# An Engel Curve for Variety

Nicholas Li*

U.C. Berkeley

November 10, 2010 – Version 1.0

*Job Market Paper*

**Abstract**

Standard approaches to measuring the welfare gains from variety take variety as exogenous to consumers. Empirically we observe an Engel curve for variety, with richer consumers purchasing a larger set of varieties than poor ones. I present a tractable model for analyzing the impact of income, relative prices and transaction costs on the welfare gains from variety. The model generates variety Engel curves through diminishing marginal utility of quantity and transaction costs. Estimates of welfare gains that assume all variety differences across households are exogenous will be biased in the presence of variety Engel curves. Variety Engel curve slopes depend on the relative price and transaction cost of marginal varieties and therefore affect the distribution of gains from variety and the measurement of real (welfare) inequality. I use data on food consumption in India to implement the model and show that variety Engel curves have significant empirical implications for measurement of the size and distribution of welfare gains from variety.

*I am grateful to Pierre-Olivier Gourinchas, Chang-Tai Hsieh, Ted Miguel, Yuriy Gorodnichenko, and Mu-Jeung Yang for valuable discussions and to participants of the Development and Macro/International Seminars and workshops at UC Berkeley. I gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada, the Institute for Business and Economics Research and the Center for Equitable Growth at UC Berkeley. All errors are my own. Comments welcome. nickli@econ.berkeley.edu

# 1. Introduction

The typical approach to measuring the welfare gains from variety in the trade and macroeconomics literature takes the set of varieties as exogenous to consumers (Feenstra (1994), Broda and Weinstein (2006), Broda and Weinstein (2010)). Consumers purchase the same set of varieties regardless of their expenditures, a set determined only by firm entry, production and export fixed costs. This assumption is common to constant elasticity of substitution (CES) preferences, which imply unlimited demand for variety independent of prices, and other homothetic preferences including homothetic translog (Feenstra (2009)) and quadratic utility with an outside good (Melitz and Ottaviano (2008)) that imply finite reservation prices.

The limitation of treating variety as exogenous is apparent when examining cross-sectional patterns of variety consumption. There is overwhelming evidence that richer households and countries consume more varieties than poor ones, e.g. an Engel curve for variety. Figure 1 plots the number of distinct food items consumed as a function of food expenditures for a cross-section of households in Spain, India and the United Kingdom.[1] Figure 2 plots the number of import varieties as a function of real GDP per capita (panel A) for a cross-section of countries.[2] While trade models with export fixed costs predict that larger markets will consume more imported varieties, the positive correlation with income per capita persists even after controlling for real GDP (panel B). Upward sloping variety Engel curves also shift over time and across areas. Figure 3 presents linear variety Engel curves for two food groups in the Indian state of Karnataka. Variety tends to be higher in urban areas and later periods for any level of real expenditure, but the slopes are also different.

I present a tractable model for analyzing variety Engel curves and their implications for welfare. Diminishing marginal utility of quantity generates a benefit to expanding the set of varieties consumed, a benefit that increases in total expendi-

[1]The data come from representative household consumption/expenditure surveys. I trim the top and bottom 1% tails of the expenditure distribution for each country and restrict the sample to households with three members. This pattern holds when looking across households in the same area and is not simply a result of geographic correlation between richer households and the supply of variety. Broda and Romalis (2010) find a similar pattern for UPCs when examining consumption of nondurables by American households, see their figure 5A.

[2]Each variety is defined as a four-digit SITC category by country of origin cell. The trade data come from the NBER-United Nations Trade Data for 2000. GDP and population data come from the Penn World Tables. Note that a similar pattern holds within most four-digit SITC categories and when restricting to goods that are obviously intended for final consumption.

tures but also depends on the asymmetry in prices and tastes across varieties. This benefit is balanced against a variety transaction cost that is increasing in the set of varieties consumed. The combination of these two forces implies a first-order condition for variety choice where consumers equate marginal benefit (which increases in expenditures) with marginal transaction cost, generating upward-sloping variety Engel curves (figure 4).

The model has two key implications for welfare analysis. First, *treating variety differences as exogenous in the presence of variety Engel curves will bias estimates of welfare gains from variety.* Suppose we compare two consumers with different expenditures ($x_2 > x_1$ in figure 4) and one consumes more variety ($n_2 > n_1$). By treating this variety difference as exogenous (equivalent to a vertical marginal cost curve in figure 4) the welfare gain from variety for the rich consumer is large (area $A + B + C$ in the figure). The poor consumer would have a smaller welfare gain from consuming the same set of varieties as the rich one (area $C$) but this gain is unrealized due to supply-side forces. In my model, the non-vertical marginal cost curve ($MC$) generated by transaction costs generates endogenous variety differences across consumers with different expenditures. This implies a smaller welfare gain from variety for the rich consumer (area $A$), and implies that the poor consumer would lose welfare (equivalent to area $B$) by consuming a larger set of varieties. Treating variety as exogenous thus amplifies the welfare differences caused by variety relative to a model where some differences are endogenous and related to movements *along* a variety Engel curve. In general, when comparing welfare gains from variety across areas or periods with different relative prices and transaction costs we still need to control for movement along variety Engel curves to avoid confounding the effects of expenditures with those of exogenous factors.

Second, *the slope of the variety Engel curve is related to the distribution of welfare gains from variety.* From figure 4 the slope of the variety Engel depends on the slope of the marginal benefit curve and marginal cost curves. A flat variety Engel curve implies lower welfare inequality, as higher expenditure consumers experience sharper diminishing returns to quantity. A steep variety Engel curve implies higher welfare inequality, allowing richer consumers to counteract diminishing returns by expanding on the extensive margin. The extreme case of exogenous but higher variety for rich households yields the greatest welfare inequality, while the case of exogenous but identical variety yields the lowest. My model permits an interpretation of the welfare gains from greater variety due to shifting variety Engel curves (e.g. figure 3) as rich-biased or poor-biased depending on the change in the slope.[3]

I provide an empirical application of the model by analyzing the massive growth of average food variety in India between 1983-2005 and the higher average food variety of urban areas. I adopt a simple parameterization of the variety Engel curve model that nests CES preferences, enabling a comparison of welfare gains under my model with those under CES preferences with exogenous variety. I find welfare gains equivalent to 9.4% of food expenditures from falling variety transaction costs over the 1983-2005 period. This is large relative to growth of food expenditures (close to zero) and total expenditures (20%-40%). Households living in urban areas have lower transaction costs than those in rural areas with a resulting welfare gain equivalent to 2.3% of food expenditures. This is large relative to an average urban-rural food expenditure gap of 14%. Under CES preferences, welfare gains over time are overestimated for food groups with rising expenditures and underestimated for food groups with falling expenditures. CES preferences systematically overstate the average urban-rural welfare gain by a factor of two as they do not account for the fact that urban households are richer on average. The distributional effects of changing variety Engel curve slopes are also sizeable, with rural households at the 90th percentile of the food expenditure distribution experiencing 20%-33% greater welfare gains from variety than those at the 10th percentile between 1983-2005. Urban variety gains are up to 70% greater for households at the 90th percentile relative to the 10th percentile in 1983, but this rich-bias has diminished over time.

As much of the estimated welfare gains from variety are related to the transaction cost parameter in the model, my findings suggest that policies to reduce these costs, including improved infrastructure, removal of barriers to modern retailing, and food market integration, could yield large improvements in consumer welfare in India and other developing countries where the cost of variety can be high. I show directly that one measure of transaction costs, shopping time, behaves as predicted by the

[3]Broda and Romalis (2010) consider distributional effects of variety but rely on heterogeneous preferences to explain the different sets of varieties consumed by rich and poor households - variety is treated as exogenous with respect to income decile. They find that the prices of varieties consumed by the rich rose more but that they also gained more welfare from new varieties than poor households. My model provides a mechanism that explains why rich households consume more (and different) varieties than poor households, how changes in relative prices affect welfare and the choice of varieties consumed, and how changes in the availability (transaction costs) of varieties consumed disproportionately by poor or rich households affects the cost-of-living with common preferences.

model. District level measures of market access and infrastructure, which would tend to reduce variety transaction costs, have a significant effect on variety conditional on expenditure but have minimal effect on the aggregate set of varieties consumed in a district.

The source of variety gains in my model is diminishing marginal utility of quantity, which is distinct from the usual source considered in the Industrial Organization and Marketing literature. Discrete choice and Lancaster models imply welfare gains from variety due to heterogeneity in tastes - greater variety allows the average consumer to get closer their "ideal" variety.[4] This approach, which may be more appropriate for studying narrow product categories or those for which households do not purchase multiple units, does not imply a household extensive margin or variety Engel curve. While is a small literature considering the interaction of income and variety choice in these models, the nature of the welfare gain is different and the relevance of the ideal variety and variety Engel curve models depends on the context.[5]

A related literature uses Engel curves to measure bias in price indexes relative to a model-implied cost-of-living index. Costa (2001) and Hamilton (2001) use downward drift in food Engel curves to measure CPI bias, while Almas (2008) uses the same technique in the cross-section to measure bias in the Penn World Tables. Bils and Klenow (2001a) use quality Engel curves to measure the extent to which the BLS overstates inflation for durables by underestimating quality improvements. Unlike the food Engel curve literature, the non-homotheticity in my model is not used to calculate welfare gains by assuming a stable functional relationship between real expenditures and variety. Variety Engel curves are used to compute bias due to endogenous variety consumption, but the welfare gains are based on conventional measures of consumer surplus. While the other studies rely on constant slopes for identification and compute a common change in the cost-of-living, in my model changes in the slope of variety Engel curves identify changes in the distribution of welfare.

In section 2 I present the theory of variety Engel curves. Section 3 describes the Indian National Sample Survey data and interprets some stylized facts in light of the

[4]It is also possible to interpret the aggregate welfare gains under CES preferences in relation to distance from ideal type. Anderson et al. (1992) show that a logit discrete choice framework can generate aggregate level CES demand.

[5]Allenby and Rossi (1991) analyze non-homothetic discrete choice, though their focus is on pricing and not welfare gains from variety. Hummels and Ludovsky (2009) consider a generalized Lancaster (1979) model where the valuation of proximity to "ideal type" rises in income, but focus on theoretical general equilibrium implications.

model. Section 4 estimates the model and uses it to analyze the welfare implications of variety for growth and inequality in India. Section 5 examines factors that influence the cost of variety and Section 6 concludes.

# 2. The Theory of Variety Engel Curves

## 2.1. CES love of variety

The standard constant elasticity of substitution (CES) model of variety has a representative agent solving

$$\max_{q_i} \left( \int_0^n q_i^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} \; s.t. \int_0^n p_i q_i di \leq X \quad (1)$$

where $\sigma > 1$ is the elasticity of substitution, p is price, q is quantity, X is total expenditure, n is the measure of varieties consumed and i indexes the varieties. Households take the set of varieties in the economy (indexed [0,n] for simplicity) as fixed and allocate quantities across the different varieties to maximize welfare. The resulting demand function is $q_j = \frac{X}{p_j}\left(\frac{p_j}{P(n)}\right)^{1-\sigma}$ where $P(n)$ is the CES price aggregator $P(n) \equiv (\int_0^n p_i^{1-\sigma} di)^{\frac{1}{1-\sigma}}$. The expenditure function gives the nominal expenditures necessary to reach utility level $U_0$ and is given by $X(U_0, \{p_i\}) = U_0 P(n)$. In this model variety counteracts the diminishing marginal utility of quantity implied by $\sigma > 1$. The expenditure function is strictly decreasing in n - hence the term ‘love of variety’ - and since n is costless, demand for varieties is unlimited. It can only be constrained exogenously by the set of varieties available (through production fixed costs). Income has no effect on demand for variety.

A simple parameterization of prices allows us to compute the marginal benefit (utility) of variety explicitly. I adopt an exponential distribution of prices given by $p_i = p\{[\psi(\sigma-1)]^{\frac{1}{1-\sigma}}\} i^{\frac{1}{\theta}}$, which emerges from a model with monopolistically competitive firms with productivity parameters drawn from a Frechet distribution. The term $p$ captures a level shift in prices, which would result if the productivity of all firms in the economy was scaled by the same factor. The second term in curly brackets is a common scaling factor that simplifies some later algebra. The parameter $\frac{1}{\theta}$ measures the dispersion of firm productivity and hence variety prices. The curvature introduced by rising prices implies a hierarchy of varieties ordered from lowest to highest price. By convention a cheaper variety has a lower index.[6] The CES price aggregator is given by $P(n) = pn^{-\psi}$. The variety elasticity of the expenditure function is constant and given by $\frac{\partial X(U_0,\{p_i\})}{\partial n}\frac{n}{X} = -\psi$, where $\psi \equiv \left[\frac{1}{\sigma-1} - \frac{1}{\theta}\right]$. A relatively higher price for the next variety in the hierarchy implies a lower marginal benefit to variety. As $\theta \to \infty$ we have constant prices and symmetric CES. The log marginal benefit of consuming an additional variety is:

$$log(MB) = log(X/p) - (1-\psi)log(n) + log(\psi) \tag{2}$$

Log marginal benefit is higher for consumers with higher expenditures, and while log MB can fall below zero for high enough n, MB itself remains positive. The elasticity $\psi$ determines the rate at which marginal benefit falls as variety increases. If varieties are not too close as substitutes ($\sigma > 1$) or there is enough curvature in prices ($\frac{\theta}{1+\theta} < \sigma - 1$) so that $0 < \psi < 1$, the log marginal benefit of variety is decreasing in variety.

Panel A of figure 5 depicts equilibrium variety choice by plotting log marginal benefit against log variety. The standard CES model gives rise to an equilibrium where the vertical bar at $\bar{n}^1$, representing the exogenous variety limit, intersects the log marginal benefit curve. If we plotted the equilibrium in marginal benefit space rather than log marginal benefit space, the area under the marginal benefit curve gives total welfare from variety.

