PRELIMINARY DRAFT: PLEASE DO NOT QUOTE

Revised March 2005

Poverty Nutrition Trap in Rural India*

Raghbendra Jha, Australian National University

Raghav Gaiha, University of Delhi

Anurag Sharma, Australian National University

ABSTRACT

The contribution of the present paper is threefold. First, we formally test whether the effect of calorie deprivation on wages is more significant/higher for the lower quantiles of workers. In the extant literature this is established through non-linear terms in the wage equation. A more satisfactory method of doing this is through quantile regressions. Second, the quantile regression approach helps us identify the exact group for which the poverty-nutrition trap holds. The extant literature is unable to establish whether there are systematic differences across different quintiles in the response of productivity/wages to nutrition. The present paper addresses this lacuna. Third, we are able to establish a critical wage level for which the PNT trap hypothesis holds. For wages higher than this the hypothesis does not hold. We then argue that this value of the wage rate should set a floor for any minimum wage for agricultural labourers.

JEL codes: I 12, I 31, I 32, J 41
Key words: nutrition, calories, wages, poverty trap, labourers

All correspondence to:
Prof. Raghbendra Jha,
ASARC, Division of Economics,
Research School of Pacific and Asian Studies,
Australian National University,
Canberra ACT 0200, Australia
Phone: + 61 2 6125 2683
Fax: +61 2 6125 0443
Email: r.jha@anu.edu.au

* The UK Department for International Development (DFID) supports policies, programmes and projects to promote international development. DFID provided funds for this study as part of that objective but the views and opinions expressed are those of the author(s) alone. The authors would like to thank DFID for supporting this research. The present draft has benefited from the advice of C. J. Bliss, D. Newbery, J. Behrman, P. Bardhan, K. L. Krishna, A. de Janvry, P. Scandizzo, M. Lipton and Anil Deolalikar. We would also like to thank Hari Nagarajan for making available to us NCAER's data set on rainfall in various agro-climatic zones of India. The paper was presented at seminars in the Australian National University and the Australian Bureau of Agricultural and Resource Economics. Jha would like to thank seminar participants in both institutions for their comments.

I. Introduction

The effect of nutritional intake on labour productivity and wage rates has been an important area of research for economists and nutritionists for some time. This found initial expression in the form of the efficiency wage hypothesis developed by Leibenstein (1957) and Mazumdar (1959) and formalized and extended by Mirrlees (1975), Stiglitz (1976), Dasgupta and Ray (1986, 1987), and Dasgupta (1993), among others. Early surveys include Bliss and Stern (1978a, 1978b) and Binswanger and Rosenzweig (1984). The efficiency wage hypothesis postulated that in developing countries, particularly at low levels of nutrition, workers are physically incapable of doing hard manual labour. Hence their productivity is low which then implies that they get low wages, have low purchasing power and, therefore, low levels of nutrition, completing a vicious cycle of deprivation. These workers are unable to save very much so their assets –both physical and human – are minimal. This reduces their chances of escaping the poverty-nutrition trap (henceforth PNT).[1]

There is a substantial literature on empirically testing for the existence of PNT.[2] Strauss (1986) models the effect of nutrition on farm productivity. He tests and quantifies the effects of nutritional status as measured by annual calorie intake on annual farm production and, hence, labour productivity using farm household level data from Sierra Leone. He finds significant and sizable effect of calorie intake on farm output, even after accounting for endogeneity. These effects are stronger at lower levels of calorie intake – although this is determined through the presence of non-linear terms rather than a quantile regression approach.

[1] In this paper we use the terms efficiency wage hypothesis and PNT interchangeably.
[2] For a comprehensive review see Strauss and Thomas (1998).

Thomas and Strauss (1997) investigate the impact of four indicators of health (height, body mass index, per capita calorie intake and per capita protein intake) on wages of urban workers in urban Brazil. They discover that even after accounting for endogeneity issues and controlling for education and other dimensions of health, these four indicators have significant positive effects on wages. The effect of the nutritional variables - per capita calorie intake and per capita protein intake – was higher at low levels of nutrition, again determined through non-linear terms. In contrast Deolalikar (1988) finds in a (panel fixed effects) joint regression of the wage equation and farm production in rural South India that calorie intake does not affect either but a measure of weight-for-height does. He concludes that calorie intake does not affect wages or productivity indicating that the human body can adapt to short-run shortfalls in calorie intake. However, the fact that weight-for-height affects wages and productivity indicates that chronic undernutrition is an important determinant of productivity and wages.

The contribution of the present paper is threefold. First, we formally test whether the effect of calorie deprivation on wages is more significant/higher for the lower quantiles (in terms of wage rates) of workers. In the extant literature this is established through non-liner terms in the wage equation. A more satisfactory method of doing this is through quantile regressions. Second, the quantile regression approach helps us identify the exact group for which the poverty-nutrition trap holds. The extant literature is unable to establish whether there are systematic differences across different quantiles in the response of productivity/wages to nutrition. The present paper addresses this lacuna. Third, we are able to establish a critical wage level for which the PNT trap hypothesis holds. For wages higher than this the hypothesis does not hold. We then argue that this value of the wage rate should set a floor for any minimum wage for agricultural labourers.

The plan of this paper is as follows. In section II, a sketch of the relationship between nutrition and work capacity is given to motivate the nutrition poverty trap. In Section III we describe the model and data. Section IV presents and discusses the results, in particular establishing that a certain proportion of the population is caught in the poverty-nutrition trap. Section V argues that the wage rate for which this occurs is an appropriate benchmark for setting the minimum wage for agricultural workers (or for pursuing alternative policies that guarantee the attainment of the level of nutrition associated with this wage). It then portrays some characteristics of the population in this category. Section VI concludes.

II. Nutrition Poverty Traps

In Figure 1, a stylised version of the relationship between work capacity and nutrition is given.[3] The vertical axis represents a measure of work capacity and the horizontal axis income. Note first that work capacity is a measure of the tasks that an individual can perform during a period, say, the number of bushels of wheat that he can harvest during a day. Income is used synonymously with nutrition in the sense that all income is converted into nutrition. Nothing of importance changes if 70 or 80 per cent of income share is spent on nutrition.

