Department of Agricultural and Resource Economics
University of California Davis

How Elastic is Calorie Demand? Parametric, Nonparametric, and Semiparametric Results for Urban Papua New Guinea

by

John Gibson and Scott Rozelle

Working Paper No. 00-022

May, 2000

California Agricultural Experiment Station
Giannini Foundation for Agricultural Economics

May 2000

How Elastic is Calorie Demand?
Parametric, Nonparametric, and Semiparametric Results for Urban Papua New Guinea

John Gibson†
University of Waikato

Scott Rozelle
University of California, Davis

Abstract

This paper seeks further evidence on the elasticity of calorie demand with respect to household resources. The case presented is for urban areas of Papua New Guinea, where just over one-half of the population appear to obtain less than the recommended amount of dietary energy. The relationship between per capita calorie consumption and per capita expenditure in urban areas of Papua New Guinea is not consistent with the view that income changes have negligible effects on nutrient intakes. The unconditional calorie demand elasticity is approximately 0.6 for the poorest half of the population, most of whom have less than the recommended 2000 calories per day available to them. Using parametric and semiparametric estimation to control for a wide range of other influences on calorie consumption does not materially reduce the size of the elasticity. Therefore, these results are not supportive of "growth-pessimism" and instead suggest that policies that increase urban household incomes will also act to reduce undernutrition.

JEL: I12, O15
Keywords: Income, Nutrition, Nonparametric, Semiparametric, Papua New Guinea

† Department of Economics, University of Waikato, Private Bag 3105, Hamilton, New Zealand. Fax: (64-7) 838-4331. E-mail: jkgibson@waikato.ac.nz.

How Elastic is Calorie Demand?
Parametric, Nonparametric, and Semiparametric Results for Urban Papua New Guinea

Inadequate nutrition is perhaps the most important problem facing the poor. Being hungry lowers productivity, hinders learning, and increases the risk of disease. All of these factors conspire to help the poor stay poor. Therefore, a key question for any policy aiming to improve human development is whether it improves nutrition? The orthodox view in development economics has been that policies which increase the incomes of the poor have beneficial effects on nutrition. However, in recent years a new literature has emerged suggesting that increases in income will *not* result in substantial improvements in nutrient intakes (Behrman and Deolalikar, 1987). Although the poor may increase their food expenditures as incomes rise, this extra spending goes on food attributes other than nutrients, for example, taste, appearance, variety, or status (Behrman, Deolalikar and Wolfe, 1988). Moreover, other factors, such as women's schooling, are claimed to be more important than incomes in determining nutrient demands (Behrman and Wolfe, 1984).

The key behavioural parameter in these studies is the elasticity of calorie demand with respect to household resources. Estimates of this elasticity range from 0.01 to 1.18 (Strauss and Thomas, 1995). The wide range is partly explained by whether the estimates come directly from calorie demand equations, or are indirect conversions from food demand equations. The estimates also are affected by whether the calories measured are those actually consumed (intake) or just those available to the household, and whether household resources are measured by income or by expenditure. The recent literature on calorie demand elasticities claims that the high elasticities that were previously believed to exist are likely to be the result of either bad data or poorly designed estimation strategies (Behrman and Deolalikar, 1987; Bouis, 1994).

Important implications follow from the belief that changes in income do not result in substantial changes in levels of nutrition. It increases “growth-pessimism” because of the wedge that is driven between affluence and an obvious component of human development – having access to an adequate level of nutrients. Believing that income changes have only small effects on nutrient consumption may also induce a sense of indifference about the nutritional effects of short-term income shocks (e.g., following a structural adjustment) because of the assumption that affected households will just downgrade the quality, rather than the quantity, of their diets. The new evidence on calorie demand elasticities also sits uneasily with Sen’s entitlement approach to famines; if calorie intakes of the poor are income unresponsive one would not expect entitlement failures to result in mass starvation (Ravallion, 1990).

Given the important policy implications of the recent views on calorie demand, it is worth seeking further evidence on the elasticity of calorie demand with respect to household resources. In this paper, the case presented is for urban areas of Papua New Guinea. This is a setting with serious nutritional problems; almost one-half of the population obtain less than the recommended amount of calories, with the poorest decile of households getting only 54 percent of the requirement (Gibson, 1995). Papua New Guinea is also interesting because the economy has been subjected to a number of recent shocks, including droughts, structural adjustment, and a fall in the value of the currency by two-thirds since 1994. The resulting falls in real incomes have greatly affected the urban population in many ways, and it is reasonable to be concerned that nutrition has suffered since urban residents are highly dependent on imports of cereals and other foods.

In addition to policy relevance for our case study country, there are other benefits of studying calorie demand in urban areas. First, the number of undernourished people living in cities will soon exceed the number living in the countryside (Ruel and Garret, 1999), but most studies of calorie-income relationships have been restricted to rural areas.[1] Another benefit is that measurement error problems may be less important in the cities because diets are based mainly on purchases rather than on self-produced items. Problems in measuring quantities and values of self-produced foods may create errors on both sides of calorie demand equations (own-food production is a component of income), causing an upward bias in the estimated effect of income on calories (Bouis and Haddad, 1992). Measurement error should be even less of a problem in our study because all members of respondent households kept detailed expenditure diaries and the enumerators also paid careful attention to transfers of food between households.

The next section of the paper examines the previous literature on the relationship between income and calories. After describing the household survey data, we explain our various methods of estimating the relationship between calories and income. The main findings are then presented and the last section offers some conclusions.

