs (State level)	**-0.03**	**-0.03**	**-0.03**	**-0.02**	-0.02	0.03
	[-3.68]**	**[-2.25]***	**[-2.39]***	**[-2.40]***	[-0.62]	[0.34]
Price of Cereals (State level)	**0.02**	**0.03**	**0.04**	**0.03**	0.01	-0.03
	[4.43]**	**[3.86]****	**[9.25]****	**[5.01]****	[1.01]	[-0.87]
N	36227	36227	36227	36227	36227	36227
	Adj. R^2	0.128; F-statistics	116.71			
NFHS 3 (2005-6) Women						
Age	**0.08**	-0.01	-0.01	**0.05**	**0.25**	**0.57**
	[7.02]**	[-0.62]	[-0.63]	**[3.65]****	**[7.14]****	**[6.13]****
Age Squared	**-0.00**	0.00	**0.00**	0.00	**-0.00**	**-0.01**

	[-1.83]+	[0.15]	**[1.82]+**	[1.08]	**[-3.07]****	**[-3.81]****
Wealth Scores	**0.12**	**0.05**	**0.08**	**0.12**	**0.17**	**0.21**
	[31.58]**	**[8.90]****	**[21.55]****	**[24.58]****	**[18.63]****	**[5.99]****
Education (Attempted or Completed Primary)[1]	**0.43**	**0.15**	**0.23**	**0.39**	**0.54**	**0.62**
	[10.32]**	**[2.12]***	**[6.33]****	**[7.90]****	**[4.21]****	**[2.01]***
Education (Attempted or Completed Secondary)	**0.33**	0.10	**0.13**	**0.30**	**0.30**	0.35
	[8.26]**	[1.50]	**[3.38]****	**[6.25]****	**[3.10]****	[1.04]
Education (Attempted or Completed Higher than Secondary)	0.06	-0.13	**0.22**	0.17	-0.39	-0.18
	[0.58]	[-0.77]	**[2.23]***	[1.49]	[-1.63]	[-0.27]
Working for family member	**-0.31**	0.03	**-0.08**	**-0.25**	**-0.83**	**-1.48**
	[-9.08]**	[0.58]	**[-2.04]***	**[-5.81]****	**[-8.72]****	**[-5.74]****
Working for someone else	**-0.25**	**0.13**	**-0.07**	**-0.25**	**-0.65**	**-0.96**
	[-6.24]**	**[2.08]***	**[-1.72]+**	**[-5.12]****	**[-6.85]****	**[-2.77]****
Self employed	**-0.23**	0.10	-0.01	**-0.24**	**-0.67**	**-1.47**
	[-4.23]**	[0.95]	[-0.20]	**[-4.06]****	**[-4.27]****	**[-4.47]****
Price of Sugar (State level)	**-0.04**	**0.07**	-0.00	**-0.05**	**-0.18**	**-0.62**
	[-1.79]+	**[1.84]+**	[-0.16]	**[-1.82]+**	**[-2.11]***	**[-3.51]****
Price of eggs (State level)	**-0.05**	**-0.03**	**-0.03**	**-0.05**	**-0.03**	0.02
	[-6.72]**	**[-3.07]****	**[-3.68]****	**[-4.98]****	**[-1.98]***	[0.32]
Price of Cereals (State level)	**0.02**	**0.01**	**0.02**	**0.02**	**0.02**	-0.01
	[5.05]**	**[2.50]***	**[8.12]****	**[6.96]****	**[2.33]***	[-0.66]
N	51888	51888	51888	51888	51888	51888
	Adj. R^2	0.173; F-statistics	236.62			

t statistics in brackets ----- + p<.10, * p<.05, ** p<.0; Base categories: [1] Education (None); [2] Work Status(Not working) [3] Marital Status (Never Married); [4] Religion (Other); [5] Social Group (Non-Backward Group); [6] Location (North); [$] Quantile regression standard errors are bootstrapped based on 100; [a] The median z-score and the corresponding percentile for the group; Only key explanatory variables are selected; Significant coefficient estimates are shown in bold.