An increase in expenditures from $X_1$ to $X_2$ shifts the log marginal benefit curve up from $MB_1(X_1)$ to $MB_2(X_2)$. Welfare increases but $\bar{n}^1$ is still binding so variety is unaffected. A decrease in $\theta$ (rise in relative prices) lowers both the slope and intercept of the log(MB) curve. This decreases welfare but the percent change in the cost-of-living index is identical for rich and poor households and variety consumption remains at $\bar{n}^1$.

An exogenous variety increase from $\bar{n}^1$ to $\bar{n}^2$ affects rich and poor households symmetrically. While the richer household receives a greater total increase in utility from variety, and would be willing to pay more than the poor household, the percentage change in the cost-of-living index is identical. In other words all households would pay a constant share of their expenditures to consume an additional variety.

[6]This formulation follows the work of Arkolakis et al. (2007) who use the Pareto distribution due to its tractability (especially with CES preferences) and its widespread use in the trade and growth literature.

## 2.2. Consumer fixed costs and endogenous variety demand

I now modify the CES framework by adding a friction to variety consumption. This leads to an endogenous "Engel Curve for Variety" with richer households consuming a larger set of varieties than poor households, consistent with figures 1 and 2. I assume that this friction takes the form of fixed utility costs incurred by consumers. In Appendix A I consider the different implications of variety Engel curves based on fixed budget costs or bounded marginal utility.

There are many costs to consuming greater variety that may be distinct from the costs of consuming greater quantity (i.e. the unit price). These include time, money, and complementary inputs related to search, shopping, travel, storage and preparation. Consumers may have to learn which varieties they prefer and how to prepare them (see Atkin (2009)). These factors may be important for imported goods as well, as many of them are not easily accessible to all households in a country. Other factors that may constrain variety consumption at the household level are indivisibilities, minimum scale of consumption, and bulk discounting, all of which may prevent the purchase of infinitesimal quantities. The exact nature of these costs will vary with the setting, and the key distinction is that variety costs are distinct from preferences and prices.[7]

The consumer problem has two stages. In stage one, consumers takes variety $n$ as fixed and allocate expenditure across varieties as if they had to homothetic CES preferences. In stage two, they choose optimal $n$ to balance gains, which come through the effect on the CES price aggregator P(n), with utility costs given by F(n). The second-stage problem is:

$$\max_n \frac{X}{P(n)} - F(n) \tag{3}$$

where the first term corresponds to standard CES utility $U$ and the second term is the utility cost of consuming measure n of varieties. I assume that the variety cost function F(n) is increasing in n. Additive separability in both the first and second stage is critical to the simple two-stage optimization. Varieties lie along a continuum and both P(n) and F(n) are smooth functions of n.

[7]More generally, one could consider part of these costs as a preference parameter. The key assumption required for analyzing welfare is that differences in variety consumption are due to differences in real costs facing households and not just preferences.

The first order condition for optimal variety choice n is given by:

$$\underbrace{-\frac{X}{P(n)^2}\frac{\partial P(n)}{\partial n}}_{MB} = \underbrace{\frac{\partial F(n)}{\partial n}}_{MC} \tag{4}$$

Unless the choice of n is constrained exogenously, households equate the marginal benefit and marginal cost of consuming additional variety. Expenditures show up in the first order condition by increasing the marginal benefit of variety, leading richer households to consume more varieties endogenously. The fixed costs limit variety choice.

I adopt $P(n) = pn^{-\psi}$ from earlier and use a similar exponential function for fixed costs, $F(n) = Fn^{\epsilon}$. The second order condition for an interior solution requires an elasticity of welfare losses from fixed costs greater than the elasticity of CES welfare gains from variety, or $\epsilon > \psi$. An advantage of this approach to bounding variety endogenously is that it generates a closed-form solution for variety choice and the expenditure function.[8]

The first-stage maximization is standard CES taking n as fixed. The second-stage maximization problem is:

$$\max_n \frac{X}{pn^{-\psi}} - Fn^{\epsilon} \tag{5}$$

and taking logs of both sides of the first order condition gives

$$\underbrace{log(\psi) + (\psi - 1)log(n) + log(X/p)}_{log(MB)} = \underbrace{log(\epsilon) + log(F) + (\epsilon - 1)log(n)}_{log(MC)} \tag{6}$$

The equilibrium that equates log marginal benefit and log marginal cost is shown in panel B of figure 5. The graph shows the case with $\psi < 1 < \epsilon$ but the only requirement for an interior optimum is that $\epsilon > \psi$. The optimal choice of variety is given by $n = \left(\left[\frac{X}{p}\right]\frac{\psi}{F\epsilon}\right)^{\frac{1}{\epsilon-\psi}}$. Households face a similar menu of prices and costs and have access to the same set of varieties, but the model implies that the rich consume more varieties than the poor. Taking logs yields a log-linear variety Engel curve with slope

[8]In theory one could substitute any homothetic expenditure function that varies in n for P(n), an idea explored in appendix A, or we could use other functional forms for fixed costs but analytic solutions and aggregation may prove difficult.

$\frac{1}{\epsilon-\psi}$

$$\ln n_h = \frac{1}{\epsilon-\psi}\ln(\frac{\psi}{\epsilon}) + \frac{1}{\epsilon-\psi}\ln X_h - \frac{1}{\epsilon-\psi}\ln p - \frac{1}{\epsilon-\psi}\ln F \tag{7}$$

where h indexes households. The log-linearity of the variety Engel curve is a special feature of the assumed exponential functional form and constant elasticities of P(n) and F(n). Log-linearity is a reasonable approximation in light of figures 1 and 2 though there is some evidence that the slope decreases for very high expenditures. The expenditure function is given by:

$$X(U_0, \eta, p, F) \equiv U_0^{1-\eta} p F^{\eta} \Psi \tag{8}$$

where $\eta \equiv \frac{\psi}{\epsilon}$ and $\Psi \equiv \left[\eta^{\eta/(1-\eta)} - \eta^{1/(1-\eta)}\right]^{\eta-1}$. The demand system is homogeneous of degree zero in prices, as scaling up by common factor $p$ increases the cost-of-living by the same percent for all households. The fixed cost scalar F affects households symmetrically. Changes in price dispersion (through $\psi$) and the variety-cost slope ($\epsilon$) do not affect households symmetrically as they affect the mapping from $U_0$ to X - the dispersion of welfare relative to expenditures - that is related to the slope of the variety Engel curve.

## 2.3. Comparative statics

### 2.3.1. Changes in expenditure (X)

The left panel of figure 6 depicts the effect of a rise in expenditure. The increase from $X_1$ to $X_2$ shifts the marginal benefit curve upwards. If marginal cost is unchanged and there is no exogenous constraint, variety consumption rises from $n_1$ to $n_2$, the point where the marginal benefit and marginal cost are once again equalized. In a cross-section of households or countries, we can think of variation in expenditure X generating endogenous movements in n, which graphically represents movement *along* a variety Engel curve in (log $X$,log $n$) space (right panel of 6). The slope of this Engel Curve - which in the parameterized case is $\frac{1}{\epsilon-\psi}$ - depends on the slopes of the marginal benefit and marginal cost curves, while the scaling factors for price ($p$) and fixed cost ($F$) only affect the intercept.

The welfare impact of higher expenditures can be broken down into two parts:

$$\frac{\partial U}{\partial X} = \frac{\partial U}{\partial X}\Big|_{n=n_0} + \frac{\partial U}{\partial n^0}\frac{\partial n^0}{\partial X} \tag{9}$$

The first part of the welfare gain is identical to the one in the standard CES model where the set of varieties is held fixed. The second part of the welfare gain comes from allowing the household to re-optimize the choice of varieties, counteracting the diminishing marginal utility of consuming higher quantities of the original variety set. The ability to re-optimize effectively increases the gap in welfare between two households for a given gap in expenditures, relative to a world where variety is exogenous. In terms of figure 6, area A is the standard gain in welfare from higher expenditures evaluated at the original $n_0$, area B is the extra gain from re-optimization.

Note that the cost-of-living function from equation 8 *already* accounts for movements along a variety Engel curve, so changes in variety induced by changes in expenditures are not relevant for comparing the effects of variety on the cost-of-living over time and space. However, the ability to re-optimize the choice of variety based on expenditures amplifies the dispersion of welfare across poor and rich households. The size of the re-optimization effect by area B is a welfare gain from variety for rich households that depends on the slope of the variety Engel curve - it is a gain *relative* to a world with flat variety Engel curves.

### 2.3.2. Changes in price dispersion ($\theta \to \psi$)

A fall in $\theta$ represents a rise in prices for all varieties and an increase in relative prices for high index varieties. This lowers the marginal benefit of variety through its effect on $\psi \equiv \frac{1}{\sigma-1} - \frac{1}{\theta}$, the CES aggregator - variety elasticity.[9] The rise in prices and consequent decline in $\psi$ has two effects on the MB curve - the intercept shifts down and the slope decreases. The left panel of figure 7 shows how this lowers optimal variety consumption, and the right panel shows that the slope of the variety Engel curve $\frac{1}{\epsilon-\psi}$ decreases.

Area A gives the standard effect of rising prices holding variety constant - $\frac{\partial U}{\partial \psi}\Big|_{n=n^0} > 0$. This has the typical homothetic CES effects, with households substituting relatively cheaper low index varieties for high index varieties and lowering quantities.

[9]Note that a mean-preserving rise in price dispersion would require a decrease in the common price term $p$ that would depend on the average level and distribution of household expenditures.

As $\frac{\partial n^0}{\partial \psi} > 0$, n decreases and the household stops consuming some higher index varieties altogether. Area B would be a welfare loss for a household forced to consume the original set of varieties ($n_1$) instead of dropping the more expensive ones whose marginal cost exceeds the now lower marginal benefit. Re-optimization of the variety set thereby decreases the negative welfare impact of rising prices.

Differentiating the expenditure function and rearranging we can derive the elasticity of the cost-of-living index with respect to the parameter $\psi$:

$$\frac{\partial X}{\partial \psi}\frac{\psi}{X} = \frac{\psi}{\epsilon}\left(log(F) - log(U_0)\right) + \frac{\partial \Psi}{\partial \psi}\frac{\psi}{\Psi} \tag{10}$$

where $\Psi \equiv \left[(\epsilon/\psi)^{-\frac{\psi}{\epsilon-\psi}} - (\epsilon/\psi)^{\frac{\epsilon}{\epsilon-\psi}}\right]^{-\frac{\epsilon-\psi}{\epsilon}}$. Only the term $U_0$ varies across households. It can be shown that this elasticity is negative so that a rise in the price slope and fall in $\psi$ increases the cost-of-living. Because the $U_0$ term enters negatively, the cost-of-living rises by more for rich households than poor households.

In the homothetic CES model the relative cost-of-living of high and low utility households is unaffected by changes in the distribution of prices. In this model high utility households with a greater budget share of high index varieties experience larger increases in their cost-of-living when the price of "luxury" varieties rises. Higher (lower) relative prices for high index varieties flatten (steepen) the variety Engel curve leading to lower (higher) welfare inequality.

Broda and Romalis (2010) also find that lower and higher priced varieties are imperfect substitutes and are consumed differentially by rich and poor households. They show that rising relative prices for varieties consumed by rich households can reduce welfare inequality relative to nominal income inequality. My model differs in that the ability to re-optimize the variety choice set - substituting away from the higher priced varieties to lower priced varieties - partly counteracts this effect. If variety choice is endogenous the set of goods consumed by rich and poor households cannot be thought of as fixed when assessing the welfare impact of price changes.

### 2.3.3. Changes in fixed costs F(n)

A fall in fixed costs through the $\epsilon$ parameter decreases the slope and intercept of the log marginal cost curve. The effect on the variety Engel curve is the mirror image of a fall in $\psi$, raising the slope and intercept. Figure 8 depicts the effect of a fall in $\epsilon$ and the associated infra-marginal gain (A) and the re-optimization gain (B). The effect on the cost-living index is given by:

$$\frac{\partial X}{\partial \epsilon}\frac{\epsilon}{X} = \frac{\psi}{\epsilon}\left(log(U_0) - log(F)\right) + \frac{\partial \Psi}{\partial \epsilon}\frac{\epsilon}{\Psi} \tag{11}$$

This expression, the mirror image of equation 10, is always positive but is greater for richer households with higher $U_0$. A fall in $\epsilon$ makes marginal varieties less costly to households. All households consume more variety but richer households consume proportionately more and their cost-of-living falls by a greater amount. This is reflected in the steepening of the variety Engel curve and implies a decrease in welfare inequality relative to expenditure inequality.

Fixed costs can also fall through the scalar F, leading to a parallel shift down in the log marginal cost curve and shift up in the variety Engel curve. The cost-of-living index (equation 8) has an elasticity of $\frac{\psi}{\epsilon}$ with respect to F, which does not depend on utility, so the cost-of-living decreases proportionately for any level of utility.

Although changes in the log marginal cost and benefit slopes $\epsilon$ and $\psi$ have symmetric effects on variety Engel curves and the cost-of-living index, there is one key difference. The CES functional form implies that we can write relative budget shares for two consumed varieties as $\ln \frac{s_i}{s_j} = (\psi[\sigma - 1] - 1) \ln \frac{i}{j}$, which varies with $\psi$ but not with $\epsilon$. A significant advantage of using consumer fixed costs to bound variety demand is that it allows changes in variety consumption *without* changes in relative budget shares for consumed varieties. Variety Engel curves in models of bounded marginal utility necessarily imply that prices simultaneously determine the shape of the variety Engel curve and the ratio of budget shares/quantities (see appendix A for more discussion).