Income and Calories: Previous Evidence

Strauss and Thomas (1995) summarise a wide ranging set of studies of income and expenditure elasticities of calorie demand and have identified four themes that emerge from the patterns of the elasticity estimates. First, the largest estimates of nutritional responses to income changes come from indirect calculations of the calorie elasticity, based on (calorie) weighted averages of expenditure elasticities for broad food groups. One problem with this method is that it assumes a constant conversion factor between expenditure on a food group and the quantity of calories obtained. If there is "quality-shading," where richer households buy more expensive, "tastier," and "more convenient" calories within the broad food groups, the elasticity of calorie *quantities* with respect to income will be overstated. For example, Behrman and Deolalikar (1987) estimate an indirect elasticity of 0.77, which is 4.5 times higher than the directly estimated calorie elasticity (0.17).

The second theme is that the calorie elasticity is lower when household resources are measured by income rather than by expenditure. Measurement error bias is the likely cause because current incomes are more volatile than current expenditure, making them a more noisy measure of permanent income (Bhalotra and Attfield, 1998).

The third theme that emerges concerns the way that calories are measured. The data commonly available to economists refer to calorie *availability*. These data, however, are derived from the same household budgets that are used to produce estimates of total household expenditure, which appear on the right hand side of calorie demand equations. Having the dependent and an independent variable created from the same data creates the possibility of common errors. Unlike the usual case where measurement error causes bias of regression coefficients towards zero, however, in this case the correlated errors may increase the size of estimated elasticities (Bouis and Haddad, 1992). For example, a household may understate its food expenditures so the calculated calories will also be understated and it will appear that low levels of expenditure (as a proxy for income) correlate closely will low calorie consumption.

A further problem with calorie availability data is that they may not adequately control for *wastage* and *leakages*. Calorie consumption is overstated for households that give away or waste relatively more food or have relatively more visitors at meals, while it is understated for households that are absent from many meals or receive food gifts. Because households in the first group are likely to be rich and those in the second group poor, uncorrected leakages will cause the calorie elasticity to be overstated (Bouis, 1994).

The fourth theme that emerges is that elasticities measured at a single evaluation point, such as the mean per capita income or expenditure level, may understate the elasticity that applies to much poorer households. This would occur if the calorie elasticity is subject to non-linearities (Ravallion, 1990). Hence, elasticities reported at mean expenditures may not be relevant for the poorly nourished, the group that is of most concern to public policy.

The strategy for this paper is to undertake a careful set of empirical exercises that look at calorie demand. In the first part of our paper we will focus on producing estimates of calorie demand elasticities for different income groups (addressing the issues of non-linearities). We also want to see if these calorie demand elasticities are robust to different estimation techniques that address some of the main statistical issues raised by Strauss and Thomas (1995) and Subramanian and Deaton (1996), examining the effect on estimates of calorie-expenditure elasticities when (a) the error term of the calorie demand equation is correlated with total expenditures; (b) when the effects of calorie shading are accounted for. Finally, our data set allows us to avoid the typical measurement problem that arises with calorie availability data because our estimates are adjusted to account for gifts of food in and out of the household and differing numbers of household members participating in meals during the survey period.

Data

The empirical work in this paper uses data from the Papua New Guinea Urban Household Survey, carried out in six provinces in 1985-87. Over 1300 households were included in this survey but the available sample is smaller because of households with missing expenditure information (*n*=84) and the removal from the sample of non-private dwellings (*n*=149). After deleting a further 61 households because the roster of people present at each meal was not completed, a sample of 1033 households was left.

An important feature of the survey is that data were collected with income and expenditure diaries, rather than by the possibly less accurate recall method. Diaries were completed by all adults (and included questions on expenditures made by children) for a 14 day period (the customary pay cycle), with interviewers normally making *daily* checks on each household. In addition to food purchases and other recurring expenses, the diaries also recorded details on own-production and informal sector sales of food. The survey paid close attention to food stocks and transfers of food among households, listing each transaction with people who were not part of the household.

The diaries also provide detailed enumeration of each household's food purchases, gifts, and own-family garden production for over 200 separate food items. The data used in this study were aggregated into 35 food groups. The aggregation, however, is unlikely to cause serious bias from unmeasured intra-group substitution (i.e., the problem raised by Behrman and Deolalikar, 1987) because it mainly affected fresh fruits and vegetables, which are not important sources of calories in the sample. The cereal and root crop staples (rice, flour, sweet potato, taro, cassava, bananas and sago), and the main sources of dietary fats (coconuts, pork, dripping) remained as separate food items in the group of 35 foods.[2]

One important feature of the survey is that it included a roster of the number of adults and children present at each meal during the diary-keeping period. These data complement the details on food gifts recorded in the diaries and allow an adjusted measure of per capita daily calorie availability to be computed. The unadjusted estimate of per capita daily calorie availability is divided by the ratio of the actual number of person-meals to the number of person-meals that would be expected given the household size. For example, residents in a household where 10 percent more person-meals were consumed than would be expected have an adjusted calorie availability that is only 90.9 percent (= 1/1.1) as high as the unadjusted calorie availability. This adjustment assumes that meals eaten by visitors (or by absent residents dining at other houses) have the same calorie content as meals eaten by residents. With this adjustment, the diary and roster data cover most leakages of calories from rich households and imports of calories into poor households, with only food wasted and food scraps fed to pets not covered. Adjusting the calorie availability estimates in this way increases the apparent calorie consumption of the poorest quartile of households by three percent and reduces the apparent consumption of the richest quartile by seven percent (Table 1).

Two other features of the sample are apparent from Table 1 that suggest that this is an interesting setting in which to examine the calorie-income relationship. Although incomes are relatively high by developing country standards, they are unequally distributed. With annual average expenditures of approximately US$1300 per person at the time of the survey, the urban PNG population is not particularly poor. Hence, calorie demand elasticities might be expected to be low if there is an inverse relationship between elasticities and income levels (Behrman and Wolfe, 1984). But, the high degree of inequality (the richest quartile has twice as many calories and seven times higher expenditures than the poorest quartile) may, however, cause a wider range of elasticities than is found in samples drawn from more homogeneous environments.