TABLE 2
Summary of Econometric Results for Rural Men: NFHS 3 (2005-6)

Explanatory Variables	(1) OLS	(2) Quantile Regression[$] Severely Underweight	(3) Under weight	(4) Normal	(5) Over weight	(6) Obese
NFHS 3 (2005-6) Men						
Age	**0.29**	**0.20**	**0.23**	**0.28**	**0.36**	**0.53**
	[23.54]**	**[8.43]****	**[18.05]****	**[18.43]****	**[11.97]****	**[6.38]****
Age Squared	**-0.00**	**-0.00**	**-0.00**	**-0.00**	**-0.00**	**-0.01**
	[-20.03]**	**[-8.41]****	**[-16.48]****	**[-15.34]****	**[-9.46]****	**[-5.23]****
Wealth Scores	**0.12**	**0.04**	**0.06**	**0.11**	**0.18**	**0.27**
	[25.75]**	**[5.15]****	**[14.67]****	**[21.95]****	**[19.39]****	**[8.79]****
Education (Attempted or Completed Primary)[1]	**0.23**	-0.03	**0.11**	**0.26**	**0.31**	0.24
	[4.61]**	[-0.34]	**[2.12]***	**[4.01]****	**[2.49]***	[0.86]
Education (Attempted or Completed Secondary)	**0.38**	0.04	**0.19**	**0.32**	**0.47**	0.46
	[7.92]**	[0.42]	**[3.89]****	**[5.45]****	**[4.54]****	[1.59]
Education (Attempted or Comp-leted Higher than Secondary)	**0.89**	**0.36**	**0.64**	**1.05**	**1.08**	0.22
	[10.29]**	**[2.34]***	**[7.53]****	**[10.73]****	**[6.93]****	[0.42]
Price of Sugar (State level)	0.00	**0.12**	**0.07**	-0.04	**-0.22**	-0.07
	[0.04]	**[2.11]***	**[2.21]***	[-1.00]	**[-2.84]****	[-0.33]
Price of eggs (State level)	**-0.05**	-0.01	-0.01	**-0.04**	-0.03	-0.08
	[-4.22]**	[-0.69]	[-1.09]	**[-3.34]****	[-1.04]	[-1.25]
Price of Cereals (State level)	0.00	-0.00	0.01	**0.01**	0.01	0.01
	[0.79]	[-0.09]	[1.34]	**[3.05]****	[1.09]	[0.52]
N	28705	28705	28705	28705	28705	28705
Adj. R^2	0.200	-	-	-	-	-
F-statistics	204.10	-	-	-	-	-
Log-likelihood	-7.0e+04	-	-	-	-	-

t statistics in brackets ----- + p<.10, * p<.05, ** p<.0; Base categories: [1] Education (None); [2] Marital Status (Never Married); [3] Religion (Other); Social Group (Non-Backward Group); [5] Location (North); [$] Quantile regression standard errors are bootstrapped based on 100; [a] The median z-score and the corresponding percentile for the group; Only key explanatory variables are selected; Significant coefficient estimates are shown in bold.

TABLE 3

Econometric Results: National Family Health Survey: Female-Rural Sample (Pseudo Panel for 1992-3, 1998-9 and 2005-6)

Explanatory variables	**(1)** Pooled	**(2)** Fixed Effects
Age	**0.15**	**2.91**
	[6.49]**	**[2.81]****
Age Squared	**-0.00**	-0.03
	[-3.64]**	[-1.62]
Wealth Score	**0.07**	**0.19**
	[10.63]**	**[1.97]***
Proportion of cohort with no education	-	**-2.39**
	-	**[-1.83]+**
Education (Attempted or Completed Primary)[1]	**0.48**	-
	[3.57]**	-
Education (Attempted or Completed Secondary)	**0.51**	-
	[4.69]**	-
Education (Attempted or Completed Higher than Secondary)	**0.42**	-
	[3.93]**	-
Self employed[2]	-0.10	-3.25
	[-0.56]	[-1.21]
Working for someone else	-0.20	**-4.48**
	[-1.37]	**[-2.42]***
Working for family member	-0.12	-1.62
	[-1.39]	[-1.32]
Price of Sugar (State level)	-0.04	-
	[-1.42]	-
Price of eggs (State level)	-0.05	-
	[-3.56]**	-
Price of Cereals (State level)	0.01	-
	[2.98]**	-
Survey round dummy (1998 -9)[6]	**4.25**	**4.61**
	[33.47]**	**[9.29]****
Survey round dummy (2006 -3)	**4.54**	**5.27**
	[68.61]**	**[5.62]****
Constant	**15.25**	**-50.36**
	[21.40]**	**[-2.79]****
N	103536	378
Adj. R^2	0.024	0.581
F-statistics	860.57	30.47
Correlation between the error and regressors	0.97	-
Hausman Test	-	41.92(0.00)
F-test (Unobserved heterogeneity=0)	-	1.26(0.07)
Test for time effect	-	45.67(0.00)

t statistics in brackets ----- + p<.10, * p<.05, ** p<.0; Base categories: [1] Education (None); [2] Work Status(Not working) [3] Religion (Other); [4] Social Group (Non-Backward Group); [5] Location (North);[6] Round Dummy(1993 -1); Only key explanatory variables are selected; Significant coefficient estimates are shown in bold.