### 2.3.4. Changes in exogenous variety limits

The analysis so far assumed an interior optimum. In general equilibrium production fixed costs (like firm entry costs) would limit the total set of varieties exogenously, which we can denote by an exogenous upper bound $\bar{n}^2$. Figure 9 shows that this bound constrains variety for the richest households with $x_2$ and $x_3$ but has no effect on a poor household with $x_1$. The portion of the variety Engel curve above $\bar{n}^2$ is flat. Suppose now the bound increases to $\bar{n}^3$. There is no change in welfare for the household $x_1$ that was previously at an interior optimum ($x_1$) but welfare rises for households with $x_2$ and $x_3$. Households for which the new constraint still binds ($x_3$) benefit the most. Compared to a model with no transaction costs, the welfare gains from this exogenous (supply-driven) variety growth are reduced, but there is still a substantial gain to the extent that the marginal benefit exceeds the marginal cost for the new varieties.

An implication of the model is that the relaxation of exogenous variety constraints never benefits poor households. It is still possible to analyze "new" varieties that disproportionately benefit poor households in terms of an increase in the relative marginal benefit of low-index varieties (a simultaneous fall in p and $\psi$), or a decline in low-index fixed costs (fall in F and rise in $\epsilon$). If variety gains are concentrated in groups that have a low across-group income elasticity poor households will also experience greater welfare gains. The discipline imposed by theory is that when demand for variety is driven by diminishing marginal utility in quantity, preferences take a hierarchical form with the rich consuming a superset of the varieties consumed by the poor.[10]

## 2.4. Comparison of homothetic vs. non-homothetic love of variety

The positive correlation between income and variety in consumption and trade data suggests that ignoring income-effects may bias predicted trade patterns and welfare gains from variety. As there are already trade models that analyze income effects as a source of trade patterns I emphasize the normative implications of variety Engel curves relative to homothetic models. [11]

[10]Discrete choice models in which welfare gains from variety come from 'proximity to ideal type' and random taste variation disallow new varieties that exclusively benefit poor OR rich households. In contrast, my model implies that there can be new varieties that only benefit the rich, but there cannot be new varieties that only benefit the poor.

[11]Most of the literature has focused on non-homotheticity and import quality rather than the extensive margin. The exceptions feature an aggregate, non-homothetic consumer with bounded marginal utility (Saure (2009) and Foellmi et al. (2010)) or a Lancaster model of ideal variety Hummels and Ludovsky (2009). Holding the fixed cost constant, my model generates similar predictions about the effects of income per capita on aggregate import variety but has additional implications for the effects of the income distribution and the distribution of welfare gains within countries. It also adds another degree of freedom with fixed costs that can shift variety demand holding income and prices constant.

### 2.4.1. Understatement of welfare due to aggregation

Models that lack a variety/extensive margin, including most non-homothetic flexible functional forms, will miss welfare gains by assumption. Changes in variety within a consumption category will be treated as a shift in tastes (change in expenditures given price and income) or an error in demand estimation, with no implications for welfare.

Homothetic models with an explicit extensive margin can also underestimate or miss welfare welfare effects when aggregate and average variety do not coincide. Suppose variety is measured as in figure 10 as the aggregate set consumed in a country, $\bar{n}$. A fall in fixed costs in my model increases variety consumption for all households (like $x_1$) except those constrained at $\bar{n}$ (like $x_2$). *Average* variety consumption increases but there is no effect on *aggregate* variety consumption. The variety Engel curve shifts up and the cost-of-living falls for all households (up to $\bar{n}$) but because the aggregate measure of variety is unchanged, methods that use aggregate variety to calculate welfare effects will find no gains.

This could occur in a trade model if a country consisted of heterogeneous agents at varying points along their variety Engel curves, each paying within country fixed costs. A fall in the consumer fixed costs would be lost in aggregation, as gains in average variety for unconstrained households are ignored.[12]

The Feenstra (1994) index calculates variety gains as a function of aggregate expenditure shares on common and non-common varieties (see Appendix B). If all varieties are common at the aggregate level there is no welfare gain from variety. By restricting the common set to a single, universally consumed variety, changes in average non-common variety could show up in the taste/quality parameter of the model through their effect on expenditure shares relative to the common variety. This would be reflected as a welfare gain, but the gain could be overstated or overstated as it ignores income effects on average variety and omits the effect of higher variety costs.

[12] In general equilibrium the aggregate set would generally rise as the richest households demand more varieties, but the welfare gains from the increase in the aggregate set still understate the gain in welfare from higher average variety. Households below the original constraint before and after the fall in fixed costs have a welfare gain unrelated to the increase in the aggregate set.

### 2.4.2. Overstatement/understatement of welfare due to endogenous variety

Homothetic models of variety lead to biased estimates of welfare gains when variety Engel curves exist and they assume variety differences are exogenous. In a variety Engel curve model there is no welfare gain to a poor household/country from consuming the set of varieties of a rich one (holding prices and fixed costs constant). This is not true for homothetic cost-of-living indexes that always imply welfare gains from consuming a superset of varieties (e.g.Feenstra (1994)). These imply that rich households/countries would have higher welfare *even if they had the same expenditures*; conversely poor households/countries would have greater welfare if they consumed the larger set of varieties consumed by the rich.

Figure 4 illustrates the bias. Suppose there are two consumers with expenditures ($x_1$) and ($x_2$) and we observe that the consumer with higher expenditures consumes more variety $n^2 > n^1$. A homothetic welfare measure interprets the difference in variety as exogenous differences a variety constraint ($\bar{n}$). The welfare gain from variety to the rich household is shaded area $A + B + C$ while the hypothetical welfare gain to the poor household would be shaded area C. However, if the difference in variety consumption were in fact due to movement along the variety Engel curve, the rich household would only gain shaded area A, as areas B and C lie under the marginal cost curve. Were the poor household to consume $n^2$ it would experience a welfare loss equal to area B.

Note that understatement of welfare gains from variety could also occur for the same reason. Suppose expenditures decline but variety remains the same. The variety Engel curve model implies that either marginal benefit must have increased or marginal cost must have decreased, so welfare is higher. Under exogenous variety there can be no welfare gain from variety when variety is constant, leading to understatement when the variety Engel curve model is valid.

Even when controlling for expenditures, neglect of the cost of variety leads to slight overstatement of welfare gains. For example, an exogenous doubling of variety lowers the cost-of-living by $\psi$ percent for CES models, while in a variety Engel curve model an equivalent increase in variety due to lower $F$ lowers the cost-of-living by $\psi - \frac{\psi}{\epsilon}$. This is essentially the difference between the shaded area $A+B+C$ in figure 4 and the shaded area $A+B$ in figure 8 (if it was a parallel shift in the MC curve caused by a change in F, rather than the change in $\epsilon$ depicted). When variety Engel curve slopes are relatively flat ($\epsilon$ large relative to $\psi$) this effect is small, so in most cases the divergence in welfare gains will be due to failure to distinguish income-driven variety differences from exogenous ones. In general, we would expect the difference in welfare gains between these models to be greater in the cross-section than over time, as income differences over short periods are small compared to cross-sectional income differences.

### 2.4.3. Distribution of welfare from variety

If the aggregate set of varieties in the economy expands due to an increase in the exogenous upper bound, which only binds on rich households, only the rich benefit. The Feenstra (1994) model interprets this as a decrease in the cost-of-living for all households, albeit a small one if the aggregate expenditure share of the rich households is small. If instead variety growth occurs through a decrease in the cost of consuming low index varieties, there could be large aggregate gains from variety but they would be poor-biased households. In both of these cases the distributional impact of variety could be very large but would be ignored by homothetic models. These models assume a flat variety Engel curve slope but the distributional impact of variety operates through this slope, with flatter slopes implying lower welfare inequality and steeper slopes greater welfare inequality.

Broda and Romalis (2010) divide American households into expenditure deciles and treat each decile as a homothetic representative agent with an exogenous variety set. They do not compare welfare from variety across deciles, but instead compare changes over time in prices and variety for each decile. By not explicitly accounting for variety Engel curves these cost-of-living measurements will be biased. Upper expenditure deciles that experienced more income growth experienced more movement along their variety Engel curves relative to exogenous changes from fixed costs or prices. Substitution away from more expensive varieties on the extensive margin would also counteract the negative impact of higher inflation for luxury varieties. If these two effects offset we would tend to overestimate the rise in real inequality. Without a variety Engel curve there is no theoretically consistent way to construct a cost-of-living index based on utility, prices, and variety effects that allows statements about nominal versus welfare inequality.

# 3. Variety in India

In this section I examine several implications of the model in light of India food consumption data, including the importance of movement along variety Engel curves, the role of relative prices and other factors in shifting variety Engel curves, and the relation between "availability" of varieties and transaction costs.

## 3.1. National Sample Survey Data

India's National Sample Survey (NSS) collects household consumption data using an interview with a 30-day recall period.[13] Like the Consumer Expenditure Survey (CEX) in the United States, a central purpose of the NSS is to compute the expenditure shares used to weight prices in a consumer price index. The number of households surveyed varies by round, but the "thick" survey rounds used in this paper - 38th (1983), 43rd (1987-88), 50th (1993-1994), 55th (1999-2000), and 61st (2004-2005) - collect data for over 100,000 households. The survey is designed to be representative for all of India's states and the rural and urban sectors. [14] I use sampling weights whenever aggregating and restrict attention to the 17 largest states and New Delhi. The lowest geographic units are first-stage strata - villages or urban blocks with 10 sampled households - but the lowest that can be mapped across years or merged with other datasets are districts.[15]

The number of different expenditure categories varies over time as the National Sample Survey Organization (NSSO) adds new goods to its list or combines goods into a single category. The general trend for food has been to consolidate the number of categories over time, which could lead to downward biased variety growth. I aggregate across categories to form a consistent set of 134 food varieties. Table 2 lists the food varieties organized into nine groups based on survey headings - grains, pulses,

[13]The 55th survey round used a 7-day recall period in addition to the 30-day period, potentially biasing up consumption estimates and measured poverty reduction. See Deaton and Kozel (2005) for an overview. The 61st round returned to a consistent reporting period and I only use data for the 55th round when total expenditure and total food expenditure are not being compared over time (e.g. calculating elasticities). Omitting this round does little to affect the results.

[14]The NSS surveys follow the Indian census and defines urban areas as (a) all statutory places with a municipality, corporation, cantonment board or notified town area committee, etc. or (b) a place satisfying the following three criteria simultaneously: (i)a minimum population of 5,000; (ii) over 75 per cent of male working population engaged in non-agriculture; and (iii) population density over 400 per sq. km.

[15]There is no district data for 1983 or 1993-94.

dairy, meat, oil, vegetables, fruits and nuts, sugar/spices, beverages/processed food. Most of the food goods are relatively homogeneous but those in the beverage and processed food category like "cold beverages," "sauces," or "cooked meals" are probably quite heterogeneous. Each group features at least one heterogeneous "other" category.

The survey records quantity data for most of the goods in the food category, imputing values of home-produced goods and gifts at ex-farm and local retail prices. Unit values can be calculated as expenditure divided by quantity. While quality effects and unobserved composition imply that unit values are different than the prices of homogeneous goods, Deaton and Tarozzi (2005) and others have argued that these quality effects are small can be used to compare the cost-of-living across states, between rural and urban areas, and over time (Deaton (2008), Deaton and Kozel (2005), Deaton and Tarozzi (2005)).[16] When interpreting unit values as prices I use medians within an area/period.

## 3.2. Variety growth

The top panel of table 1 documents the rise in average variety and the large gaps between urban and rural households. The extent of variety increase varies considerably across the nine food groups. The bottom panel documents large differences in real expenditures. I construct real expenditures by deflating nominal expenditures by a Tornqvist price index with 1987-1988 rural India as the base, using median unit values as prices. I do this separately for each food group but for total expenditures I use the set of all goods with unit values (67% to 83% of aggregate expenditure).

Figure 11(a) presents non-parametric variety Engel curves for all varieties (including non-food), restricted to five person households and the middle 98% of the real expenditure distribution. Figure 11(b) presents a similar plot for food expenditures and food varieties. The variety Engel curves are roughly log-linear and have significant upward shifts over time and for urban relative to rural areas. Average food variety in 2004-2005 is 30-40% higher than in 1983 conditional on real expenditure and about 15-20% higher in urban areas in any given year. The circles plot median variety and real expenditures for each sector/period, and indicate that comparing average variety without controlling for real expenditure levels (movement along the

[16]See Deaton and Tarozzi (2005) for a discussion of the advantages and disadvantages of household surveys as a source of price data compared to official government statistics.

variety Engel curve) would lead to overstated welfare gains over time for all varieties and overstated gains for urban versus rural households for food.

### 3.2.1. Which varieties?

Figure 11(b) hides significant heterogeneity across groups and distributional effects. Figure 12 compares variety Engel curves by plotting log variety on log real expenditure for the ten food groups. There are significant gains for urban households, and for rural households over time, in grains, vegetables, fruit, sugar/spices, and beverages/processed food, mostly concentrated at the top. There appears to be a slight decrease in variety for oil and milk groups.

Instead of looking at number of food varieties per household, we can look at the number of households consuming each good. Panel A of figure 13 plots the share of rural households consuming each variety in 1983 and 2004-2005. Most of the points lie above the 45 degree line indicating broad variety growth.[17] Panel B documents a similar pattern for rural and urban households in 2004-2005.

The variety Engel curve model implies that the intensive margin (share of budget) and extensive margin (share of households consuming) margins are positively correlated, as variety growth is driven by diminishing marginal utility from quantity. I calculate the mean group budget share (conditional on purchase) and the share of households consuming for each state/sector in 2004-2005. Figure 14 plots the relationship. The slope is significantly greater than zero for all groups except meat, but there are varieties that are very important to the few households that consume them.