Methods

Non-linearities may characterise the relationship between calories and income because the least well-nourished persons are likely to make the largest nutritional responses as their budgets shift (Ravallion, 1990). Thus, identifying the full range of nutritional responses, rather than aggregating them into a single estimate, may be important because social concern about the size of the nutritional responses may vary with the degree of malnourishment. To better understand the relationship between calorie demand and income, nonparametric regression may be an appropriate tool because it makes no assumptions about functional form, allowing the data to 'speak for themselves' (Delgado and Robinson, 1992).

Nonparametric regression estimates the function, $m(x)=E(y|x)$, by computing an estimate of the location of y within a specific band of x. If this band maintains a constant number of observations, the estimator is a "nearest neighbour" estimator while if it maintains a constant width it is a "kernel" estimator (Strauss and Thomas, 1995). We use a nearest neighbour estimator, known as LOWESS (Cleveland, 1979), because the distribution of income (measured by per capita expenditures (PCE)) is skewed even after a log transformation. Thus, a kernel estimator may not give robust results because the fixed width bands will have few observations in the upper tail. The details of the estimator are explained in Appendix 1.

Although nonparametric regression techniques help us to explore non-linearities in the relationship between calories and income, they have the disadvantage of being restricted to bivariate relationships. Ideally, we would like to discover the effect of income on calories after controlling for relevant covariates because otherwise the calorie elasticity may be biased. For example, household size may negatively affect both per capita expenditures and calories (because of scale economies and the lower calorie needs of children), so the exclusion of this relevant explanatory variable will cause an upward bias in the estimated coefficient on per capita expenditures.[3]

Previous studies have used two approaches to introducing factors other than incomes into calorie demand models. Subramanian and Deaton (1996) indirectly account for household size by splitting their sample into eight different household-size types and estimating nonparametric calorie-income regressions within each subsample. Strauss and Thomas (1995) use the nonparametric LOWESS estimator to explore the shape of the non-linearities in the calorie—income relationship and then based on these results, search for a parametric functional form (the log-inverse quadratic) that can approximate the shape that they find in their nonparametric work. The advantage of using the parametric approximation is that extra covariates can be added to the model.

In this paper we also search for a parametric specification that approximates the shape of our nonparametric calorie demand curve but we also use a new method of incorporating extra covariates into a nonparametric model – semiparametric estimation. A semiparametric estimator combines both nonparametric and parametric components:

$$y_i = z_i \beta + q(x_i) + \varepsilon_i$$

where z_i is a $1 \times p$ vector of explanatory variables of known (or asssumed) functional form and x_i is the explanatory variable of unknown functional form (Robinson, 1988). Appendix 1 also describes the estimation method in more detail.

Results

Nonlinearities in the Calorie-expenditure Relationship

Nonparametric Estimates

The first results reported are the nonparametric estimates of the regression of the logarithm of per capita calories, adjusted for meals fed to guests and meals received, on the logarithm of per capita expenditure (PCE). Figure 1 shows the locally weighted smoothed scatterplots between calories and expenditures for four different bandwidths: 100, 300, 500, and 700 observations. Each of these curves is based on the full sample of 1033 households, and the bandwidth refers to the number of observations used to form the smoothed scatterplot point for each household. Although an algorithm for the optimal choice of the smoothing parameter (and therefore number of nearest neighbours in the band) has been suggested by Cleveland (1979), the advantage of presenting several smoothed scatterplots is that it shows whether the main features of the results emerge, regardless of the level of smoothing chosen.

All four of the smoothed scatterplots in Figure 1 show that the relationship between calories and expenditures flattens somewhat at higher per capita expenditure levels. The changing slope is clearest in the scatterplot with a bandwidth of 700 observations because the finer level variation is suppressed. It appears that the slope of the regression function falls at a point near the median level of per capita expenditure, which is when predicted per capita calorie availability reaches about 2100 per day. However, the curve does not completely flatten out. In this respect, Figure 1 resembles the calorie availability – expenditure curves for Bukidnon in the Philippines presented by Strauss and Thomas (1995) and for Maharashtra presented by Subramanian and Deaton (1996). The interesting feature of the current results is that they are for urban households who are less poor than the Indian and Filipino households in the previous studies.

Figure 1 also shows standard errors for the nonparametric regression functions. These have been obtained by bootstrapping (Efron and Tibshirani, 1993). Random samples were drawn, with replacement, from the residuals of the original regression and used to generate 100 vectors of new dependent variables, conditioning on log PCE. The LOWESS regression was then re-estimated on each of these 100 new samples, giving 100 predicted values of log per capita calories for every single value of log PCE. The standard deviation over these 100 replications is used as an estimate of the standard error for each point on the nonparametric regression curve.[4] The intuition behind this bootstrapping is that if we knew the population distribution we could obtain the sampling distribution of any statistic by simulation: draw random samples (with replacement) of size N=1033, calculate the statistic and make a tally of the values that the statistic takes for each sample. But, not knowing the true population distribution, bootstrapping instead uses the observed distribution of the sample in its place.

Figure 2 shows the slope of the curve in Figure 1 (or the elasticities of per capita calorie consumption with respect to per capita expenditures) for the LOWESS regressions estimated with a bandwidth of 700 observations. For the poorest one-quarter of the population the calorie elasticity is approximately 0.60. In the second quartile the elasticity falls from 0.57 to 0.48, with a larger fall to 0.34 in the third quartile. The calorie elasticity then hovers around 0.30 for the richest one-quarter of the population. The fall in the size of the elasticity is similar to what Subramanian and Deaton (1996) find but the decline in the elasticity is less smooth, falling rapidly amongst the middle two population quartiles, which is the region where food energy requirements of approximately 2000 calories per person per day are achieved and then surpassed. The range of the elasticity in Figure 2 is not consistent with a constant elasticity relationship between calories and income. In other words, the least-squares estimated coefficient of 0.42 (standard error 0.02) from the regression of log per capita calories on log PCE would not be a very accurate summary of the relationship between calories and income across the different parts of the sample.