The shape of the capacity curve requires an explanation. It is assumed here that much of the nutrition goes into maintaining the body's resting metabolism. This refers to the energy required to maintain body temperature, sustain heart and respiratory action, and to support the ionic gradients across cell membranes. For the "reference man" of the Food and Agriculture organisation (FAO)- a European

[3] The following exposition is based on Ray (2004).

male weighing 65 kg-the requirement is 1700 calories per day. Of course the requirement varies with the individual and the environment in which he lives. In the case of India Gopalan et al. (1971) indicate that for men doing sedentary, moderate and heavy work the calorie requirements per day are, respectively 2400, 2800 and 3900. A higher body mass, for example, raises resting metabolism. Another significant component is energy required to carry out physical labour. The FAO's estimate, applied to their reference man, prescribed 400 kcal per day for "moderate activity". It is of course arguable that for the poor in developing countries this may be an underestimate. Once resting metabolism is taken care of, however, there is a marked increase in work capacity, as the bulk of the energy input goes into work. This phase is followed by a phase of diminishing returns, as the body's frame restricts conversion of nutrition into work capacity.

Figure 1 here

Assume that income is generated by working in a labour market, where piece rates are paid. A piece rate, then, appears as a relationship between the number of tasks performed and the total income of a person. Using these assumptions, a supply curve of labour could be constructed that shows different quantities of labour supplied at different piece rates. Aggregation across individuals yields an aggregate supply curve, as shown in Figure 2.

Figure 2 here

At a piece rate of v_3 there is a gap in labour supply and a discontinuous jump. Introducing a downward sloping demand curve, an interesting case is that in which the demand curve passes through the dotted supply curve. If the piece rate is larger than v^*, there is excess supply, which lowers this rate. On the other hand, if the piece rate is lower than v^*, there is excess demand, so that wages rise.

Note, however, that a piece rate of v^* is an equilibrium wage, provided we allow for unemployment.

Figure 3 here

The gap in labour supply could be filled by having some people work and restricting labour market access to others. Those rationed out will be relatively undernourished. This completes the vicious cycle of poverty. Lack of labour market opportunities results in low wages and consequently low work capacity; a low work capacity feeds back by lowering access to labour markets. It is easy to show that higher non-labour assets (e.g. land) lead to higher wage incomes. Thus the poor without assets are doubly disadvantaged: not only do they not enjoy non-labour income but also have restricted access to labour market opportunities.

Note that nutritional status depends on both current consumption of nutrients (e.g. calories) and the history of that consumption. In the analysis that follows, we focus on the effects of differences in calorie intake.[4]

III. The Model and Data

The essence of an empirical test for the PNT Hypothesis is the specification of a wage equation conditional on energy intake and control variables as:

$$Lnw_h = f(calorie_h, p_1, p_2, p_3, p_4, X) \qquad (1)$$

where w_h and 'calorie' represents the wage and calorie intake of the h^{th} individual.[5] p_i is the probability of being occupied in the ith occupation with i =1 indicating employment in agriculture, i=2 employment in non-agriculture, i=3 self employment and i = 4 other employment. This set of variables controls for labour market participation. 'X' represents control variables such as prices of various food products, income of the household from the non-agricultural sector,

[4] For critiques of PNT hypothesis, see Srinivasan (1994), and Subramaniam and Deaton (1996).
[5] In actual estimation it is common to add a square term in calories as well.

some household characteristics as well as some regional dummies. The probabilities are taken as the control variables to incorporate the impact of labour market participation on the wage rate. It is thus argued that the wage rate of the worker depends on his nutrition proxied as his energy intake, which in turn depends on his wages. Hence the wage rate and nutritional intake are endogenous in this model.

In view of this endogeneity the empirical strategy followed in this paper is as follows: The empirical analysis is done using the instrumental variables approach where calorie intake is assumed endogenous and hence instrumented in the fist stage. The second stage subsequently uses the estimated value of the calorie intake to estimate (1). The probabilities of labour market participation are predicted from a Maximum Likelihood Multinomial Logistic Regression (multi-logit) model discussed next.

Multi Logit Model:

The polychotomous dependent variable *employed* takes four values: 1 if worker employed in agriculture, 2 if worker employed in non-agriculture, 3 if worker self-employed and 4 if worker employed in other sectors. When there is no risk of confusion we will refer to agriculture, non-agriculture, self-employment and other sectors as occupation 1, 2, 3, and 4 respectively. The independent variables for the analysis can be broadly classified into two categories: Household level variables (which mainly include household characteristics) and Location Dummies to incorporate the role of regional disparity. These household and other variables are summarized in Table 1.

Table 1 here.

The predicted probabilities of participating in the labour market are calculated from the above regression and used subsequently in the instrumental variables regression discussed next.

Instrumental Variable Estimation:

W_t = f (calorie, Hhcar, P1, P2) (2)

The IV estimation is a two-stage regression.

Fist Stage IV Regression:

The first stage regresses calories and calories squared on *Food price* (Price of various food products) and *Nonlab* = Income of the HH from non- agriculture sector, in addition to the other independent variables used in (2). The IV technique then calculates *caloriehat* and *caloriehat2.*

Thus the following regression is used to calculate the predicted value of calorie intake:

Enepchat = f(edfem_edu_2, _edfem_edu_3, _edfem_edu_4, headage, headage2, amale, afemale, hhsize, hhgrp, _relreligi~1,_relreligi~2, _relreligi~3, relreligi~4, _relreligi~6, p1, p2,price_pulses, price_gur, _~r, price_edib~l, price_milk, Nonlabincome)
(3)

where Enepchat is per capita energy consumption within the household and the independent variables (other than those appearing in Table 1) can be broadly categorised as in Table 2.

Table 2 here.

An equation similar to (3) is estimated[6] for enepchat2 (square of consumption of energy per capita).

Second Stage IV regression

The second stage IV regression uses the predicted value of calorie intake to estimate (1) i.e.