### 3.2.2. Relative prices

How much of the increase in variety can be explained by relative prices? Figure 15 plots the change in share of households consuming over 1983-2005 against the ratio of prices in 2005 over prices in 1983 for each variety/state/sector. Surprisingly this relationship is positive and significant at the 5% level for 7 of the 9 groups though the $R^2$ is low and there is significant dispersion. That prices rose faster for varieties with more extensive margin growth is at odds with models that only feature variety gains through declining relative prices. While relative prices (and hence marginal

[17] The main exceptions below the 45 degree line are coarse grains, curd (buttermilk) and ghee, and groundnut oil. See Deaton and Dreze (2009) for a discussion of buttermilk and coarse grains, and Finnis (2007) for coarse grains.

benefit) play a role in the extensive margin, other factors like increased expenditures (for groups with rising real expenditures like milk, oil, fruit,vegetables and processed food) and changes in fixed costs/availability are central to observed patterns of variety growth.

I also examine whether the Pareto distribution adopted earlier is a reasonable approximation for relative expenditure shares. I order varieties by expenditure within each group for each household and regress the log expenditure on log rank for each state/sector, controlling for household fixed effects. I test the null that this relationship is linear against a quadratic alternative with log rank squared. For three of the ten groups the share of state/sectors with a significant log rank squared term (at the 5% level) is below 25%. For the others I reject the null that there is no quadratic term for over 60% of state/sector combinations. The large sample size makes it easy to reject linearity, but the change in adjusted $R^2$ is small in all cases.

### 3.2.3. Availability

One way to measure availability of variety is to use the total number of varieties consumed at different levels of geographic aggregation. Table 3 reports the count of varieties consumed at the state, region, district, village/block and household levels for 1987-1988 and 2004-2005. There has been no growth in the aggregate measure of availability at the state or region level but modest growth at the district level and below.

There are several reasons for analyzing fixed costs as the source of variety differences rather than exogenous limits based on aggregate availability measures. First, measures of availability based on aggregate varieties are affected by sampling and are endogenous to expenditures, as discussed later and shown in 8. Second, even if we had perfect data on all varieties sold in every store, it is not obvious what is the right aggregate set to use when calculating the welfare gains from variety. In terms of table 3 the right measure of "availability" is somewhere between the varieties consumed by the household and the total set of varieties sold in the entire world. Third, even if we assumed that all varieties consumed in a village in our sample were available without cost, we would need some transaction cost or friction to explain why typically (a)the richest household consumes less than the full set of varieties and (b)poor households consume less varieties than rich ones.

In light of these considerations I ignore exogenous variety limits in the subsequent analysis and model variety differences that are not due to marginal benefit (relative prices) or expenditures as due to differences in the transaction costs/fixed costs. Note that exogenous variety limits effectively represent a vertical kink in the marginal cost curve and horizontal kink in the variety Engel curve (as in figure 9). As I do not model them explicitly, they will show up as a flatter slope for the linear variety Engel curves I estimate hence a higher value for the marginal cost of variety parameter ($\epsilon$)..

# 4. Estimation

Calculating the size and distribution of welfare gains from variety with my model requires four parameters: $\sigma$, $\theta$, $\epsilon$ and F. I first describe estimation of these parameters and then use the model to analyze welfare effects over the 1983-2005 period and between rural and urban households in 2004-2005.

## 4.1. Elasticity of substitution - $\sigma$

The $\sigma$ parameter in a CES model is equivalent to the own-price elasticity in more general models and captures the size of welfare gains from variety. I use the method of Feenstra (1994) which achieves identification through functional form and heteroskedasticity in a panel. As identification in this case relies on functional form assumptions and aggregate data, I also consider an alternative own-price elasticity estimate at the household level using a flexible functional form and spatial variation in prices based on Deaton (1988).

Feenstra (1994) begins with CES demand and considers demand (in share form) for variety relative to a base variety k:

$$\ln S_i/S_k = (1-\sigma)\ln p_i/p_k + e_i^k \tag{12}$$

These relative shares are differenced over time, yielding error term $e_i^{kt} = \Delta_t e_{it}^k = \Delta_t(\ln S_{it}/S_{kt}) + (\sigma - 1)\Delta_t(\ln p_{it}/\ln p_{kt})$. Feenstra (1994) allows for a non-horizontal supply-curve through relative supply equation $\delta_i^{kt} = -\frac{p}{1+p}\Delta_t(\ln S_{it}/S_{k,t}) + \Delta_t(\ln p_{it}p_{kt})$. The demand elasticity $\sigma$ can be identified if the relative demand and supply shocks $e_i^{kt}$ and $\delta_i^{kt}$ satisfy the independence condition $E_t(e_i^{kt}\delta_i^{kt}) = 0$. By multiplying the shocks together we get:

$$Y_i^{kt} = \theta_1 X_{1i}^{kt} + \theta_2 X_{2i}^{kt} + u_i^{kt} \tag{13}$$

where $Y_i^{kt} = (\Delta_t \ln p_{it}/p_{kt})^2$, $X_{1i}^{kt} = (\Delta_t \ln s_{it}/s_{kt})^2$, and $X_i^{kt} = (\Delta_t \ln s_{it}/s_{k,t})(\Delta_t \ln p_{it} - p_{kt})$ and $u_i^{kt} = e_i^{kt}\delta_i^{kt}$. $u_i^{kt} = e_i^{kt}\delta_i^{kt}$ is correlated with the regressors, but by averaging over time the independence condition implies we can use $E_t(\bar{u_i^{kt}}) = 0$ for identification. The time-averages $\bar{Y_i^k}, \bar{X_{1i}^k}, \bar{X_{2i}^k}$ are the second moments of the changes in prices and expenditure shares. As T approaches infinity, under weak conditions $plim(\bar{u}_i) = 0$ so the error $\bar{u}_i$ vanishes. The estimates of $\theta_1$ and $\theta_2$ obtained by OLS are used to solve for the demand elasticity.[18]

In addition to the independence of error terms and functional form assumptions, this method requires at least three varieties for identification since there are two parameters in equation 13 and only (n-1) observations after time-averaging and differencing with respect to a base variety. Consistency requires that the regressors are not collinear, so the true demand and supply variances must differ across varieties.

The Feenstra (1994) approach is consistent with the CES part of my parameterized model except for an aggregation issue. Let expenditures on variety i be given by $p_i q_i = X v_i \left(\frac{p_i}{P(n)}\right)^{1-\sigma}$ with $v_i$ a common, variety-specific taste shock. The log aggregate expenditure share of i is

$$\ln S_i = (1-\sigma)\ln p_i + \ln\left(p^{\sigma-1}\nu^{(1-\sigma)\psi}\right) + \underbrace{\frac{\int_{X_i}^{X_{max}} X^{1+\frac{\psi(1-\sigma)}{\epsilon-\psi}} dF_X}{\int_{X_{min}}^{X_{max}} X dF_X}}_{aggregation\ term} + \ln v_i \qquad (14)$$

using the formulas earlier ($\nu = \left(\frac{\psi}{F\epsilon}\right)^{\frac{1}{\epsilon-\psi}}$). The minimum expenditure level of a household purchasing variety i is $X_i$. $F_X$ is the CDF of expenditures with support $[X_{min}, X_{max}]$. In Feenstra (1994) measurement error in prices and correlation of taste shocks and prices (due to non-horizontal supply curves) imply that standard OLS on equation 12 is biased. In my model the error term $e_i^k$ also contains the aggregation term that depends on minimum cutoffs $X_i$ and $X_k$ and the distribution of expenditure. If $X_k$ and $X_i$ are below $X_{min}$ then these terms are equal and the aggregation term drops out of equation 12. If this is not the case, shocks to the distribution of expenditures ($F(X)$) and fixed costs ($F$,$\epsilon$) can also affect relative shares.

To the extent that these effects operate like demand shocks given prices, the Feenstra (1994) approach should still provide valid estimates of $\sigma$. I use the most popular

[18]See Feenstra (1994) for details such as the formulas for obtaining $\sigma$ and $\rho$ from the $\theta_1$ and $\theta_2$ parameters, as well as the extension by Broda and Weinstein (2006) that uses quantity weights.

variety as the base (k) and combine the five survey rounds, each divided into 62 regions and 4 quarters. I difference across quarters within a survey round and treat each region as equivalent to another time-period, which means that T is as high as 930 for computing the (n-1) average variances. Following Broda and Weinstein (2006) I weight each variety by the number of periods T but multiply this by the share of households consuming rather than quantity. Measurement error in prices (median unit values) is likely to be related to the number of sample households rather than quantity. The size of demand shocks from changing reservation expenditures or the expenditure distribution is likely to be smaller for widely consumed varieties, with an aggregate elasticity closer to the average elasticity. If the true own-price elasticities are different across varieties, putting more weight on widely consumed varieties gets closer to the average elasticity facing consumers.[19]

As a check on the validity of the Feenstra (1994) methodology I use an alternative framework for estimating price elasticities from Deaton (1988). An almost ideal demand system is estimated using a share equation for each variety i, household h, and cluster c:

$$w^i_{hc} = \alpha_i + \beta^i \ln X_{hc} + \gamma^i z_{hc} + \sum_{j=1}^{n} \theta^i_j \ln p_{jc} + f^i_c + u^i_{ch} \tag{15}$$

where w is budget share, X is expenditure, z is a vector of household controls, p is the cluster price and f is a cluster fixed effect. The (average) own-price elasticity in this system is $e_i = \frac{\theta^i_i}{\bar{w}_i}$ where $\bar{w}_i$ is the sample average budget share. Clusters are defined as areas with constant prices. Estimation proceeds in two stages, with the first stage using cluster fixed effects to estimate the $\beta$ and $\gamma$ parameters. The second stage regresses the cluster average of $y^i_c = w^i_{hc} - \beta^i \ln X_{hc} + \gamma^i z_{hc}$ on prices across clusters, generating estimates of $\theta$ used to compute price elasticities.

I implement the corrections for quality effects on unit values and first-stage measurement error in Deaton (1988), treating villages and urban blocks as clusters and using region/quarter fixed effects. Identification of price elasticities relies on the assumption that prices are constant within clusters, that prices vary across clusters within a region/quarter and that they are independent of taste or demand shocks tastes within a region/quarter. If a price is missing from a cluster it is imputed from the median region/quarter median to avoid dropping many clusters. I use

[19]Using only T to weight varieties leads to slightly higher elasticity estimates for some varieties and lower estimates for others, with overall welfare effects similar in magnitude.

household size and male and female adult ratios as additional controls. I assume group-separability and estimate the demand system for each group separately using group expenditures as the X variable. Own-price elasticities are then aggregated to the group level using aggregate expenditure shares.

## 4.2. Price/taste hierarchy - $\psi$

The parameter $\psi \equiv \frac{1}{\sigma-1} - \frac{1}{\theta}$ governs the welfare gains from variety, and reflects both the elasticity of substitution and the degree of asymmetry across varieties captured by $\theta$.[20] I use a simple procedure to estimate $\psi$ by constructing the Feenstra (1994) index for each household relative to a household consuming just the base variety. This index is given by:

$$P_g^F \equiv \left(\frac{x_{1g}}{X_g}\right)^{\frac{1}{\sigma_g - 1}} \tag{16}$$

and depends only on group expenditure shares ($X_g$), expenditures on the base variety ($x_{1h}$) and $\sigma$. The index equals one for households consuming only the base variety and is lower for households consuming additional varieties.

The index $P_g^F$ is equivalent to $P(n) = pn^{-\psi}$ in the continuous CES case. I use OLS to estimate:

$$\ln P_{ghc}^F(n) = \gamma_c - \psi_g \ln n_{ghc} + u_{ghc} \tag{17}$$

where $\gamma_c$ is an area/period fixed effect and $n_{ghc}$ is the number of varieties consumed by the households. The CES demand structure implies that $\psi_g$ is positive, with a slope that depends on the extent of diminishing returns to quantity ($\sigma$) and the asymmetry across varieties relative to the base. If the true relationship is non-linear in n, my procedure puts greater weight on the local marginal benefit of varieties around the sample median n. For each state/sector I use the most widely consumed variety as the base variety and restrict the sample to households for which it has the highest budget share. [21]

[20]If we define the CES part of utility as $\int_0^n d_i^{\frac{1}{\sigma}} q_i^{\frac{\sigma-1}{\sigma}} di$ and parameterize the taste $d_i = i^{\frac{1-\sigma}{\theta_2}}$ and price $p_i = zi^{\frac{1}{\theta_1}}$ then the $\theta$ in the formula $\psi \equiv \frac{1}{\sigma-1} - \frac{1}{\theta}$ is given by $\theta = \frac{\theta_1+\theta_2}{\theta_1\theta_2}$. The marginal benefit is then a function of both the asymmetry in prices and in tastes/quality across the different varieties, and the lowest index varieties (with the highest expenditure share) need not be the cheapest. For example, rice could be more expensive than some of the coarse grains, but tastes so much better that it is still the most important variety for both the intensive and extensive consumption margins.