Our results can also demonstrate the dangers of reporting the value of the calorie-expenditure elasticity estimated by nonparametric methods at a single evaluation point (Table 2). The calorie elasticity for the household with mean log PCE is 0.42 (row 1). However, this household is richer than the median household (the underlying distribution of per capita expenditure is more positively skewed than is a true log-normal distribution) and consequently has a lower calorie elasticity (row 1 versus row 2). Evaluating the elasticity at the median household is also not a good idea because there is an inverse correlation between household size and per capita expenditure. The median household, ranked in terms of PCE, contains people at the 60th percentile of the population; the median person is found in a household at the 40th percentile of the household ranking. Hence, the elasticity evaluated at the median household (0.43) is lower than the elasticity applying to the median person (0.48, from row 3).

Parametric Estimates

Regression results from traditional functional forms using parametric methods demonstrate the difficulties in capturing the changes in the calorie elasticity as income changes. Figure 2 shows the elasticity curve when per capita calories are regressed on the inverse of log PCE and its square. This parametric specification is used by Strauss and Thomas (1995) to approximate the shape of their nonparametric calorie elasticity curve. But this parametric specification does not give a good approximation to the nonparametric results in the current data from PNG, because it misses the rapid fall in the size of the calorie elasticity in the middle quartiles and overstates the elasticity for the poor and understates it for the rich.

The closest we came to isolating the changing relationship between per capita calories and expenditures was with an income spline function. Initially, we tried four segments, each corresponding to a population quartile based on the distribution of per capita expenditures. Statistical tests, however, suggested that the first two segments could be collapsed into one (p<0.84). From the resulting model, the calorie elasticity for the first two quartiles was 0.62 (0.05); for the third quartile it was 0.41 (0.20); for the fourth quartile it was 0.24 (0.05), where standard errors are in the parentheses. A Wald test of linearity ($H_{0:}$ slope dummy variables for all quartiles equal zero) suggests significant structural differences in the calorie-expenditure relationship, with $\chi^2_{(4)}$=24.2.

Choosing the Covariates

With a suitable parametric functional form chosen (i.e., the income spline function), the next step in refining our characterisation of the income-expenditure relationship is to decide which variables to add to the model (to give us the specification of the semiparametric model—see next section below). Household size is likely to matter, as explained above. Demographic composition variables also may be important if there are differences in calorie consumption according to age and gender. The age and gender of the household head and education levels, especially for women, may affect nutrient intakes (Behrman and Wolfe, 1984), as may the ethnicity of the household head and their main economic activity. Food prices might be important to include as a way of ensuring that non-linearities are not just due to excluded price effects, because low-income consumers may have the largest nutrient response to price changes (Alderman, 1986). Cluster-level fixed effects are also candidates for inclusion because there may be community influences on eating patterns that are not captured by the individual and household level variables. Finally, although the households in the current sample are urban, many of them have food gardens and this may influence their calorie consumption.

Table 3 contains the results of regressing log calories on a log PCE spline function plus various sets of covariates (columns 1 to 4). Controlling just for household size in column (1) gives elasticities of calories with respect to per capita expenditure of 0.57 for the poorest two quartiles; 0.29 for the third quartile; and 0.13 for the richest quartile. These are somewhat smaller elasticities than when household size is excluded, with the reduction being largest at high expenditure levels. Thus, in our parametric analysis adding household size to the model increases the non-linearity in the calorie-income relationship. The results in column (2) are generated by a model that adds further covariates for characteristics of the household head, household demographic composition, education levels, and access to gardens. The added variables, however, only affect the estimates for the third quartile, where the calorie elasticity rises to 0.35. Among the new variables, the preponderance of negative signs on the four child demographic ratios confirms the fact that children consume fewer calories than adults. The age of the household head also has a positive and significant effect. The ethnicity and gender of the household head, however, has no significant effect on calorie demand. The type of income earning activity of the household head also appears to have little impact on calorie demand, unless the household grows a garden, which negatively affects calorie consumption, a result that might reflect the fact that the crops grown by PNG urban residents in their gardens are mostly vegetables, which have a lower calorie content than the cereals that non-gardening households would tend to buy.[5]

Column (3) reports the results of adding the prices of seven major foods which contribute two-thirds of available calories. Adding prices to the model causes only a small rise in the elasticity of calories with respect to per capita expenditures over all segments of the spline function. The minimal reduction in the non-linearity between calories and expenditures suggests that it is unlikely that excluded price effects are producing the non-linear relationship between calories and expenditures in Figure 2. The sign patterns of the price elasticities appear plausible, with increases in the price of cheaper sources of calories reducing calorie availability and increases in the price of more expensive sources of calories, especially that for sugar, increasing calorie availability.

Column (4) reports the results of adding 282 dummy variables, one for each cluster in the sample (and dropping prices, which do not vary within clusters). While the addition of the cluster effects does not attract a very large *F*-statistic, the calorie elasticity for the poorest half of the population falls to 0.51 and the elasticity for the richest quarter rises to 0.23. Excluded community effects appear to have exaggerated the non-linearity in Figure 2. The excluded community effects, however, are not the sole cause; the within-cluster calorie elasticity for the poorest half of the sample is still two times higher than for richest quarter of the sample. Moreover, the overall responsiveness of calorie demand to income is not affected by the presence of the cluster effects: the coefficient on log PCE in a constant elasticity version of the column (4) equation is 0.38, which is only slightly higher than the elasticity of 0.35 estimated without cluster effects. In fact, the most important result of Table 3 is that the income elasticity of calories is high in all of the equations, especially for the poor.