Ln w_h =f(enepchat, Enepchat2 , _edfem_edu_2, _edfem_edu_3, _edfem_edu_4, headage, headage2, amale, afemale, hhsize, hhgrp,_relreligi~1, _relreligi~2, _relreligi~3, _relreligi~4, _relreligi~6, p1,p2) (4)

We estimate this model using data from the National Council for Applied Economic Research (NCAER). This data were collected through a multi-purpose household survey spread over six months, from January to June 1994. The data were collected using varied reference periods based on some conventional rules. The wage data used is that for harvesting since the number of data points for wages for other occupations were limited. To be consistent the analysis works with data associated with positive harvesting wage. This amounts to observations on 4640 adult male workers.

In the context of this data set this estimation strategy closely resembles the extant approach to estimating the PNT. We then go one step further by estimating the last stage as a set of quantile regressions using the same variables. This has the effect of considerably sharpening the identification of the incidence of the PNT. Furthermore, quantile regressions have certain inherent advantages over the least squares alternative (Buchinsky 1998, Yu et al. 2003).[7] We find that the PNT holds

[6] Results on this estimation are not reported here to conserve space.

[7] At least six such advantages are mentioned in the literature. (i) The quantile regression model can be used to characterize the entire distribution of the dependent variable for a given set of regressors. (ii) The qunaitle regression model has a linear programming interpretation, which makes estimation easy. (iii) The objective function for quantile regression is the weighted sum of absolute deviations. This gives a robust measure of the position of the observations considered in the sample and hence the estimated coefficients are not sensitive to outlier observations of the dependent variable. (iv) Quantile regression estimators are more efficient than least squares

only up to the bottom 60th percentile of the workers considered here and, for the rest, it does not.

"Quantile" is a generic term for dividing the population into segments of population, e.g., percentiles, and deciles. Suppose that the ?th quantile of a population is m? where 0<?<1. If F_N is the cumulative distribution function of y, the variable of interest, then m? is defined by:

$$q = \Pr[y \le m_q] = F_N(m_q) \qquad (5)$$

For a sample the analogous expression for defining $\hat{m}_q$ is

$$\hat{m}_q = \inf[y : F_N(y)] \ge q \qquad (6)$$

Hence, for example, these equations say that in a group of workers with varying wages, a worker earns more than the proportion ? of the reference group of workers and worse than the proportion (1-?). The median case is where ? = ½

Qunatile regression is a generalization of the concept of ordinary quantiles. Consider a sample (y_i, x_i), i=1,…,n from a population where x_i is an Kx1 vector of regressors. Then it is assumed that:

$$y_i = x_i' b_q + u_{qi} \qquad (7)$$

where u?i is the error term such that Quant?(u?i)=0. Thus,

$$Quant_q[y_i / x_i] = x_i' b_q \qquad (8)$$

where Quant?(y_i/x_i) represents the conditional quantile of y_i , conditional upon the set of independent variable vector x_i. The assumption that Quant?(u?i/x_i)=0 implies that the distribution term u?i only satisfies the assumption that the ?th

estimators if the error terms are not normally distributed. (v) Different solutions at distinct quantiles may be inferred as differences in the response of the dependent variable to changes in the regressors at various points in the distribution of the dependent variable. (vi) Quantile regression is more stable than mean regression for analysing contaminated data. Yu and Jones (1998) show that the variance of a typical kernel smoother is greater than the variance of a smooth quantile regression curve.

quantile of v_i, i.e., (y_i - x_iß?), conditional upon the vector of regressors, is equal to zero – no assumption need be made about the distribution of the error term. The ?[th] quantile regression result is the solution to the following minimization problem:

$$\min_b \frac{1}{n}[\sum_{i:y_i \geq x_i'b} abs(q\, y_i - x_i'b) + \sum_{i:y_i < x_i'b}(1-q)abs(y_i - x_i'b)] \qquad (9)$$

where *abs* refers to absolute value.

IV. <u>Results</u>

Results on the Multinomial logit regression are reported in Table 3.

Table 3 here

P3 is the omitted category in this regression. The results show that the older the head of the household the lower probability of being employed in occupations 1, 2, and 4. In each case, however, this effect works at a decreasing rate. The gender composition of the household is not a significant determinant except in occupation 4. Religious affiliations are not a significant determinant of occupational choice nor are Bimaru, coastal and non-coastal dummies. However, land owned increases (at a decreasing rate) the probability of being employed in occupations 1, 2 and 4. This is intuitively plausible as the greater the amount of land owned the higher the probability of cultivating one's own land to the exclusion of other occupations. As expected, rainfall positively affects P1 (as well as P4) but not P2.

Results on the first stage of the analysis (for enepc) for male_harvest wage are reported in Table 4. Having adult female with only primary education does not significantly affect calorie consumption whereas if the adult female has a higher level of education then calorie consumption increases. Per capita energy consumption rises (but at a decreasing rate) with the age of the head of the household. Per capita calorie consumption goes down with an increase in the number of adult males and adult females in the household. Religious affiliation 1,2 and 3 negatively affect per capita calorie consumption whereas the effect of religion 6 is insignificant. Higher participation rates in occupations 1 and 2 positively affect calorie consumption. Of the prices retained in the regression prices of pulses and edible oil negatively affect calorie consumption whereas the price of gur has a positive and significant effect and the price of milk has an insignificant effect. Non-labour income is also insignificant.

Table 4 here.

Results on the second stage of regression of the wage rate on predicted value of energy per capita and other control variables are reported in Table 5.

Table 5 here.

Table 5 indicates that female education is an insignificant determinant of the wage whereas the wage drops (at a decreasing rate) with the age of the head of the household (older workers are perceived to be able to exert less effort). An increase in the number of adults (male or female) in the household reduces the wage rate. Religious affiliation does not have a significant effect on the wage rate whereas the participation rates in agricultural and non-agricultural occupations are significant and positive. But, the most important result in Table 5 is that the coefficients of both predicted energy per capita as well as predicted energy pre capita squared are insignificant. If this was indeed true then the results indicate that PNT hypothesis does not hold in the case of rural India and confirm the conclusions of Deolalikar (1988).