[21]Changing tastes for marginal varieties are allowed in this model provided they represent a real increase in welfare (and are not offset by a decrease in utility from the base variety). This is analogous

## 4.3. Variety Engel curves - $\epsilon$ and F

Fixed costs can be estimated using the variety Engel curve equation 7:

$$\ln n_{gh} = \omega_g + \beta_g \ln x_{gh} + u_h \tag{18}$$

The slope of the Engel curve $\beta_g$ corresponds to $\frac{1}{\epsilon_g - \psi_g}$ in the model, so $\epsilon_g$ can be calculated given an estimate of $\psi_g$. If $\psi_g$ is high, indicating a large benefit from variety, but the Engel curve is flat, then $\epsilon$ must be high - fixed costs rise quickly to choke off the larger marginal benefit to richer households. The variety Engel curve equation gives the level of fixed costs $F_g$ as a function of the intercept $\omega_g$ and other parameters:

$$\ln F_g = \ln \frac{\psi_g}{\epsilon_g} - (\epsilon_g - \psi_g)\omega_g - \ln p \tag{19}$$

I estimate variety Engel curves by OLS, restricting the sample to households that consume the base variety. The final parameter required to back out $F_g$ is the price level term $p$, for which I use a Tornqvist index over all common varieties with aggregate share weights.[22]

## 4.4. Welfare measurement

Table 4 presents some of the estimated model parameters. Groups like grains and beverages/processed food have very low elasticities of substitution while other like milk and oil have high elasticities, but the high estimates are not precisely estimated. The estimates based on Deaton (1988) are lower for most groups and generally fall within the 95% confidence interval of the other estimates. Sugar is the most notable exception, with an elasticity below one for the alternate method. I focus on the Feenstra (1994) estimates for comparability with conventional CES welfare measures but consider the alternate elasticities as a robustness check. The average $\psi$ parameter rose

to the assumption in the Feenstra (1994) model that taste for the base set of varieties is constant, which allows rising relative expenditures on varieties outside the base set due to either lower relative prices, higher quality, or higher tastes to lower the cost of living.

[22]Using the price of the base variety has little impact on the results. Technically the price of the base good is $p_1 = \frac{\int_0^1 p_i q_i di}{\int_0^1 q_i di} = p[\psi(\sigma-1)]^{\frac{1}{1-\sigma}} \frac{1-\frac{\sigma}{\theta_1}}{1-\frac{\sigma-1}{\theta_1}} di$ which depends both on relative taste/quality (through $\theta_2$) and relative prices ($\theta_1$). As $\theta_1$ is not observed an exact derivation of $p$ would require additional information on $\theta_1$ or $\theta_2$, rather than the $\theta$ parameter that combines both and can be derived from my estimate of $\psi$.

for two thirds of the groups, though only slightly in most cases. Greater marginal benefit thus plays some role in variety growth, but as the case of grains makes clear it is insufficient to explain shifts in variety Engel curves, and for most groups there is a large decline in fixed costs ($\epsilon$ and $F$).

I compute four different measures of welfare gains from variety, expressed as a percentage reduction in the cost-of-living (COL) index. The first measure uses aggregate variety, pooling all households in a state/sector (when comparing across rounds) or a sector (comparing within a state) and calculating the Feenstra (1994) index (see appendix B). This measure will lead to underestimation of welfare gains from variety when there is a large increase in average variety with little change in aggregate variety.

The second measure uses average household variety and evaluates welfare gains using the formula $1-\left(\frac{\bar{n_g^a}}{\bar{n_g^b}}\right)^{-\psi_g}$, with $\psi$ estimated by pooling across periods or sectors (but allowed to vary across states). Groups are aggregated using Sato-Vartia group expenditure share weights similar to the first measure. If period $a$ has higher variety than period $b$ the cost-of-living will be lower if $\psi$ is positive. This measure will overstate welfare gains when expenditures increase and average variety is higher due to movement along the variety Engel curve, and because of variety costs.

The third measure is the "level" welfare effect from the variety Engel curve model, calculated by estimating a common $\psi$ and $\epsilon$ across periods and forcing $F$ to capture the average distance between variety Engel curves.[23] This imposes constant welfare gains (as percentage of expenditure) across the range of group subutilities. The group level variety welfare gains can be expressed as:

$$1-\left(\frac{F_g^a}{F_g^b}\right)^{\eta_g} \tag{20}$$

with $\eta_g=\frac{\psi_g}{\epsilon_g}$. I also aggregate the gains across groups using the Sato-Vartia index, which ignores differences in group budget shares across the expenditure distribution. This focuses attention on the role of non-homotheticity within groups on measurement of average welfare gains.

The fourth measure captures the distributional effects of variety Engel curves on welfare. For this measure I estimate $\psi$ and $\epsilon$ separately across rounds/sectors, allow-

[23]To the extent that the curves are not parallel, this procedure will put more weight on households near the median expenditures.

ing the variety Engel curve slope to change. Welfare gains at the group level are then indexed to particular group utility levels (corresponding to expenditures in the base period). I focus on changes in fixed costs and evaluate the welfare gains using $\psi_g$ from the base period, so the welfare gain is for a household with the same expenditures, tastes, and relative prices as the base period that experiences the variety costs of the comparison period.[24] The group level welfare gains can be expressed as

$$\frac{X_g^a(U_g^0, F_g^a, \epsilon_g^a, \psi_g^a)}{X_g^b(U_g^0, F_g^b, \epsilon_g^b, \psi_g^a)} - 1 \tag{21}$$

I use two different specifications for aggregation and indexation of group utility levels $U_g^0$. The first assumes across-group homotheticity and uses aggregate group expenditure shares, indexing group utility to the implied group expenditure of households at the 10th, 50th, and 90th percentiles of the food expenditure distribution. Differential welfare gains in this case only come from non-homotheticity within groups. The second specification estimates group demand as $w_g = \alpha_g + \beta_g \ln(X)$ (where X is food expenditure) and defines group share-weights and subutility levels according to the implied group expenditures of households at the 10th, 50th, and 90th percentiles of the food expenditure distribution. The use of expenditure-dependent share weights allows differential welfare gains if the gains within a group are uniform.

Although in either specification the assumption of across-group separability ensures the validity of group-level welfare gains, total welfare gains depend heavily on the structure of across-group demand. Although I compare welfare gains assuming prices are unchanged, households can adjust budget shares in response to differential variety welfare gains across groups so welfare gains are still under-stated. More generally the welfare calculations only pertain to food expenditures and assume separability of food and non-food, so total welfare calculations will depend on the price and income elasticities of food versus non-food and the particular model adopted.

[24]Changes in $\psi$ create infra-marginal gains on consumed varieties with no change in variety, so I ignore them for this exercise except for estimating fixed costs. In the data changes in the marginal benefit of variety sometimes increase welfare from variety (beverages/processed food) and sometimes decrease them (grains).

## 4.5. Results

Table 5 presents the average welfare gains (across 35 state/sectors) between 1983-2005. As expected we find almost no welfare gains when using the aggregate Feenstra (1994) index since the set of overlapping varieties at the state/sector level is almost complete. This is an underestimation of welfare gains result. Conversely, welfare gains calculated using average variety consumption are large, over 10% of food expenditures. The gains estimated using my model are only 1% smaller. The reason for this is that there was little change in overall food expenditure over this period (see table 1) so movement along the variety Engel curve is not distorting measurement.[25] However at the group level there are large differences. Accounting for variety Engel curves increases the gains for grains and decreases them for beverages/processed food because while average variety increased for both groups, real expenditures decreased for grains and increased for beverages/processed food. There is significant variation in welfare gains across states and sectors, with urban sectors gaining considerably more than rural sectors. Almost two-thirds of the gains are being driven by the beverages/processed food category due to its very low $\sigma$ (high $\psi$) and the large increase in variety over the period.

The final three columns of 5 present the distribution results. Note that the magnitudes are not strictly comparable to the level results for several reasons, most notably the use of the base period $\psi$. Using aggregate group share weights, households at the 90th percentile of the food expenditure distribution gained 1.9% more from variety than those at the bottom. Rich-biased gains over time exists within all groups but milk, vegetables and sugar/spices. The degree of rich-bias is larger for rural households. This difference is only driven by non-homotheticity *within* groups and ignores differential group shares. The last row show that using group expenditure weights that vary by percentile leads to a significant reversal for urban areas, with poor households experiencing over 5% greater welfare gains, but variety growth in rural areas remains rich-biased by 1%.

Table 6 presents the average welfare gains (across 17 states) for urban relative to rural areas in 2004-2005. The aggregate CES variety measures show no welfare gains, but the average CES measures shows very large gains of 6% for the average

[25]Note that while the gain from extra varieties is smaller in the variety Engel curve model than the CES model with exogenous variety, when variety growth is driven by a fall in fixed costs there is a large welfare gain from lower fixed costs for infra-marginal variety.

state in the sample. The level welfare gains from the variety Engel curve model reduce these gains by over half to 2.3%. This contrasts greatly with the results over time, and the reason is that urban households have significantly higher food expenditures on all groups except grains, implying that the average urban household is further to the right of a variety Engel curve. This effect is particularly large for processed/food beverages, as urban households spend almost twice as much on these varieties. Dropping beverages/processed food only reduces the urban welfare gains slightly from 2.3% to 1.9%. The overall distribution of welfare gains for urban areas is slightly rich-biased (0.3%) as many of the group-biases cancel out. Using group expenditure weights that vary by percentile tends reduces the rich bias further.

Table 7 examines the robustness of these findings, focusing on the average welfare gain and the differential 90th/10th percentile welfare gains (using aggregate or percentile specific group weights). The first column re-estimates the welfare gains using the (generally lower) elasticities calculated using the Deaton (1988) methodology. The gains are about 25% higher over time and over twice as large for urban versus rural areas. Static urban vs. rural gains are still rich biased, while gains over time for urban areas remain poor-biased. The second column uses the upper 95th percentile Feenstra (1994) elasticities and finds lower but still sizeable gains. The third column uses a four group aggregation scheme by combining grains/pulses, milk/meat/oil, vegetables/fruit, sugar/spice/beverages/processed food. All parameters are re-estimated[26] at the four-group level. The gains are a bit smaller but follow a similar pattern. Finally I estimate welfare gains over time between each survey round separately, and for urban versus rural by survey round. Welfare gains are greatest in the earlier and later periods but are relatively stable for urban versus rural areas.[27]

## 5. Fixed costs

The nature of the frictions that limit variety consumption matters for welfare and policy analysis. If differences in variety (conditional on marginal benefit $\psi$) are only

[26] The elasticites and 95% confidence intervals from the Feenstra (1994) estimator are 1.88(1.27,2.49), 17.31 (3.94,31.25), 5.63(4.55,6.70), and 1.59 (0.93,2.25).

[27] Note that the comparisons with the 55th round are likely to be biased by the addition of a 7-day recall period, so the implied losses over 1994-2000 (and large gains over 2000-2005) may be the result of upwardly biased expenditure estimates.

due to taste then the welfare gains I calculate may be overstated or understated. If they are not, then understanding what drives them is key to understanding what policies will benefit consumers. Welfare gains of 2-3% for urban areas and 10% over time are large relative to differences and growth in real food expenditure (see 1) and the effects may be bigger still for non-food varieties. Rural infrastructure, modern retailing, and liberalization of internal and international trade are major issues facing India and other developing countries in the next century, and large welfare gains from variety should be incorporated alongside the usual employment, income, and price effects considered in cost-benefit analysis.

## 5.1. Tastes: Immigrants and Cohort analysis

While it is difficult to conclusively reject heterogeneous tastes as a source of different consumption patterns, I present two pieces of evidence that tastes alone do not explain variety differences.

I first compare the consumption patterns of migrants and non-migrants under the assumption that migrants bring their tastes with them. [28] If urban-rural variety differences are only caused by tastes, rural migrants to urban areas should consume less varieties than their urban counterparts. The 1987-1988 NSS data allows identification of migration status of individuals, defined as a current place of residence different than the the last usual place of residence (where they must have resided at least six months). I define a "rural to urban" migrant household as one whose head migrated from a rural to an urban area and an "urban to rural" household as the reverse. I exclude rural-rural and urban-urban migrants, so the comparison groups are non-migrant urban and rural households.

Figure 16 plots the food variety Engel curves for non-migrants and migrants by sector.[29] Rural to urban migrants behave virtually identically to urban non-migrants. On the other hand, urban to rural migrants only resemble rural non-migrants at lower levels of expenditures. The number of these migrants is small[30] and they may differ from the typical rural household in other ways (e.g. they are less likely to be involved

[28] This follows the work of Atkin (2009) who finds that migrants have similar tastes to households from their location of origin using the same India data.

[29] I trim the 1% tails of the food expenditure distribution, restrict the sample to households with five members, and deflate urban expenditures using a rural-urban food price index for 1987-88.

[30] 1.6% of the sample is urban to rural migrants, compared to 8.6% rural to urban, 8% rural to rural, and 6.0% urban to urban

in agriculture).[31]

I next examine variety consumption over time for a single cohort. If tastes for food variety are formed during childhood and persist over time, they could not lead to a rise in variety over time for a particular cohort. To avoid confounding life-cycle effects with changes over time I first pool the cross-sections between 1983 and 2005. Figure 17(a) presents head of household age dummies from a regression of log food variety on age dummies, log food expenditure, and year fixed effects.[32] For households with no children there is a small decrease in variety over the life-cycle, while those with three children experience a bigger decline (possibly due to the age of the children).

These life-cycle effects are dwarfed by the cohort effects in figure 17(b) which plots variety Engel curves over time for the cohort born between 1953 and 1962 (restricted to households with five members). This cohort experiences large increases in variety consumption similar in magnitude to the general population, and the life-cycle effects actually bias down this increase. While I cannot rule out that this cohort's taste for variety changed over time (and in the opposite direction of the life-cycle effect), the trend in variety consumption is not driven by different tastes across cohorts formed in childhood.

## 5.2. District-level infrastructure and market density

By matching districts in the 1987-88 survey round to data from the World Bank India Agriculture and Climate Dataset I examine the correlation between variety consumption and district characteristics. There are 271 districts that can be linked across 13 Indian states with variables including population density, road density and distance to the coast. From the NSS data I construct the share of households in each district that use electricity as the main source of lighting, another indicator of local infrastructure.