Semiparametric Estimates

Figure 3 shows the semiparametric estimates of the elasticity of calories with respect to per capita expenditure for the models using the covariates in columns 1, 3 and 4 of Table 3. The elasticity curve from the nonparametric estimates (from figure 2) is also presented for comparison. Most significantly, all of the semiparametric elasticity curves have the same basic shape as the nonparametric curve, although the addition of covariates results in estimated elasticities that are slightly smaller at all income levels (since the nonparametric curve in Figure 3 is higher at all points than any of the semiparametric estimates). Adding household size to the model reduces the size of the calorie elasticity by approximately 10 percent for the poorest households, and by over 30 percent for the largest households (the light solid line, Figure 3). Adding the rest of the covariates (demographics, economic activity, schooling, food prices) shifts the elasticity curve upwards slightly, especially at higher expenditure levels (compared to the semiparametric model with only household size—the dark solid line, Figure 3). Adding cluster effects (and deleting prices) to get the 'within-cluster' model (the dashed line, Figure 3) causes calorie elasticity estimates to fall to approximately 0.5 for the poorest quarter of the population. However, it should be stressed again, that while this discussion has highlighted the differences between the nonparametric and semiparametric estimates, perhaps the most significant finding is that regardless of the estimating technique, the estimates for the poor are high. In other words, calorie consumption by urbanites in PNG positively responds to income changes at all levels of income, and the poor are especially responsive.

Instrumental Variables Estimates

The results presented thus far have been based on estimators that assume zero correlation between per capita expenditure and the error term. There are two possible reasons for questioning this assumption. The first is that household incomes, and hence expenditures, could be constrained by nutrition. If so, the coefficient on the per capita expenditure variable would be biased. Realistically, however, there are reasons to doubt this relationship in the current setting because of the very low cost of purchasing the extra calories needed for labouring activity (Subramanian and Deaton, 1996). In PNG at the time of the survey, it cost just two percent of the minimum daily wage to buy the 600 calories per day (in the form of rice) needed for a person to do active work as opposed to just surviving.

Second, and perhaps most seriously, any random errors in measuring food expenditures are transmitted (by construction) both to calorie availability and total expenditures, resulting in correlated measurement errors. Bouis and Haddad (1992) suggest that for a linear model the upward bias from the correlated errors will typically outweigh the usual downward (attenuation) bias that results when an explanatory variable is measured with error. Subramanian and Deaton (1996) study this problem for a log-linear model and show that using nonfood expenditure as an instrumental variable (IV) for household total expenditure will give a lower bound to the true value of the calorie elasticity, whether or not correlated measurement error is present. The reason is that, conditional on the true value of income, a positive regression error implies that food expenditure is above its predicted value so nonfood expenditure must be below its predicted value. Hence, noise in the instrument is correlated with the regression disturbances (violating the requirements for an ideal instrument), so the IV estimates are biased downwards.

Table 4 contains OLS and IV estimates of the calorie elasticity for four different specifications of a constant-elasticity (log-log) model.[6] Two sets of IV estimates, one using nonfood expenditures as an instrument for total expenditures (as in Subramanian and Deaton, 1996) and one using household total income as an instrument are presented.[7] Household income is a valid instrument because, as discussed above, the feedback from nutrition to income is likely to be small given the trivial cost of purchasing additional calories needed for physical activity. Measurement errors in income also should be uncorrelated with errors in expenditures because the two types of data were collected in separate sections of the survey and refer to different time periods. Respondents kept expenditure diaries for the two weeks after the first interview whereas they were asked about their income for the week prior to the interview.

The elasticities reported in the first row of Table 4 suggest that the lower bound elasticity for the bivariate relationship between per capita calories and per capita total expenditure is 0.31, in the presence of correlated errors, with an upper bound of 0.42 (row 1). Both lower and upper bounds fall by about 10 percentage points when household size is introduced as an additional covariate (row 2). The smallest estimate of the lower bound is 0.18, for the within-cluster model (column 2, row 4). However, this same set of covariates also gives the highest elasticity estimate (0.52) when household income is used as the instrument. Durbin-Wu-Hausman tests suggest that the IV estimates using nonfood expenditure as the instrument (column 2) are all significantly different from the corresponding OLS estimates.[8] But with the exception of the within-cluster model, when household income is the instrument (column 3), the IV estimates are not significantly different from OLS estimates.

Food Expenditures and Calorie Quality and Indirect Estimates of the Calorie Elasticity

In this final section, we illustrate that the high expenditure elasticity of calories holds up even when calorie shading is accounted for. The fact that the price paid for calories rises with per capita expenditure in the nonparametric regression results presented in Figure 4, indicates that diets do shift towards higher quality and costlier sources of nutrients as incomes rise. The results of the nonparametric LOWESS analysis (of the log of the price of per capita calories on the log of per capita expenditures) illustrates that the richest households pay approximately three times more per calorie than their poorer counterparts.

The average elasticity of the price paid per calorie with respect to per capita expenditure is approximately 0.35. As seen by the upward sloping line in figure 4, the elasticity rises slightly for rich households. As shown in Behrman and Deolalikar (1987), the elasticities of calorie quantity and calorie price can be added together to give the elasticity of food expenditure with respect to total expenditure. This food expenditure elasticity ranges from 0.92 for the poor to 0.69 for the rich, but the composition of the elasticity in terms of quantity and quality components varies with income level. Two-thirds of the extra food expenditures made by the poor go on increasing quantities of calories; only one third go on increasing calorie quality. For the rich, the ratios are roughly opposite.