This tentative conclusion, however, does not stand up to scrutiny when we do quantile regressions. Using the fitted values enepchat and enepchat2 from stage 1 of the regression we ran quantile regression[8] of (4) for various quartiles of male harvest wage. Upto the 60th quartile the sign of enepchat is positive[9] and significant. After that it turns negative and insignificant. We report results for the 60th quantile in Table 6 and for some other quantiles in the Appendix.

Table 6 here.

The pattern of dependence varies among the various quantiles but the positive and significant dependence of the wage rate on energy consumption remains unabated until the 60th quantile. Further, once this positive association is broken (at the 61st quantile) it is not restored for higher quantiles.

V. A Minimum Wage for Agricultural Workers

In Table 7 we characterize the population above and below the 60th percentile of wages according to the means of select criteria.

Table 7 here.

At their respective means both groups of labourers appear to be undernourished in terms of both calories and protein. There are substantial differences between the two groups in terms of assets – semi pucca or pucca[10] homes, livestock and bicycles. There are also differences between the two groups in terms of social status – religious and caste background - and gender of household head.

We argue in this paper that if there has to be a minimum wage in agriculture[11] it should be such that it guarantees that workers working at this wage should not fall into the PNT. Hence we argue that the wage rate at the 60th percentile should be set as the minimum wage rate. For 1994, the year of the

[8] STATA's "sqreg" command was used for the estimation of the quantile model. We tested (separately) for inequality of coefficients on enepchat and enepchat2 across different quantiles and found these to be different. These results are not reported here for lack of space.
[9] The coefficient of enepchat2 is negative and significant indicating that there are diminishing returns to increases in calorie intake.
[10] A pucca home is one made of bricks and cement.
[11] More precisely in harvesting.

sample with which we are working, this amounts to Rs. 30 per day. Scaled up by the Consumer Price Index for Agricultural Labourers (CPIAL) this amounts to Rs. 45.41 per day in 2003-04. This is just below US$1 per day at the then prevailing exchange rate. A number of commentators (e.g. Lipton 2001) have emphasized the need for setting agricultural wages high enough to avoid nutrition-poverty trap type situations.

An alternative to setting the minimum wage would be to provide subsidies to labour employment in the tradition of the efficiency wage literature. Further, following the poverty-nutrition trap model sketched in section II, diversification of non-farm opportunities, land redistribution and a more effective public distribution system targeted towards the poor would also enable large sections of the poor to break out of the trap.

VI. Conclusions

The possibility that when workers are acutely under-nourished they may not be able to exert sufficient effort so that their wages remain low which then leads to further poor nutritional outcomes has been known in the literature for almost fifty years now. A number of authors have tried to empirically test for this existence of this trap but none has been able to establish unambiguously that this holds for a subset of the working population and not the whole.

This paper has attempted to quantify and formally test for the presence of PNT in rural India. It outlines a methodology that can identify the impact of energy consumption on wage rates, even in the presence of mutual endogeneity of the two and identify the segment of the population for which the PNT holds. It identifies all workers earning upto the 60th percentile of the wage rate as the group of workers for whom the PNT hypothesis holds.

This paper has an important policy implication in that it argues that if a minimum wage has to be set in agriculture it must be adequate to ensure that workers are not caught in the poverty-nutrition trap. In the case of this sample this is equal to the 60^{th} percentile wage, i.e., Rs. 30 per day. Scaled up by the CPIAL this comes out to be Rs. 45.41per day in 2003-04, which is just below the poverty line for one day's expenditure. The paper also characterizes differences in the sample of agricultural labourers in the segments of the population for which the PNT holds and does not hold and identifies asset ownership and social background as principal attributes of these differences.

References

Behrman J. and A. Deolalikar (1987) “Will Developing Country Nutrition Improve with Income? A Case Study of Rural South India” *Journal of Political Economy,* vol.95, no.1, pp.108-138.

Binswanger, J. and M. Rosenzweig (1984) “Contractual Arrangements, Employment and Wages in Rural Labour Markets: A Critical Review” in Binswanger, J. and M. Rosenzweig (eds.) *Contractual Arrangements, Employment and Wages in Rural Labour Markets in Asia,* New Haven, Conn.: Yale University Press.

Bliss, C. and N. Stern (1978a) “Productivity, Wages and Nutrition: Part I: The Theory” *Journal of Development Economics*, vol.5, no.2, pp. 331-362.

Bliss, C. and N. Stern (1978b) “Productivity, Wages and Nutrition: Part II: Some Observations” *Journal of Development Economics*, vol.5, no.2, pp. 363-398.

Buchinksy, M. (1998) “Recent Advances in Quantile Regression Models – A Practical Guide for Empirical Research” *The Journal of Human Resources,* Vol.33, no.1, pp. 88-126.

Dasgupta, P. and D. Ray (1986) “Inequality as a Determinant of Malnutrition and Unemployment: Theory” *Economic Journal*, vol.96, no.4, pp. 1011-1034.

Dasgupta, P. and D. Ray (1987) “Inequality as a Determinant of Malnutrition and Unemployment: Policy” *Economic Journal*, vol.97, no. 1, pp. 177-188.

Dasgupta, P. (1993) *An Inquiry into Well-Being and Destitution,* Oxford: Oxford University Press.

Deolalikar, A. (1988) “Nutrition and Labour Productivity in Agriculture: Estimates for Rural South India” *Review of Economics and Statistics*, vol.70, No.3, pp. 406-413.

Gopalan, C., Sastri, B. and S. Balasubramanian (1971), *Nutritive Values of Indian Foods,* Hyderabad: National Institute of Nutrition, Indian Council of Medical Research.

Leibenstein, H. (1957) *Economic Backwardness and Economic Growth: Studies In the Theory of Economic Development*, New York: Wiley and Sons.

Lipton, M. (2001) *Successes in Anti-Poverty*, ILO Geneva, Discussion Paper

Mazumdar, D. (1959) “The Marginal Productivity Theory of Wages and Unemployment” *Review of Economic Studies*, vol.26, no.3, pp.190-197.