Table 8 presents results of OLS regressions of three dependent variables on a set of district covariates. "Mean variety" is average food variety in the district, "mean resid-

[31]Selection into migration is tough to address though the results are robust to controlling for occupation and already control for expenditures. Using state-specific rural-urban price indexes and/or intra-state migrants gives similar results. Using only recent migrants (those that moved in the last three years) yields similar results though variety consumption of rural to urban migrants drops below that of urban non-migrants below the 25th percentile of the food expenditure distribution.

[32]I restrict the sample to households with one adult male and one adult female. The omitted category is average adult age 21-30, and the dummies are for ages 31-40, 41-50, 51-60, 61-70 and 71-80.

ual variety" is average variety after netting out the effects of variety Engel curves (and is thus closely related to the residual fixed cost), and "aggregate variety" is the total number of varieties consumed in the district. In addition to the district covariates listed above I include a measure of average real food expenditure[33] and the number of sample households in the district.

Population density, road density, and electrification are all positively correlated with mean variety, while distance from the coast is negatively correlated. The effects on residual variety are similar, implying that these variables have effects conditional on real food expenditures. The aggregate variety measures only varies with real food expenditures and the number of sample households in the district, suggesting that sampling plays a significant role. The residual fixed costs in the model that we think are related to market access, density, and general "availability" of varieties are thus correlated with these variables in the cross-section.

## 5.3. Shopping and transport costs

If the fixed costs represent non-financial transaction costs, they should show up in shopping patterns. The model predicts that (1)households with higher expenditures should spend more time shopping, and (2)households that face lower fixed costs should spend more time shopping. Suppose shopping time is proportional to the costs incurred in the model, $n^{\epsilon}F$. As higher expenditures lead to greater n (holding F constant), higher spending households spend more time shopping. If fixed costs $F$ decrease, the direct effect is to lower shopping time but the indirect effect (by increasing n) raises shopping time. With $\epsilon > \psi$ the indirect effect dominates, so households facing lower fixed costs have higher shopping time conditional on expenditures.

I test these predictions with data from the India Time-Use Survey collected by the National Sample Survey Organization. The data cover six states over 1998-1999 and record time-use for each household member aged six or older in 15 minute interval over the preceding 24 hours. The survey also asks respondents for abnormal days in the last week, including market days and weekend activities. I focus on the time-use categories that correspond most closely to the time costs of consuming greater variety - "shopping for goods and non-personal services: capital goods, household appliances, equipment, food and various household supplies" and "travel related to

[33] I use district level Tornqvist price indexes with the aggregate of all districts for comparison shares and prices.

household maintenance, management and shopping." I also examine a measure of non-shopping travel time that includes travel for work or school and may proxy for the remoteness of an area.

Table 9 presents the results from regressions of total weekly time spent on shopping and related travel on household expenditure and size. Panel A uses village/block fixed effects and shows that within narrowly defined geographic areas households with higher expenditures spend more time shopping, as predicted by the model. Panel B adds an urban dummy to see whether residents of urban areas (with lower transaction costs according to the model) with comparable expenditures spend more time shopping. Panel C instead uses a more refined classification from the survey, with villages and towns classified as small, medium or big.[34]

The results indicate that rich households spend more time shopping (conditional on area) and denser markets lead to more shopping time (conditional on expenditure). Time spent on non-shopping travel is considerably lower in the densely populated areas, indicating that unlike shopping, activities like work and school are unavoidable (i.e. have a low elasticity of time use with respect to time cost/remoteness.) As a proxy for transaction costs, shopping time behaves as the model predicts and is likely to be an important transaction cost for generating variety Engel curves and shifts in variety Engel curve slopes.

## 5.4. Caloric requirements

Another interpretation of the fixed costs is related to caloric requirements. Deaton and Dreze (2009) document a large decline in calorie consumption in India over the 1983-2005 period (conditional on total or food expenditures) and a similar gap for urban and rural households. As shifts in calorie Engel curves mirror shifts in food variety Engel curves the connection merits consideration, though shifts in variety Engel curves for non-food imply this cannot be the only factor.

Suppose quantity is measured in calories with $p_i$ the price per calorie of variety i.

[34]For villages in the rural sector, small corresponds to under 400 residents, medium corresponds to 400 to 1200, and large corresponds to over 1200. For towns in the urban sector small corresponds to less than 50,000, medium corresponds to between 50,000 and 200,000, and large corresponds to over 200,000.

Calorie consumption E is

$$E \equiv \int_0^n q_i di = X^{1-\frac{1}{(\epsilon-\psi)\theta}} F^{\frac{1}{(\epsilon-\psi)\theta}} Z \tag{22}$$

with Z a function of the parameters. Conditional on food expenditures calorie consumption is positively correlated with fixed costs F and hence negatively correlated with food variety. Households trade-off cheap calories and the 'taste' benefits of a more varied diet.

Within sector/state/rounds I find that conditional on log food expenditure and household size there is a negative and significant relationship between log food variety and log calories with elasticity 4.7%.[35] After netting out the effects of real food expenditures on calories and variety (using within area Engel curves) I find a larger negative and significant relationship between residual variety and calories across state/sector/rounds ranging from 44% (no demographic controls) to 98% (restricting to two adult households and controlling for household size).

These are not causal estimates, as lower calorie intake could be the cause and/or effect of food variety. Li and Eli (2010) impute caloric requirements from time-use data and show that they can explain all of the urban-rural difference in calories per food expenditure and over half of the changes over time. If we took the residual calorie measure as fully exogenous (i.e. due only to changing caloric requirements) the effects are still small relative to the differences in food variety. Caloric intake has fallen at most 10% for rural households over the 1983-2005 period but food variety increased by 48% for these households. The rural-urban food variety gap in 1983 is 25% but the calorie gap is less than 10%. While the elasticity of variety to calories would have to be bigger to entirely explain these effects, the variety Engel curve provides a framework to evaluate welfare gains from lower caloric requirements. These gains are different than those implied by an Engel equivalence scale, as they would occur even if food expenditure was held constant.

## 6. Conclusion

This paper shows that CES utility can lead to both understatement and overstatement of welfare gains from variety in the presence of variety Engel curves, and provides

[35]See Li and Eli (2010) for the construction of total calories.

a framework for analyzing the distribution of welfare gains across heterogeneous agents. A natural extension is to apply this framework to international trade data. Modeling the extensive margin using importer versus exporter fixed costs has important implications for welfare measurement, especially in the cross-section. In the variety Engel curve model a poor country like India may gain little from consuming the wider set of varieties of a rich country like the United States, depending on how much of the difference is driven by importer/exporter fixed costs, relative prices, and income. The welfare gains over time for the United States found by Broda and Weinstein (2006) may also be overstated given significant growth in income and import quantities and expenditures.

The general equilibrium implications of the model are also worth exploring as there would be a feedback between the income distribution and real inequality through the (now endogenous) upper variety constraint. There has been recent work in this area (e.g.Foellmi et al. (2010)) and my approach provides an alternative way to incorporate non-homothetic demand for variety and its distributional effects. When exogenous upper-limits are the only restriction on variety then by assumption trade and growth will disproportionately benefit the wealthy, but if transaction-costs for more basic varieties play a role then there is potential for poor-biased variety growth. My approach could also be used to study the effect of internal trade liberalization (lower fixed costs) on international trade patterns and the effect of geographical differences in the variety of local amenities and (possibly income-selected) migration patterns.

Comparing the variety Engel curve framework with non-homothetic discrete choice (e.g. Allenby and Rossi (1991)) could be fruitful. Sheu (2010) carries out a similar comparison for homothetic CES and discrete choice and finds that relaxing the IIA property of CES demand can lead to different results. As the source of variety gains in my model is different it would be interesting to compare the models and their implications for welfare measurement.

## References

Allenby, Greg M. and Peter E. Rossi, "Quality Perceptions and Asymmetric Switching Between Brands," *Marketing Science*, 1991, *10(3)*, 185–204.

Almas, Ingvild, "International Income Inequality: Measuring PPP Bias by Estimating Engel Curves for Food," *Luxemburg Income Study Working Paper No. 473*, 2008.

Anderson, Simon P., Andre de Palma, and Jacques-Francois Thisse, *Discrete Choice Theory of Product Differentiation*, MIT Press, 1992.

Arkolakis, Costas, Svetlana Demidova, Peter J. Klenow, and Andres Rodriguez-Clare, "Endogenous Variety and the Gains from Trade," *Working Paper*, 2007.

Atkin, David, "Trade, Tastes and Nutrition in India," *Working Paper*, 2009.

Bils, Mark and Peter J. Klenow, "Quantifying Quality Growth," *American Economic Review*, 2001a, *91*, 1006–1030.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 2006, *121(2)*, 541–585.

_ and _ , "Product Creation and Destruction: Evidence and Price Implications," *American Economic Review*, 2010, *100(3)*, 691–723.

_ and John Romalis, "Welfare Implications of Rising Price Dispersion," *Working Paper*, 2010.

Costa, Dora L., "Estimating Real Income in the US from 1888 to 1994: Correcting CPI Bias Using Engel Curves," *Journal of Political Economy*, 2001, *109(6)*, 1288–1310.

Deaton, Angus, "Quality, Quantity and Spatial Variation in Price," *American Economic Review*, 1988, *78*, 418–430.

_ , "Prices trends in India and their implications for measuring poverty," *Economic and Political Weekly*, 2008.

_ and Alessandro Tarozzi, "Prices and Poverty in India," *Working Paper*, 2005.

_ and Jean Dreze, "Food and Nutrition in India: Facts and Interpretations," *Economic and Political Weekly*, 2009, *44(7)*, 42–65.

_ and Valerie Kozel, "Data and Dogma: The Great Indian Poverty Debate," *World Bank Research Observer*, 2005.

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 1994, *84(1)*, 157–177.

_ , "Measuring the Gains from Trade under Monopolistic Competition," *Working Paper*, 2009.

Finnis, Elizabeth, "The political ecology of dietary transitions: Changing production and consumption patterns in the Kolli Hills, India," *Agriculture and Human Values*, 2007, *24(3)*, 343–353.

Foellmi, Reto, Christian Hepenstrick, and Josef Zweimuller, "Non-homothetic preferences, parallel imports and the extensive margin of international trade," *Working Paper*, 2010.

Hamilton, Bruce W., "Using Engel's Law to Estimate CPI bias," *American Economic Review*, 2001, *91(3)*, 619–630.

Hausman, Jerry, "Sources of Bias and Solutions to Bias in the Consumer Price Index," *Journal of Economic Perspectives*, 2003, *17(1)*, 23–44.

Hummels, David and Vova Ludovsky, "International Pricing in a Generalized Model of Ideal Variety," *Journal of Money, Credit and Banking*, 2009.

Lancaster, Kelvin, *Vatiety, Equity, and Efficiency*, Columbia University Press, 1979.

Li, Nicholas and Shari Eli, "In Search of India's Missing Calories," *Working Paper*, 2010.

Melitz, Marc J. and Gianmarco I.P. Ottaviano, "Market Size, Trade, and Productivity," *Review of Economic Studies*, 2008, *75*, 295–316.

Saure, Philip, "Bounded Love of Variety and Patterns of Trade," *Working Paper*, 2009.

Sheu, Gloria, "Price, Quality, and Variety: Measuring the Gains from Trade in Differentiated Products," *Working Paper*, 2010.

# 7. Tables and Figures

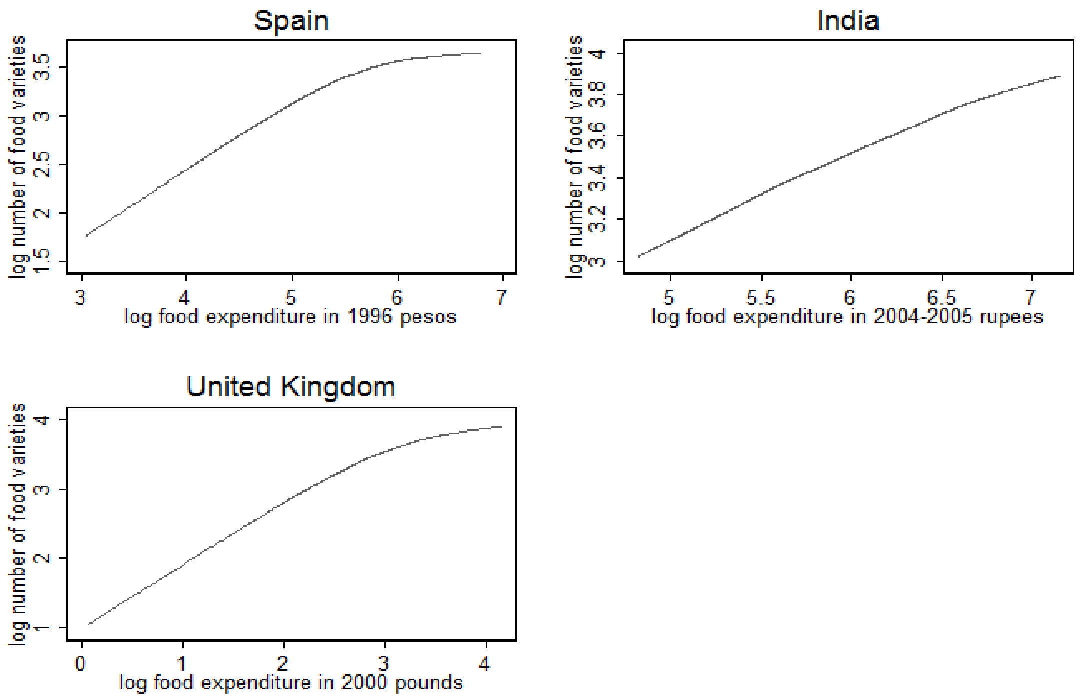



Figure 1: Local linear regression of log (number of food varieties) on log (food expenditure) for households with 3 members.