Indirect estimates of the calorie elasticity, calculated from systems of food demand equations, do not show as large an upward bias as found by Behrman and Deolalikar (1987). Estimates of the total expenditure elasticities of demand for 35 foods reported by Gibson (1996) were combined with estimates of average calorie shares to yield an indirect estimate of the expenditure elasticity of calorie demand of 0.56. This indirect estimate is 1.3 times higher than the directly estimated elasticity from the least-squares regression of log per capita calories on log per capita expenditure. Although this is a bias – due in part, presumably, to calorie shading – that investigators would like to avoid by directly estimating calorie demand equations, the scale of the bias is much smaller than the three- to four-fold error that Behrman and Deolalikar (1987) report.

One possible explanations for why our results differ so much from those of Behrman and Deolalikar (1987) is that they used much broader food groups (grains, sugar, pulses, vegetables, milk, and meat) than the ones that we used. However, we can show that this is not the reason. When we re-estimate the expenditure elasticities for a much more aggregated set of five foods (cereals; meat and fish; fruit, vegetables and nuts; root crops; and other foods), the indirectly estimated calorie elasticity rises only slightly (to 0.63). This suggests that in urban PNG most of the dietary substitution that occurs as households get richer is *between* the broad food groups (e.g., from cereals to meat) rather than *within* them.

Conclusions

The relationship between per capita calorie consumption and per capita expenditure in urban areas of Papua New Guinea is not consistent with the view that income changes have negligible effects on nutrient intakes. The unconditional calorie elasticity is approximately 0.6 for the poorest half of the population, most of whom have less than the recommended 2000 calories per day available to them. Using parametric and semiparametric estimation to control for a wide range of other influences on calorie consumption does not materially reduce the size of the elasticity. Therefor, our results are not supportive of "growth-pessimism" and instead suggest that policies that increase urban household incomes will also act to reduce under-nutrition. The results also are consistent with the idea that poorly nourished persons make larger nutritional responses to changes in income than do well nourished persons.

These results also suggest the need for serious attention to be paid to the adverse nutritional effects of real income shocks, such as the stabilisation programs and structural adjustments that Papua New Guinea is currently undergoing. The estimated elasticity of 0.6 for the poorest half of the urban population suggests that per capita calorie availability may have declined by over by over 10 percent during the period of falling real incomes since 1994. Given the already existing degree of under-nutrition, such a decline will almost certainly have serious consequences.

In terms of methodology, the current results show that considerable non-linearities in the calorie-expenditure relationship may be revealed when a parametric structure is not imposed on the data. Also, the bias from using indirect methods to calculate the elasticity of calorie demand with respect to total expenditures appears less serious in these data than in some earlier studies.

Appendix 1

Nonparametric and Semiparametric Estimating Framework

Nonparametric Framework

For each point (x_i, y_i) on the scatterplot, the LOWESS estimator forms the smoothed point $(x_i, \hat{y}_i)$ from a locally weighted regression of a first order polynomial. The weights come from a "tricube" function:

$$w_k(x_i) = \left(1 - \left(\frac{|x_k - x_i|}{\max_{\aleph(x_i)} |x_j - x_i|}\right)^3\right)^3$$

which decreases for points further away from (x_i, y_i), becoming zero at the boundary of neighbourhood $\aleph(x_i)$. A new set of weights, δ_i is then defined for each (x_i, y_i) based on the size of the residual $y_i - \hat{y}_i$. Larger residuals have smaller weights, to guard against outliers distorting the smoothed plots. New fitted values are computed as before, but with $w_k(x_i)$ replaced by $\delta_i w_k(x_i)$. The calculation of new weights and new fitted values can be repeated several times to get the robust locally weighted regression. Details can be found in Cleveland (1979).

Semiparametric Regression

The semiparametric estimator is based on the model described by Robinson (1988):

$$y_i = z_i\beta + q(x_i) + \varepsilon_i \qquad E(\varepsilon_i | z_i, x_i) = 0 \qquad (i = 1, 2, \ldots, n) \tag{1}$$

where y_i is the logarithm of per capita calories for the ith household, z_i is a $1 \times p$ vector of explanatory variables of known (or asssumed) functional form, β is a $p \times 1$ vector of regression coefficients, and x_i is a $1 \times k$ vector of explanatory variables of unknown functional form. Equation (1) can be rewritten as

$$y_i - E(y_i | x_i) = (z_i - E(z_i | x_i))\beta + \varepsilon_i \tag{2}$$

suggesting that $q(x)$ can be estimated in a three-step procedure. First, the unknown conditional means, $E(y_i | x_i)$ and $E(z_i | x_i)$ are estimated using a nonparametric estimation technique. Second, these estimates are substituted in place of the unknown functions in equation (2) and ordinary least squares is used to estimate β from

$$y_i - \hat{E}(y_i | x_i) = \left(z_i - \hat{E}(z_i | x_i)\right)\beta + \varepsilon_i^*,$$

with these estimates denoted $\hat{\beta}^*$. Noting that equation (1) can be rewritten as

$$y_i - z_i\beta = q(x_i) + \varepsilon_i \tag{3}$$

the third step is to insert the $\hat{\beta}^*$ into equation (3) so that $q(x)$ can be estimated by a nonparametric regression of $(y_i - z_i\hat{\beta}^*)$ on x. This final nonparametric regression should identify the relationship between calories and household resources, taking account of the other covariates that entered via the parametric part of the model. Examples of this approach can be found in Anglin and Gencay (1996) and Bhalotra and Attfield (1998).

References

Alderman, Harold, 1986, Effects of price and income increases on food consumption of low income consumers, *mimeo* International Food Policy Research Institute, Washington.

Anglin, Paul and Ramazan Gencay, 1996, Semiparametric estimation of a hedonic price function, *Journal of Applied Econometrics* 11(5): 633-648.

Behrman, Jere and Barbara Wolfe, 1984, More evidence on nutrition demand: income seems overrated and women's schooling underemphasized, *Journal of Development Economics* 14(1): 105-128.