TABLE 4
Econometric Results: NCAER data– Rural Sample (2005)[1]

Explanatory variables	(1) OLS[£]	(2) OLS[$]	(3) IV[α]	(4) OLS (Based on Heckman wage regression)[π]
Log of Wage	**0.43**	**0.40**	**1.84**	-
	[6.24]**	**[5.69]****	**[3.12]****	-
Predicted log of Wage	-	-	-	**1.40**
	-	-	-	**[3.05]****
Sex(Male)	-0.12	-0.03	**-0.66**	**-0.55**
	[-0.56]	[-0.15]	**[-2.00]***	**[-1.79]+**
Age years	**0.07**	**0.07**	**0.06**	0.02
	[2.75]**	**[2.55]***	**[2.34]***	[0.68]
Age years squared	**-0.00**	**-0.00**	**-0.00**	-0.00
	[-2.21]*	**[-2.07]***	**[-2.04]***	[-0.83]
Education (Lower Primary)[2]	-0.01	-0.03	-0.03	-0.09
	[-0.06]	[-0.23]	[-0.22]	[-0.78]
Education (Upper Primary)	0.11	0.08	0.02	-0.12
	[1.10]	[0.80]	[0.14]	[-0.93]
Education (Secondary)	**0.32**	0.17	0.00	-0.25
	[2.09]*	[1.08]	[0.01]	[-1.17]
Education (Higher Secondary)	-0.13	-0.51	**-0.90**	**-1.10**
	[-0.42]	[-1.58]	**[-2.45]***	**[-2.77]****
Education (Undergraduate)	-0.34	-0.89	**-1.64**	**-1.53**
	[-0.53]	[-1.20]	**[-1.87]+**	**[-1.93]+**
Education (Graduate)	**0.89**	0.28	-0.61	-0.93
	[2.86]**	[0.74]	[-1.08]	[-1.52]
Scheduled Caste	**-0.17**	**-0.17**	**-0.23**	**-0.18**
	[-2.44]*	**[-2.38]***	**[-2.93]****	**[-2.51]***
Household size	**-0.14**	**-0.14**	**-0.11**	**-0.14**
	[-3.03]**	**[-3.04]****	**[-2.28]***	**[-2.95]****
Household Size Squared	**0.01**	**0.01**	**0.01**	**0.01**
	[2.80]**	**[2.82]****	**[1.99]***	**[2.70]****
Distance to market	**-0.01**	**-0.01**	**-0.01**	**-0.01**
	[-1.92]+	**[-1.95]+**	**[-2.25]***	**[-1.64]**
BIMARU[2]	**-0.69**	**-0.63**	0.05	-0.18
	[-3.87]**	**[-3.50]****	[0.14]	[-0.65]
South	**-0.80**	**-0.75**	-0.14	**-0.42**
	[-4.69]**	**[-4.31]****	[-0.44]	**[-1.78]+**
East	**-0.94**	**-0.88**	-0.24	**-0.51**
	[-5.18]**	**[-4.83]****	[-0.73]	**[-1.99]***
Others	**-1.33**	**-1.40**	**-1.61**	**-1.52**
	[-1.75]+	**[-1.76]+**	**[-1.98]***	**[-1.88]+**
Price of cereals	**-0.01**	**-0.01**	**-0.02**	-0.01
	[-1.74]+	**[-1.82]+**	**[-2.51]***	[-0.99]
Activity Intensity	-	-0.44	0.19	**-0.58**
(Clerical/Sales/Services)[3]	-	[-1.28]	[0.41]	**[-1.69]+**
Activity	-	**-0.84**	0.30	**-1.09**
Intensity(Farmers/Fishermen)	-	**[-2.45]***	[0.49]	**[-3.26]****
Activity Intensity (Laborers and	-	**-1.01**	-0.18	**-1.21**
Production workers)	-	**[-2.94]****	[-0.34]	**[-3.57]****
Constant	18.95	19.91	16.39	19.47
	[36.62]	[31.55]	[10.17]	[24.93]
N	5811	5803	5790	5791
Adj. R^2	0.021	0.023	-0.045	0.019
F-statistics	8.54	7.93	5.70	6.74
Hausman Test	-		9.30(0.98)	
Under identification test	-		78.91(0.00)	
Weak identification test	-		48.87	-
Over identification test	-		0.51(0.48)	-