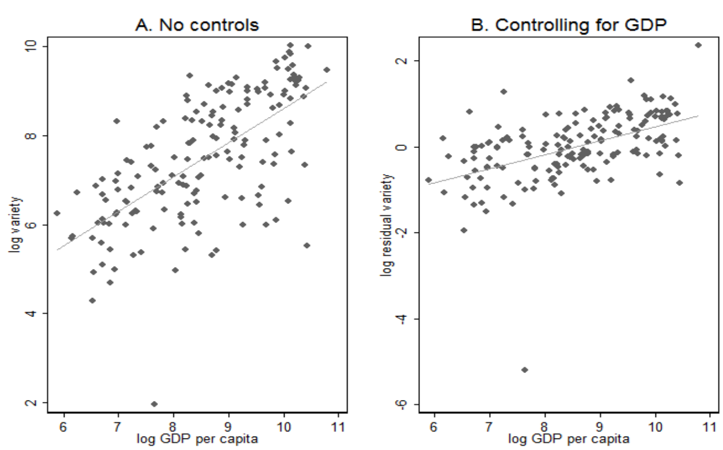



Figure 2: Log number of import categories (4-digit SITC x country) on log income per capita.

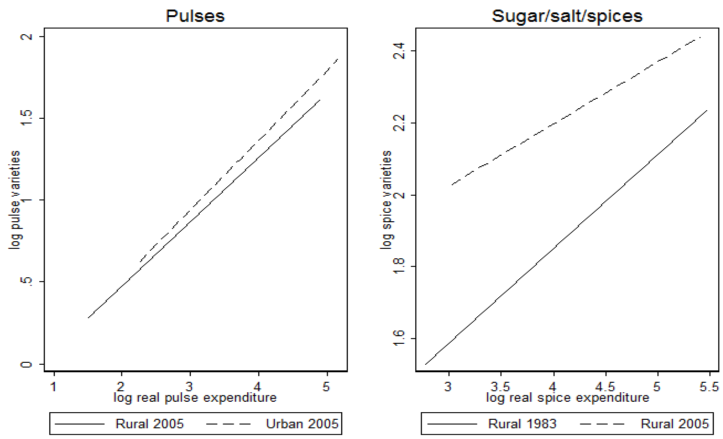



Figure 3: Shifts in variety Engel curves for Karnataka (two adult, three child households)