Mirrlees, J. (1975) “A Pure Theory of Underdeveloped Economies” in L. Reynolds (ed.) *Agriculture in Development Theory*, New Haven: Yale University Press, pp. 84-108.

Ray, D. (2004) *Development Economics*, New Delhi: Oxford University Press.

Srinivasan, T. N. (1994) “Destitution: A Discourse”, *Journal of Economic Literature*, vol. 32, no.4, pp. 1842- 1855.

Stiglitz, J. (1976) “ The Efficiency Wage Hypothesis, Surplus Labour and the Distribution of Income in LDCs”, *Oxford Economic Papers*, vol. 28, no.2, pp. 185-207.

Strauss, J. and D. Thomas (1998) “Health, Nutrition and Economic Development” *Journal of Economic Literature*, vol.36, no.3, pp.766-817.

Strauss, J. (1986) “Does Better Nutrition raise Farm Productivity?” *Journal Of Political Economy*, vol. 94, no.2, pp.297-320.

Subramaniam, S. and A. Deaton (1996) “The Demand for Food and Calories”, *Journal of Political Economy*, vol. 104, no. 1.

Thomas, D. and J. Strauss (1997) “Health and Wages: Evidence on Men and Women in Urban Brazil” *Journal of Econometrics*, vol. 77, no.1, pp.159-185.

Yu, K., Lu. Z. and J. Stander (2003) “Quantile Regression: Applications and Current Research Areas” *The Statistician*, vol. 52, no.3, pp. 331-350.

Yu, K. and M. Jones (1998) “Local Linear Regression Quantile Estimation” *Journal of American Statistical Association*, vol.93, no.2, pp. 228-238.