Behrman, Jere and Anil Deolalikar, 1987, Will developing country nutrition improve with income? A case study for rural South India, *Journal of Political Economy* 95(3): 492-507.

Bhalotra, Sonia, and Cliff Attfield, 1998, Intrahousehold resource allocation in rural Pakistan: a semiparametric analysis, *Journal of Applied Econometrics* 13(4): 463-480.

Bouis, Howarth, 1994, The effect of income on demand for food in poor countries: are our food consumption databases giving us reliable estimates?, *Journal of Development Economics* 44(2): 199-226.

Bouis, Howarth, and Lawrence Haddad, 1992, Are estimates of calorie income elasticities too high?: A recalibration of the plausible range, *Journal of Development Economics* 39(4): 333-364.

Cleveland, William, 1979, Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association* 74(368): 829-836.

Davidson, R., and MacKinnon, J. (1993). *Estimation and Inference in Econometrics*. (Oxford University Press, New York).

Delgado, Miguel and Peter Robinson, 1992, Nonparametric and semiparametric methods for economic research, *Journal of Economic Surveys* 6(3): 201-249.

Efron, Bradley and Robert Tibshirani, 1993, *An introduction to the bootstrap* (Chapman and Hall, New York).

Gibson, John, 1995, *Food Consumption and Food Policy in Urban Papua New Guinea*, Institute of National Affairs, Port Moresby, PNG.

Gibson, John, 1998, Urban demand for food, beverages, betelnut and tobacco in Papua New Guinea, *Papua New Guinea Journal of Agriculture, Forestry and Fisheries* 41(2): 37-42.

Gibson, John, 1997, Women's education and child growth in urban Papua New Guinea, *Papua New Guinea Journal of Education* 33(1): 29-31.

Gujarati, Damodar, 1988, *Basic Econometrics* (McGraw-Hill, New York).

Ravallion, Martin, 1990, Income effects on undernutrition, *Economic Development and Cultural Change* 38(3): 489-515.

Robinson, Peter, 1988, Root-*N*-consistent semiparametric regression, *Econometrica* 56(4): 931-954.

Ruel, M. T. and J. L. Garrett. 1999. "Urban Challenges to Food and Nutrition Security in the Developing World: Overview," *World Development* 27 (11 November): 1885-1890.

Strauss, John and Duncan Thomas, 1990, The shape of the calorie-expenditure curve, *mimeo* New Haven, Conn. Yale Univ.

Strauss, John and Duncan Thomas, 1995, Human resources: empirical modelling of household and family decisions, in: J. Behrman and T.N. Srinivasan, eds., *Handbook of Development Economics Vol. III*, (North-Holland, Amsterdam) 1883-2023.

Subramanian, Shankar and Angus Deaton, 1996, The demand for food and calories, *Journal of Political Economy* 104(1): 133-162.

Wolfe, Barbara and Jere Behrman, 1983, Is income overrated in determining adequate nutrition? *Economic Development and Cultural Change* 31(3): 525-549.

Table 1: Per capita expenditures and calorie availability

	Real per capita total expenditure (kina/fortnight)[a]	Food share of total expenditure (%)	Per capita daily calorie availability[b]	Adjusted per capita daily calorie availability[c]
Expenditure quartile				
Poorest	15.66	56.3	1481	1525
II	27.97	53.5	2148	2189
III	44.98	50.2	2842	2755
Richest	103.01	38.8	3614	3370
ALL	47.90	45.1	2524	2463

Source: Papua New Guinea Urban Household Survey, 1985-87

[a] 1985 prices (Kina 1.00 = US$1.00 in that year).

[b] Derived from food expenditures, with allowances for self-produced foods, net gifts of food (excluding meals), and net food stock changes (not measured for minor foods).

[c] Adjusted for meals fed to non-residents and meals received in other households.

Table 2: Calorie elasticities evaluated at potential reporting points

	Household rank *(%-ile)*	Person rank *(%-ile)*	Calorie elasticity (*LOWESS estimates)*
Household at mean ln (PCE)	53.1	64.3	0.42
Median household	50.0	60.7	0.43
Household containing the median person	40.2	50.0	0.48

Note: Ranks are from lowest to highest per capita expenditure.

Table 3: OLS estimates of calorie availability regressions

	(1)		(2)		(3)		Within cluster (4)	
	β	|*t*|	β	|*t*|	β	|*t*|	β	|*t*|
ln PCE	.566	(10)	.566	(10)	.586	(11)	.506	(8.6)
ln PCE * Q3 Dummy	-.281	(1.4)	-.214	(1.1)	-.173	(.9)	-.150	(.7)
ln PCE * Q4 Dummy	-.433	(6.0)	-.433	(5.8)	-.409	(5.5)	-.277	(3.3)
ln household size	-.222	(11)	-.183	(7.3)	-.163	(6.1)	-.179	(5.7)
rf15+			.030	(.4)	.051	(.6)	.040	(.4)
rm714			-.083	(.9)	-.066	(.7)	-.123	(1.1)
rm06			-.001	(.0)	.021	(.2)	.123	(1.1)
rf714			-.191	(1.8)	-.190	(1.8)	-.288	(2.4)
rf06			-.078	(.7)	-.048	(.4)	-.063	(.5)
Expatriate head			.068	(1.0)	.048	(.7)	.015	(.2)
Highlands head			-.014	(.4)	-.009	(.3)	-.036	(.8)
Female head			-.080	(1.2)	-.089	(1.4)	-.101	(1.4)
Age of head			.002	(1.7)	.002	(1.8)	.003	(2.0)
Wage job			.008	(.2)	.023	(.7)	.067	(1.8)
Formal business			.059	(1.4)	.090	(2.1)	.083	(1.6)
Informal business			.058	(1.4)	.055	(1.3)	.037	(.8)
Female school years			-.004	(1.0)	-.006	(1.4)	-.005	(1.0)
Male school years			-.003	(.9)	-.003	(.8)	-.003	(.7)
Has a garden?			-.118	(4.8)	-.063	(1.3)	-.203	(.7)
ln price of:								
Bread and biscuits					.088	(.3)		
Rice					-.053	(.1)		
Flour					-.753	(1.3)		
Banana					.125	(1.7)		
Coconut					-.108	(.9)		
Sweet potato					-.168	(1.6)		
Sugar					1.090	(1.8)		
$\bar{R}^2$	.451		.468		.479		.548	