t statistics in brackets ---- + p<.10, * p<.05, ** p<.01; [1] The sample has been restricted to BMI less than 25; [2] Reference category for education is No education; [3] Base group for activity intensity is Professionals/Managers; [£] Basic OLS without activity intensity variables; [$] Basic OLS with activity intensity variables; [α] The instruments for correcting for wage endogeneity are the available of a trade union in the village and state level gini coefficient [π] Exclusion restriction variables used for the probability get employed is household's infant dependency ratio.; Significant coefficient estimates are shown in bold.

Appendix: First Stage IV and Heckman Selection Results

	(1)	(2)	(3)
	IV	Heckman	
		Participation Eq.	Outcome Eq.
Exp. Var. / Dep. Var.	Log of Hourly Wages	Employed or otherwise	Log of Hourly Wages
Availability of Trade Union in a village	0.16	-0.04	0.17
	[18.15]**	[-3.01]**	[18.19]**
District level consumption inequality	0.36	-0.10	0.47
	[13.25]**	[-2.18]*	[15.62]**
Infant dependency Ratio	-	0.64	-
	-	[16.78]**	-
Gender (whether male)	0.35	1.00	0.47
	[51.07]**	[91.63]**	[16.25]**
Age	0.02	0.14	0.03
	[18.60]**	[59.37]**	[6.85]**
Age Squared	0.03	-0.00	0.04
	[3.40]**	[-58.63]**	[3.30]**
Education (Lower Primary)[1]	0.11	-0.24	0.16
	[13.39]**	[-13.19]**	[11.41]**
Education (Upper Primary)	0.20	-0.40	0.32
	[20.65]**	[-29.25]**	[15.52]**
Education (Secondary)	0.27	-0.66	0.50
	[17.94]**	[-41.80]**	[17.40]**
Education (Higher Secondary)	0.31	-0.87	0.60
	[6.12]**	[-39.06]**	[9.11]**
Education (Undergraduate)	0.49	-1.24	1.02
	[24.53]**	[-19.06]**	[40.83]**
Education (Graduate)	-0.00	-0.63	-0.00
	[-13.81]**	[-22.86]**	[-4.70]**
Scheduled Caste	-0.00	0.43	0.01
	[-0.54]	[35.89]**	[1.08]
Household Size	-0.00	-0.09	-0.02
	[-0.33]	[-19.32]**	[-4.00]**
Household Size Squared	0.00	0.00	0.00
	[2.64]**	[5.70]**	[4.92]**
Distance to Market	-0.00	-	-
	[-6.86]**	-	-
BIMARU[2]	-0.40	0.24	-0.41
	[-37.16]**	[14.36]**	[-30.11]**
South	-0.28	0.52	-0.37
	[-27.40]**	[33.32]**	[-20.60]**
East	-0.38	0.33	-0.39
	[-34.93]**	[19.31]**	[-25.99]**
Others	-0.01	0.31	0.03
	[-0.19]	[5.09]**	[0.59]
Price of cereals	0.01	-	-
	[11.00]**	-	-
Activity Intensity (Clerical/Sales/Services)[2]	-0.40	-	-
	[-21.74]**	-	-
Activity Intensity(Farmers/Fishermen)	-0.89	-	-
	[-50.50]**	-	-
Activity Intensity (Laborers and Production	-0.60	-	-
workers)	[-34.35]**	-	-
Constant	1.73	-2.74	0.67
	[42.93]**	[-46.13]**	[5.47]**
Correlation between the error terms in the	-	-	2.89
participation and outcome equations	-	-	[0.00]**
N	30852	80707	80707
Adj. R^2	0.443	-	-
F-Statistics	1066.57	-	-

t statistics in brackets + p<.10, * p<.05, ** p<.01; [1] Reference category for education is No education; [2] Base group for activity intensity is Professionals/Managers.