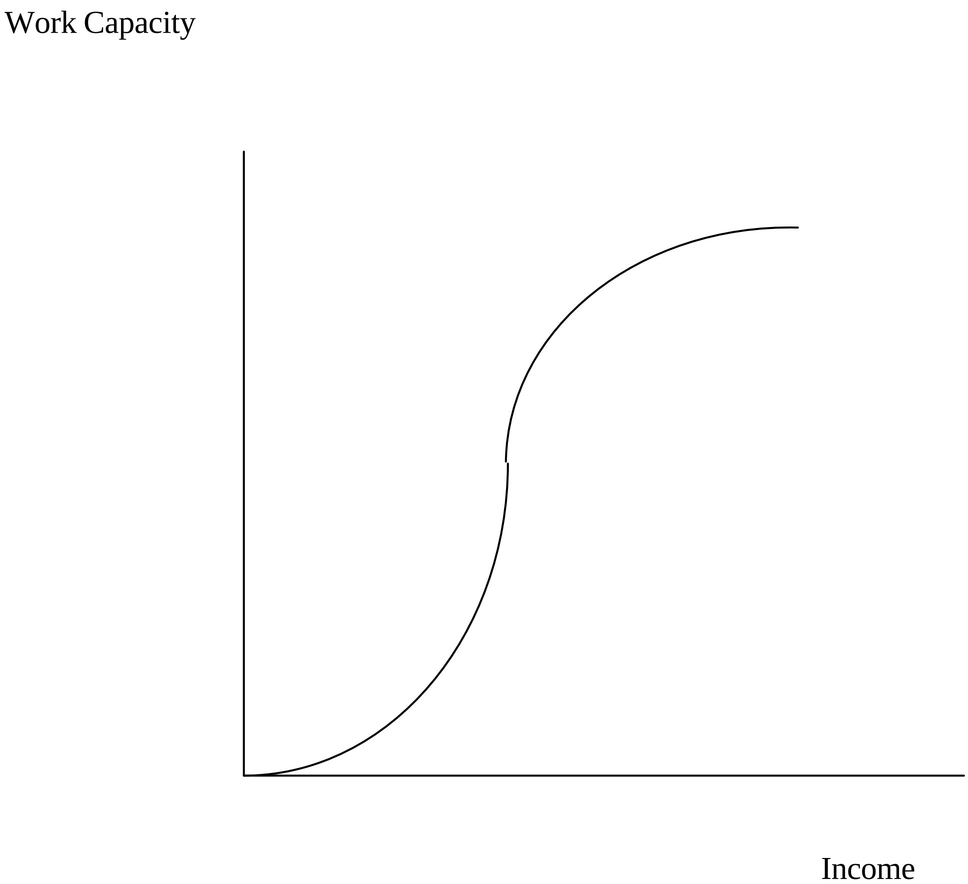

Figure 1: The Capacity Curve

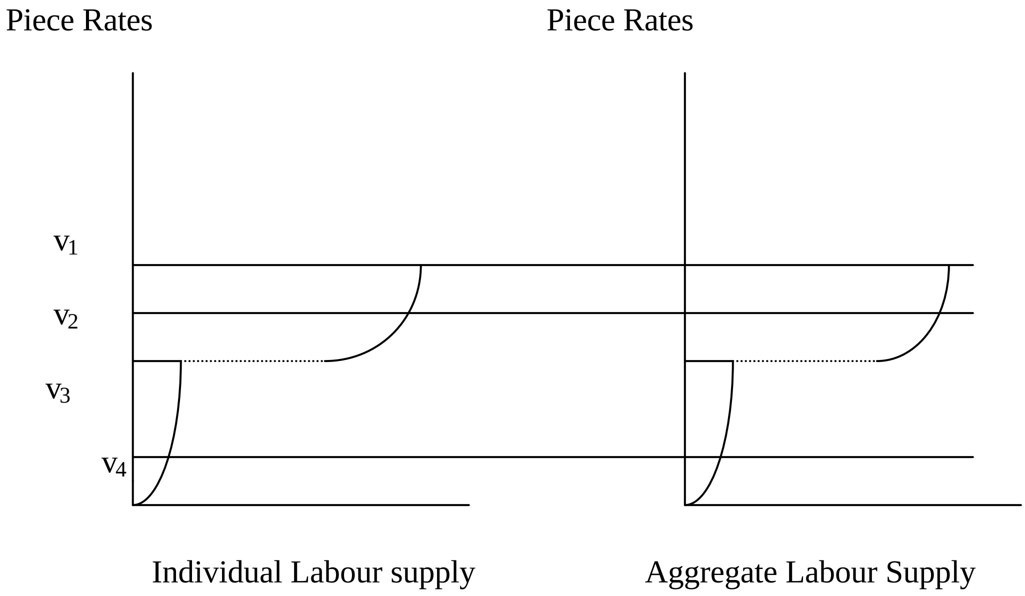

Figure 2: Individual and Aggregate Labour Supply

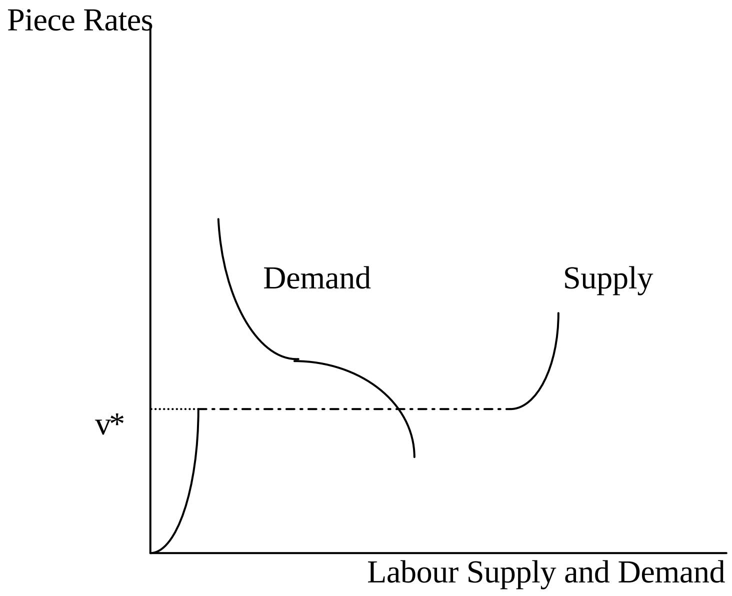

Figure 3: “Equilibrium” in the Labour Market

Source: Ray (2004).

Table 1: Independent Variables in Multi-logit model

Household Level Variables	
Variable Name	Variable Description
headage	Age of Household Head
headage2	Square of Age of Household Head
amale	no. of adult males in HH
afemale	no. of adult females in HH
hhgrp	HH Group Dummy Variable 1 if SC/ST HH and 0 Otherwise
_relreligi~1	Religion dummy: 1=Hindu, 2=Muslim, 3= Christian, 4=Sikh, 5= Buddhist, 6=Tribal 7=Jain 8=Others
_sexheadse~2	Gender of HH head 1= male 2=female
Land_own	Land Owned in Acres
Land_own2	Square of Land Owned
Other Variables	
Rain3_index	Rainfall Index (actual - normal rain fall) for agroclimatic zones.
bimaru	Dummy for Bimaru states (Bihar, Madhya Pradesh, Rajasthan, Uttar Pradesh)
Non-coastal	Dummy for non coastal districts

coastal	Dummy for Coastal districts

Table 2: Independent Variables in the First Stage of IV regression[12]

Variable	Description
nonlabincome	Non Labour Income
p1	Predicted probability for employed =1 agriculture[13]
p2	Predicted probability for employed =2 non agriculture
_edfem_edu_2	Female Education 2= Primary, 3= Middle, 4 = Matric and 1 Base case = Below primary
price_pulses	Price of Pulses
hhsize	Household size. The size of the household to ascertain per capita magnitudes is measured in adult male equivalent terms.
price_gur_~r	Price of Gur
price_edib~l	Price of Edible Oil
price_milk	Price of Milk

[12] Other food prices were included but were dropped by STATA because of collinearity problems.

[13] P4 was used as default category. STATA dropped all probabilities except

Table 3: Results on Multinomial Logit

Multinomial logistic regression	Number of obs =3621	
Log pseudo-likelihood = -2745.4961	Pseudo R2=0.244	
employed	Coef.	P>z
p1		
headage	-0.36936	0
headage2	0.003712	0
amale	0.681601	0
afemale	0.080296	0.613
hhgrp	-0.45149	0.013
_relreligi~1	-12.7557	0.58
_relreligi~2	-11.7256	0.59
_relreligi~3	7.18113	0.32
_relreligi~4	23.95511	0.43
_relreligi~6	22.24652	0.36
Land_own	0.245913	0
Land_own2	-0.00072	0
Rain3_index	0.00274	0
bimaru	2.03336	0
coastal	-12.5835	0.37
noncoastal	-14.1632	0.42
_sexheadse~2	21.9722	0
_cons	35.95229	0.45
p2		
headage	-0.21744	0.002
headage2	0.002011	0.01
amale	0.584373	0
afemale	-0.21152	0.172
hhgrp	-0.0533	0.771
_relreligi~1	4.663138	0.54
_relreligi~2	5.657916	0.51
_relreligi~3	26.35033	0.17
_relreligi~4	-6.2605	0.21
_relreligi~6	0.192044	0.18
Land_own	0.123287	0.025
Land_own2	-0.00037	0.022
Rain3_index	0.001882	0.001
bimaru	-1.04161	0.113
coastal	-15.3586	0.24
noncoastal	-12.3967	0.45
_sexheadse~2	19.75788	0.63
_cons	14.03695	0.73
p4		
headage	-0.44792	0
headage2	0.005756	0

amale	0.751971	0
afemale	0.499874	0.006
hhgrp	-0.03629	0.884
_relreligi~1	-13.9835	0.81
_relreligi~2	-12.9693	0.56
_relreligi~3	4.558774	0.51
_relreligi~4	-19.591	0.73
_relreligi~6	-20.1372	0.46
Land_own	0.19227	0.001
Land_own2	-0.00059	0
Rain3_index	0.002947	0
bimaru	2.076297	0.004
coastal	-14.0196	0.36
noncoastal	-14.5635	0.12
_sexheadse~2	25.72411	0
_cons	32.47181	0

Table 4: First Stage IV regression

First Stage Regression

IV regression With robust standard Errors.