Note: Models also contain an intercept and intercept dummies for the third and fourth population quartiles. Variables beginning with *r* are demographic ratios, so that e.g., rf714 is the ratio of females aged 7-14 to total household members. The omitted group is male adults. The omitted ethnic group is household heads from the lowlands. The omitted economic activity group is household heads who are unemployed. The within cluster regression contains 282 dummy variables for clusters. The *F*-test for the exclusion of the cluster effects is 1.64 with 282 and 729 degrees of freedom.

Table 4: OLS and IV estimates of calorie demand elasticity

Covariates	OLS Estimates	Instrumental Variables estimates: ln g as instrument[a]	Instrumental Variables estimates: ln y as instrument[b]
ln PCE	0.419	0.307	0.414
	(0.017)	(0.019)	(0.025)
		[R^2=0.86]	[R^2=0.47]
ln PCE, ln n	0.325	0.204	0.304
	(0.018)	(0.020)	(0.030)
		[R^2=0.88]	[R^2=0.53]
ln PCE, ln n, demographics, economic activity, schooling, food prices	0.375	0.199	0.402
	(0.023)	(0.026)	(0.055)
		[R^2=0.90]	[R^2=0.60]
ln PCE, ln n, demographics, economic activity, schooling, cluster effects	0.376	0.176	0.519
	(0.028)	(0.033)	(0.079)
		[R^2=0.93]	[R^2=0.76]

Notes: Standard errors in ( ). R^2 from the first stage regression in [ ].

[a] ln g is the logarithm of per capita non-food expenditures.

[b] ln y is the logarithm of per capita income.

Endnotes

[1] An exception is the studies in the special issue of *World Development* (vol. 27, no. 11, 1999) devoted to food security and nutrition in urban areas.

[2] More disaggregated data (58 food groups) were available for one province. When these were used to form calorie availability estimates, the correlation with the availability estimates from the 35 food group data was 0.999. The elasticity of per capita calorie availability with respect to per capita expenditure was 0.598 using the 35 food groups, and 0.596 using the 58 food groups. This suggests that there is only a very slight upward bias in the elasticity when it is estimated from the more aggregated data. The Pacific Islands Food Composition Database was used to compute the calorie quantities from the food quantity data. One item where food quantities were not available was cooked meals eaten out of the home; calories from this source were derived as the average "price" each household paid for all other calories plus a 50 percent premium to reflect processing margins (Subramanian and Deaton, 1996).

[3] The direction of bias for the coefficient on the included variable when a relevant variable is excluded from the model is given by the product of the coefficient on the excluded variable (for the true model) with the correlation coefficient between the excluded and included variables (Gujarati, 1988, p.403). Both the correlation coefficient and the (excluded) regression coefficient on household size are presumed to be positive in the current example, so the effect will be to cause a positive bias in the expenditure elasticity of calorie demand if household size is excluded from the model.

[4] The bootstrapping did not take account of the two-stage sample design which was used by the Urban Household Survey. In this two-stage design, approximately 300 census enumeration areas, containing 50 households on average, were first selected, and at the second stage households within these areas were selected. In principle a bootstrapping experiment could be designed to preserve this feature of the sampling design but the difficulty is that a variable number of households (between two and ten) were selected from each enumeration area, to ensure self-weighting. It would be very complex to replicate this variable cluster size in a resampling experiment, unlike the sample design confronting Subramanian and Deaton, which had a fixed number of households drawn from each cluster. Ignoring the two-stage sampling should not understate the standard errors greatly because the average number of households selected per cluster was only 3.8, and Deaton and Subramanian found only a small design effect using a sample where 10 households were selected from each cluster.

[5] One of the most interesting results in column (2) is for average schooling levels, disaggregated by gender. In contrast to Behrman and Wolfe (1984), the conditional effect of female education on calorie consumption is negative and statistically insignificant. This is notable because the same schooling variables suggest that female education has a very favourable effect on the growth of young children in these same households (Gibson, 1997). Thus, it may be that female schooling improves the efficiency of child health production, rather than shifting resource allocations within the household towards more calorie intensive budgets.

[6] Instrumental variables estimation of a spline model was also attempted but the results did not seem to be consistent with the other estimates. Using nonfood expenditure as the instrument and ln PCE as the only covariate gave estimated elasticities of 0.058, -3.285, and -0.077 for the first two, the third, and the fourth quartiles. As Strauss and Thomas (1995) suggest, interactions with functional form may wreak havoc with IV estimators. Moreover, the properties of the estimator with nonfood expenditure as the instrument have only been worked out for a log-linear structure, rather than a non-linear function.

[7] The table also contains estimates of the explanatory power of the instruments in the first stage regression because results may not be robust if the first stage regression has little explanatory power (Strauss and Thomas, 1995).

[8] These tests are based on the added-variable approach, with the residuals from the first stage regression of log PCE on the instrument(s) added to the second stage model and a *t*-test on the coefficient on these added residuals indicates whether IV and OLS results differ significantly (Davidson and MacKinnon, 1993, pp.232-242). When non-food expenditure was used as the instrument, these *t*-values (with one degree of freedom) were between 14.3 and 18.9, depending on the other covariates used in the model.