Source	SS	df	MS
Model	5.04E+08	20	25191483
Residual	1.90E+09	2751	689258.5
Total	2.40E+09	2771	866106.1

Number of obs	=	2772		
F(20, 2751)	=	36.55		
Prob > F	=	0		
R-squared	=	0.2099		
Adj R-squared	=	0.2042		
Root MSE	=	830.22	t	P>t
_edfem_edu_2	51.7318	45.59639	1.13	0.257
_edfem_edu_3	345.7321	72.45965	4.77	0
_edfem_edu_4	267.0127	88.64169	3.01	0.003
headage	-1.42934	7.576323	-0.19	0.85
headage2	0.377545	0.085287	4.43	0
amale	-163.319	19.83593	-8.23	0
afemale	-259.425	28.9522	-8.96	0
hhsize	-20.1058	10.99291	-1.83	0.068
hhgrp	98.42552	34.49993	2.85	0.004
_relreligi~1	-923.16	342.257	-2.7	0.007
_relreligi~2	-1019.35	351.949	-2.9	0.004
_relreligi~3	-747.13	357.6389	-2.09	0.037
_relreligi~4	(dropped)			
_relreligi~6	-398.903	750.761	-0.53	0.595
p1	773.1989	140.8782	5.49	0
p2	1152.235	167.9557	6.86	0
Price_pulses	-5.23984	1.773365	-2.95	0.003
Price_gur_~r	24.67676	6.266519	3.94	0
Price_edib~l	-4.51442	0.935782	-4.82	0
Price_milk	7.137409	4.022952	1.77	0.076
nonlabincome	-0.00152	0.001272	-1.19	0.232
_cons	2418.369	391.3151	6.18	0

Table 5: Second Stage IV regression

Second Stage Regression

IV	(2SLS)	regression	with	robust
Number of obs	=	2772		
F(16, 2754)	=	.		
Prob > F	=	0		
R-squared	=	.		
Root MSE	=	0.69381		

lmale_harv~t	Coef.	Robust Std. Err.	t	P>t
Enepchat	9.02E-05	0.00136	0.07	0.947
Enepchat2	-1.26E-07	2.17E-07	-0.58	0.561
_edfem_edu_2	0.029591	0.049287	0.6	0.548
_edfem_edu_3	0.136639	0.213737	0.64	0.523
_edfem_edu_4	0.054721	0.130527	0.42	0.675
headage	-0.02641	0.010501	-2.51	0.012
headage2	0.000579	0.000133	4.34	0
amale	-0.11899	0.050487	-2.36	0.018
afemale	-0.20266	0.054565	-3.71	0
hhsize	0.000833	0.015001	0.06	0.956
hhgrp	0.097997	0.08237	1.19	0.234
_relreligi~1	-0.81903	0.763543	-1.07	0.284
_relreligi~2	-1.0396	0.783365	-1.33	0.185
_relreligi~3	-0.67669	0.820546	-0.82	0.41
_relreligi~4	(dropped)			
_relreligi~6	-0.72079	0.954871	-0.75	0.45
p1	0.812421	0.349326	2.33	0.02
p2	1.78727	0.309855	5.77	0
_cons	4.315495	1.917396	2.25	0.024

Table 6: Results on the Wage equation for the 60th Quantile

60th Quantile

Simultaneous quantile regression		Number of obs	=	2779
bootstrap(20) SEs		.60 Pseudo R2	=	0.1429
		Bootstrap		
lmale_harv~t	Coef.	Std. Err.	t	P>t
q60				
enepchat	0.002187	0.00048	4.55	0
enepchat2	-2.41E-07	6.67E-08	-3.62	0
price_pulses	0.004561	0.001842	2.48	0.013
price_gur_~r	-0.01299	0.00199	-6.53	0
price_edib~l	0.003042	0.000354	8.59	0
price_milk	0.004915	0.005469	0.9	0.369
nonlabincome	5.90E-06	2.67E-06	2.21	0.027
headage	-0.02748	0.003591	-7.65	0
headage2	0.000131	2.88E-05	4.54	0
amale	0.028787	0.02036	1.41	0.158
afemale	0.267412	0.024269	11.02	0
hhsize	0.008721	0.003525	2.47	0.013
hhgrp	-0.09177	0.020146	-4.56	0
_relreligi~1	0.902046	0.078127	11.55	0
_relreligi~2	0.74667	0.078443	9.52	0
_relreligi~3	0.28671	0.126857	2.26	0.024
p1	-0.59612	0.090752	-6.57	0
_edfem_edu_2	-0.03871	0.010575	-3.66	0
_edfem_edu_3	-0.48151	0.086716	-5.55	0
_edfem_edu_4	-0.17384	0.05794	-3	0.003
_cons	0.016206	0.754912	0.02	0.983

Table 7: Comparison of Households above and below 60th Quartile of wage

Variables	Above 60th Quantile	Below 60th Quantile
No of Households	3075	1565
HH owning Property at any other place (%)	9.07	4.08
HH owning Bicycle (%)	78.37	32.58
SC/ST HH (%)	59.34	52.58
Minorities (%)	5.78	12.07
House Owners (%)	99.8	96.54
HH owning Livestock (%)	83.05	54.12
HH Size	5.63	5.25
Mean Calorie Consumption (cal/day)	2292.94	2209.8
Mean Protein Consumption (Grams/ Day)	67.01	55.84
Mean Per Capita Annual Income (Rs.)	3659	3022
Proportion of HH where head illiterate (%)	68.61	69.9
Proportion of Female headed HH (%)	2.08	4.98
Proportion of HH in Kucchha House (%)	26.24	63.13

Appendix : Regression Results for Select Quantiles

20th Quantile

Simultaneous quantile regression		Number of obs =		2779
bootstrap(20) SEs		.20 Pseudo R2 =		0.2827

lmale_harv~t	Coef.	Bootstrap Std. Err.	t	P>t
q20				
enepchat	0.001994	0.000798	2.5	0.012
enepchat2	-1.85E-07	1.21E-07	-1.53	0.126
price_pulses	0.003494	0.001784	1.96	0.05
price_gur_~r	-0.03787	0.007795	-4.86	0
price_edib~l	0.007739	0.000831	9.32	0
price_milk	-0.00045	0.004421	-0.1	0.919
nonlabincome	1.96E-06	1.16E-06	1.69	0.091
Headage	-0.04064	0.005669	-7.17	0
Headage2	0.000173	4.33E-05	3.99	0
Amale	0.107819	0.034348	3.14	0.002
Afemale	0.145179	0.034505	4.21	0
Hhsize	0.064481	0.005593	11.53	0
Hhgrp	-0.07731	0.029581	-2.61	0.009
_relreligi~1	0.621409	0.211347	2.94	0.003
_relreligi~2	0.478833	0.235292	2.04	0.042
_relreligi~3	0.360898	0.187255	1.93	0.054
p1	-0.45992	0.085542	-5.38	0
_edfem_edu_2	0.049496	0.032694	1.51	0.13
_edfem_edu_3	-0.34185	0.134813	-2.54	0.011
_edfem_edu_4	-0.27752	0.110859	-2.5	0.012
_cons	0.253572	1.408446	0.18	0.857

30th Quantile

Simultaneous quantile regression		Number of obs =		2779
bootstrap(20) SEs		.30 Pseudo R2 =		0.2656

lmale_harv~t	Coef.	Bootstrap Std. Err.	t	P>t
q30				
enepchat	0.002326	0.000852	2.73	0.006
enepchat2	-2.24E-07	1.29E-07	-1.74	0.082
price_pulses	0.004899	0.00176	2.78	0.005
price_gur_~r	-0.03423	0.006401	-5.35	0
price_edib~l	0.00645	0.000363	17.78	0

price_milk	0.002209	0.005298	0.42	0.677
nonlabincome	4.35E-06	1.84E-06	2.36	0.018
Headage	-0.0387	0.005679	-6.82	0
headage2	0.000147	4.22E-05	3.48	0.001
Amale	0.087914	0.018781	4.68	0
Afemale	0.247316	0.026465	9.35	0
Hhsize	0.056688	0.008479	6.69	0
Hhgrp	-0.13362	0.035786	-3.73	0
_relreligi~1	0.733863	0.14656	5.01	0
_relreligi~2	0.649886	0.198595	3.27	0.001
_relreligi~3	0.266019	0.166431	1.6	0.11
p1	-0.61161	0.114142	-5.36	0
_edfem_edu_2	-0.00892	0.019407	-0.46	0.646
_edfem_edu_3	-0.50639	0.14503	-3.49	0
_edfem_edu_4	-0.4082	0.122474	-3.33	0.001
_cons	-0.27633	1.451238	-0.19	0.849

40th Quantile

Simultaneous quantile regression	Number of obs =		2779
bootstrap(20) SEs	.40 Pseudo R2 =		0.2272

lmale_harv~t	Coef.	Bootstrap Std. Err.	t	P>t
q40				
enepchat	0.002941	0.001002	2.93	0.003
enepchat2	-2.92E-07	1.41E-07	-2.08	0.038
price_pulses	0.008105	0.002543	3.19	0.001
price_gur_~r	-0.04187	0.005943	-7.04	0
price_edib~l	0.005618	0.000635	8.84	0
price_milk	0.00396	0.005508	0.72	0.472
nonlabincome	6.91E-06	1.06E-06	6.53	0
Headage	-0.03502	0.00364	-9.62	0
headage2	7.85E-05	2.39E-05	3.29	0.001
Amale	0.091388	0.020511	4.46	0
Afemale	0.384064	0.03633	10.57	0
Hhsize	0.024795	0.006502	3.81	0
Hhgrp	-0.06584	0.035798	-1.84	0.066
_relreligi~1	0.937246	0.231192	4.05	0
_relreligi~2	0.954733	0.247488	3.86	0
_relreligi~3	0.191276	0.183	1.05	0.296
p1	-0.7564	0.128735	-5.88	0
_edfem_edu_2	0.01637	0.010427	1.57	0.117
_edfem_edu_3	-0.56988	0.18182	-3.13	0.002
_edfem_edu_4	-0.37042	0.130113	-2.85	0.004
_cons	-1.32407	1.847921	-0.72	0.474

70th Quantile

Simultaneous quantile regression		Number of obs =		2779
bootstrap(20) SEs		.70 Pseudo R2 =		0.0823

lmale_harv~t	Coef.	Bootstrap Std. Err.	t	P>t
q70				
enepchat	0.001554	0.001252	1.24	0.215
enepchat2	-1.68E-07	1.84E-07	-0.91	0.361
price_pulses	0.001503	0.002043	0.74	0.462
price_gur_~r	-0.01001	0.001329	-7.53	0
price_edib~l	0.002079	0.000335	6.2	0
price_milk	0.002573	0.005557	0.46	0.643
nonlabincome	4.16E-06	1.10E-06	3.78	0
headage	-0.01511	0.003469	-4.35	0
headage2	4.31E-05	1.61E-05	2.67	0.008
amale	0.044444	0.021108	2.11	0.035
afemale	0.189157	0.030589	6.18	0
Hhsize	0.003877	0.002914	1.33	0.183
Hhgrp	-0.05936	0.044002	-1.35	0.177
_relreligi~1	0.86705	0.214333	4.05	0
_relreligi~2	0.702935	0.213041	3.3	0.001
_relreligi~3	0.434611	0.246505	1.76	0.078
p1	-0.37655	0.127887	-2.94	0.003
_edfem_edu_2	-0.02098	0.004982	-4.21	0
_edfem_edu_3	-0.42339	0.19929	-2.12	0.034
_edfem_edu_4	-0.17597	0.082244	-2.14	0.032
_cons	0.822792	2.004651	0.41	0.682