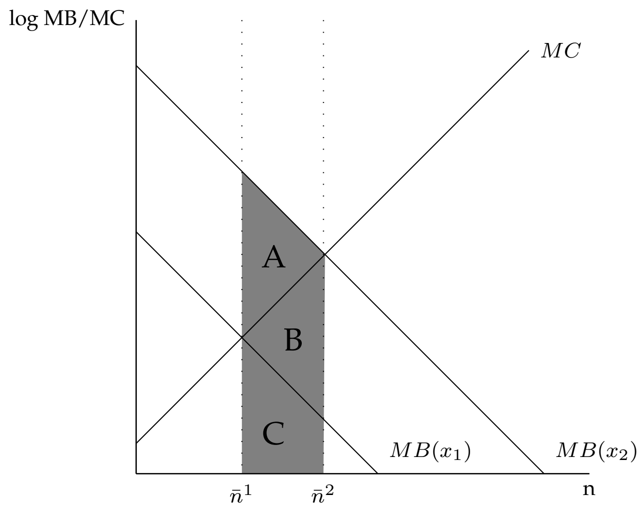



Figure 4: Exogenous versus endogenous variety gains and bias

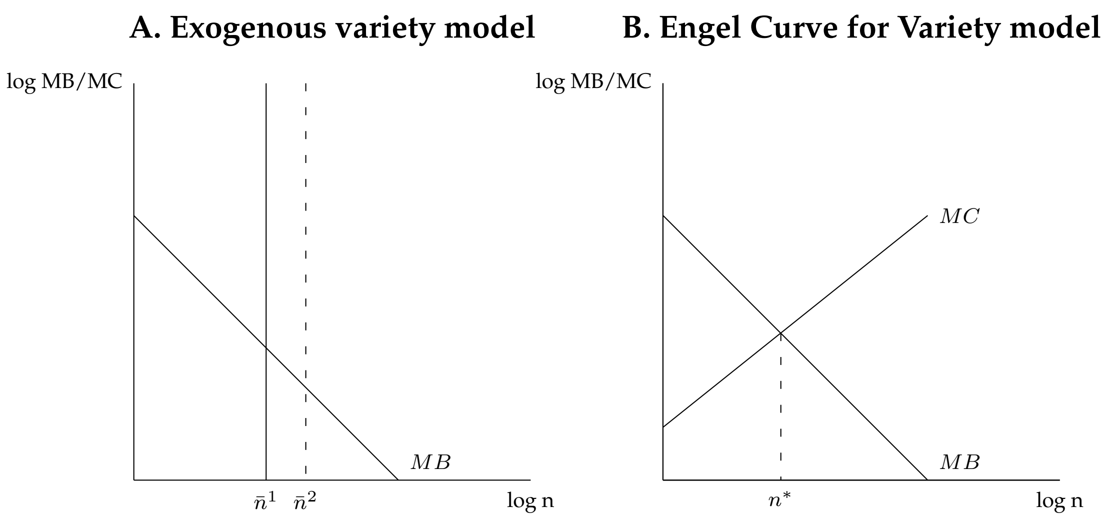



Figure 5: Standard (exogenous variety) model and Engel Curve for Variety model

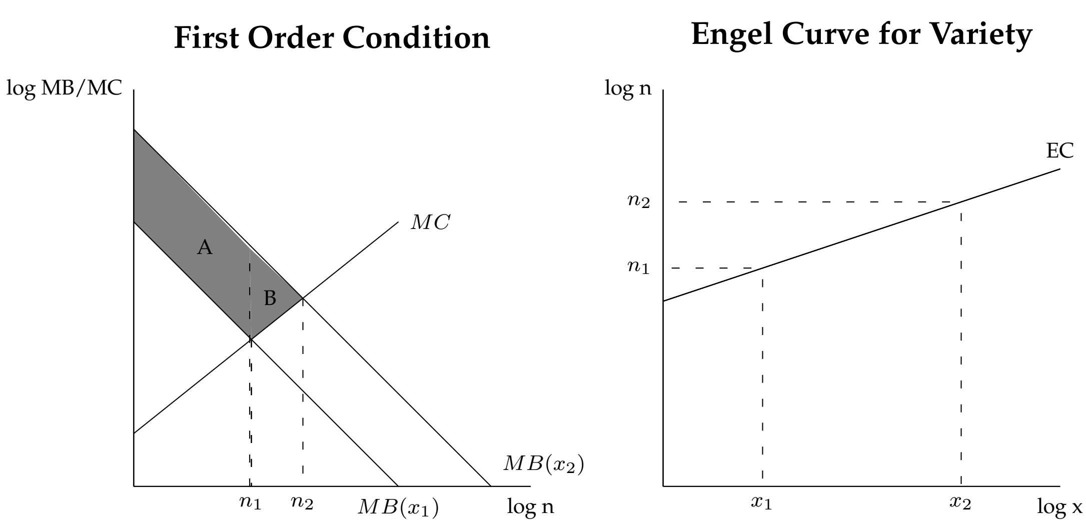



Figure 6: Rise in expenditures

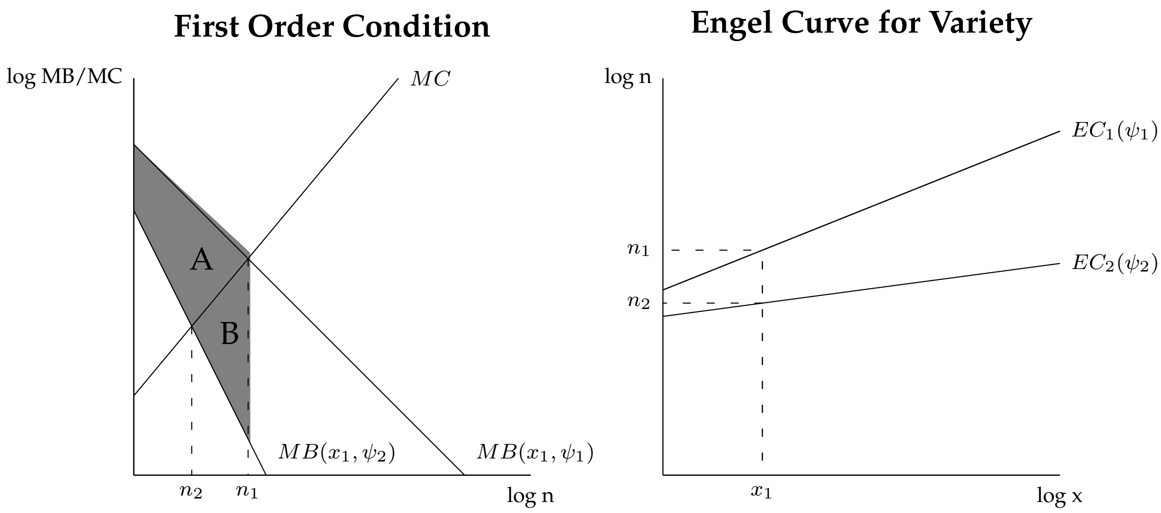



Figure 7: Rise in price slope

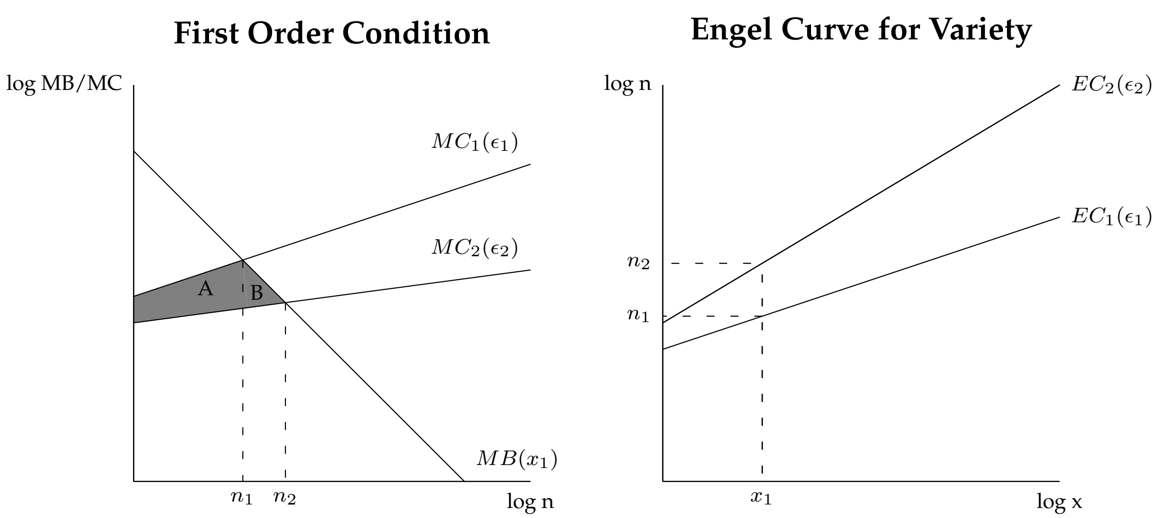



Figure 8: Fall in fixed cost slope

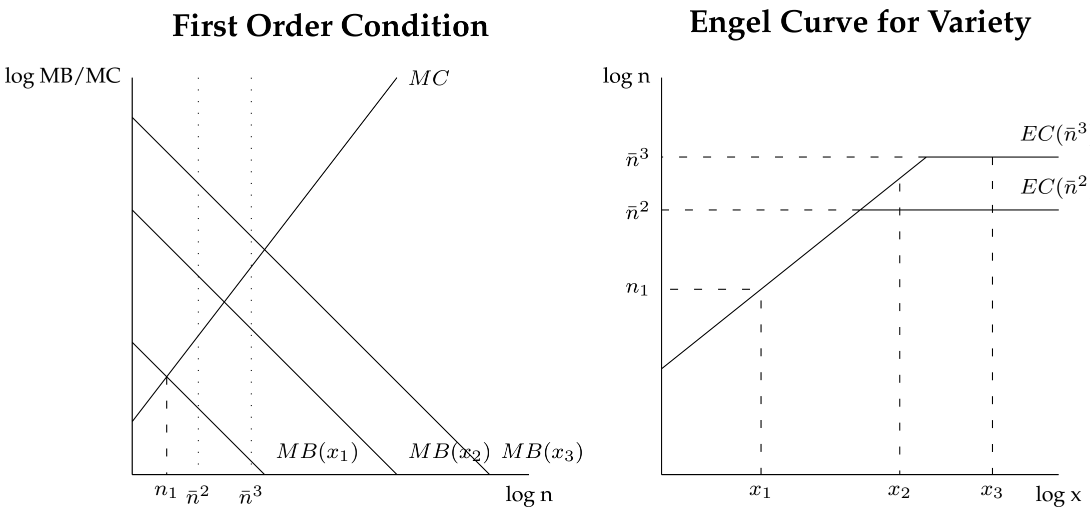



Figure 9: Rise in exogenous variety limit

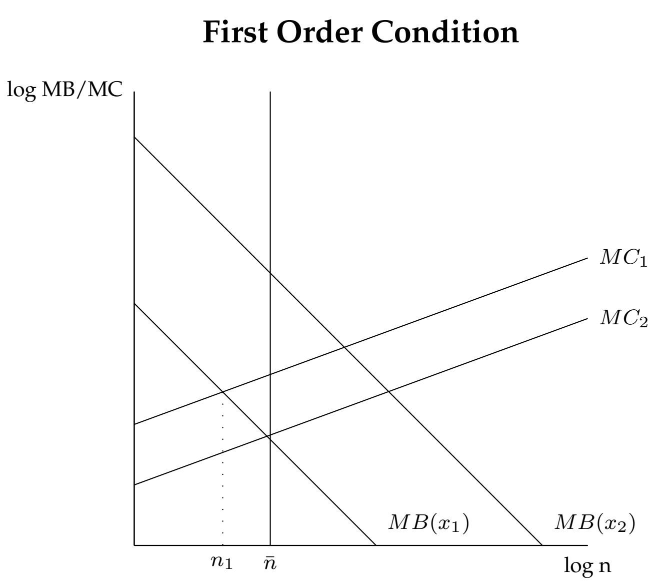
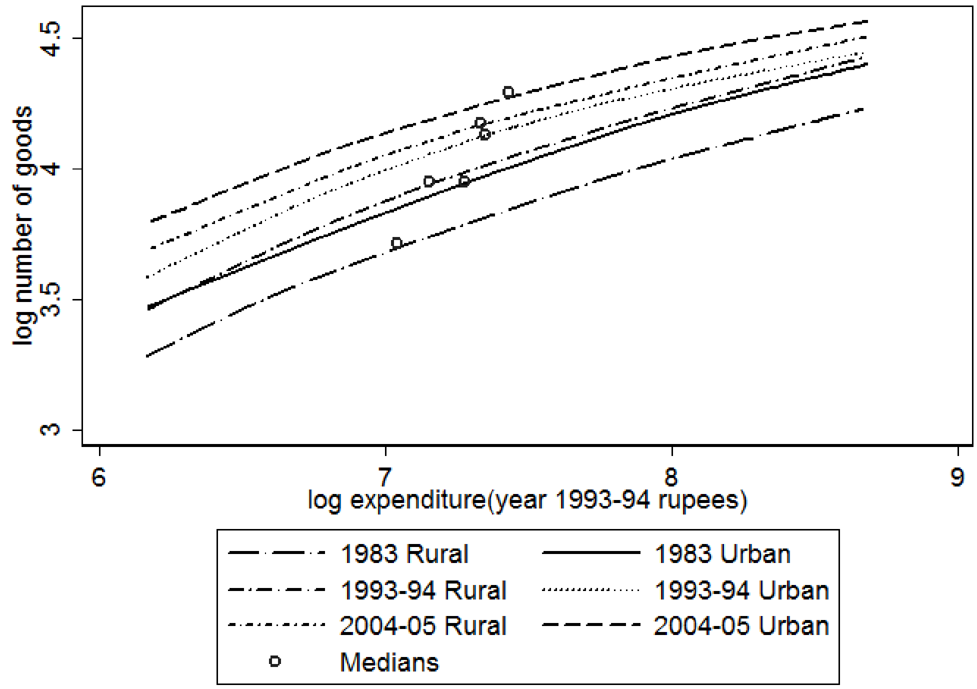
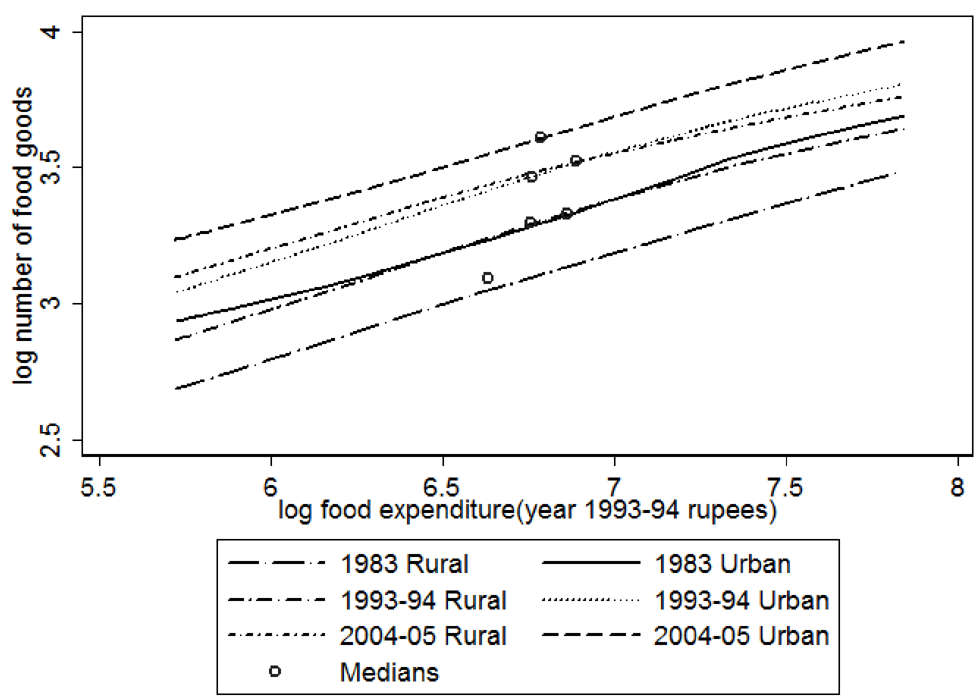



Figure 10: Underestimation of variety gains




(a) All varieties




(b) Food varieties

Figure 11: Non-parametric variety Engel curves for five person households

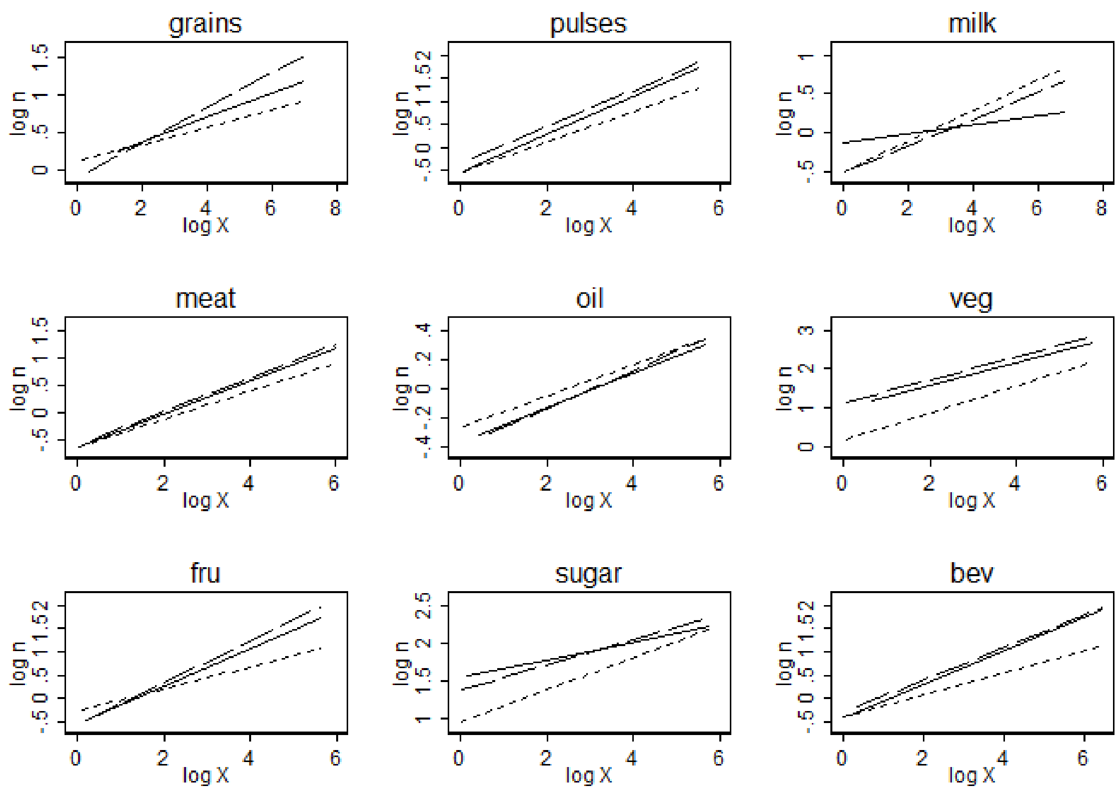



Figure 12: Group variety Engel curves: short dash (Rural 1983), solid (Rural 2004-05), long dash (Urban 2004-05)

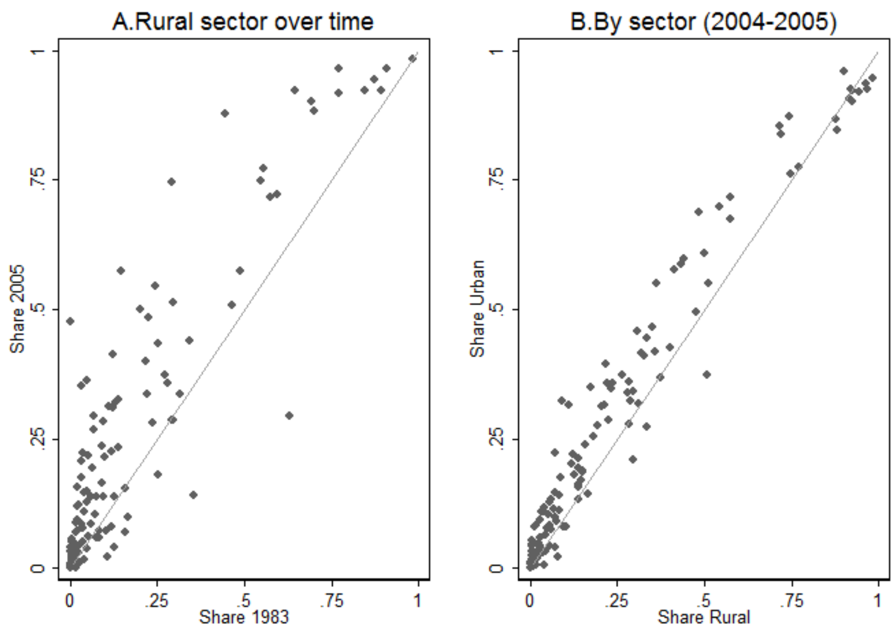



Figure 13: Differences in share of households consuming by variety

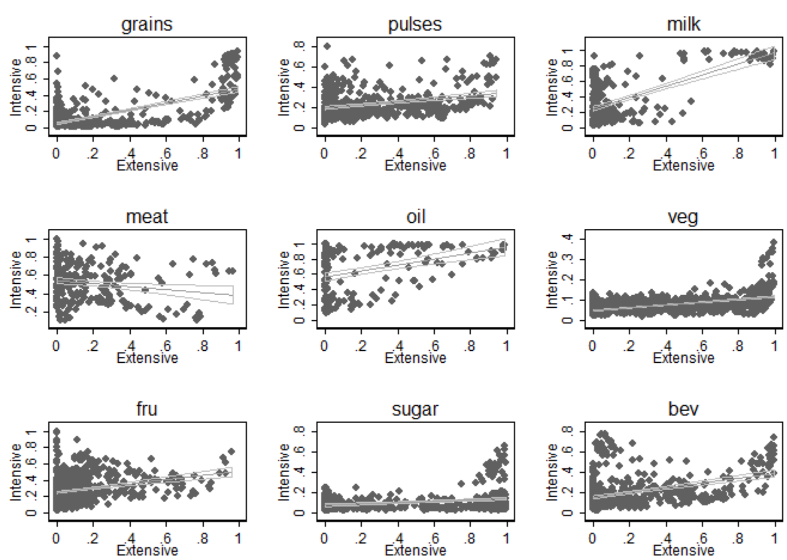



Figure 14: By state/sector/variety, 2005: Extensive (share of households consuming) vs. Intensive (budget share conditional on consuming) margins

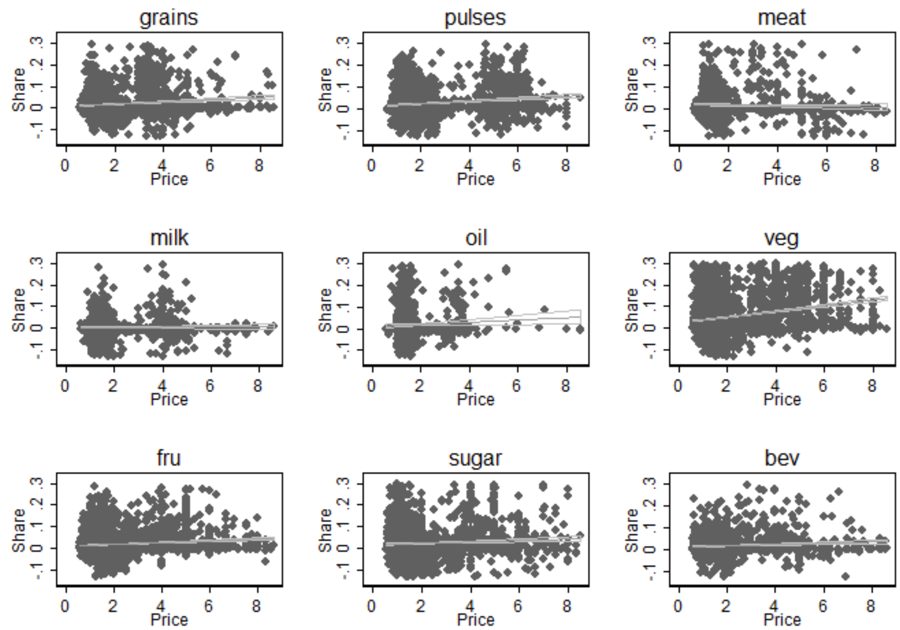



Figure 15: By state/sector/variety, 1983-2005: Change in share of households consuming vs. price 2005/price 1983

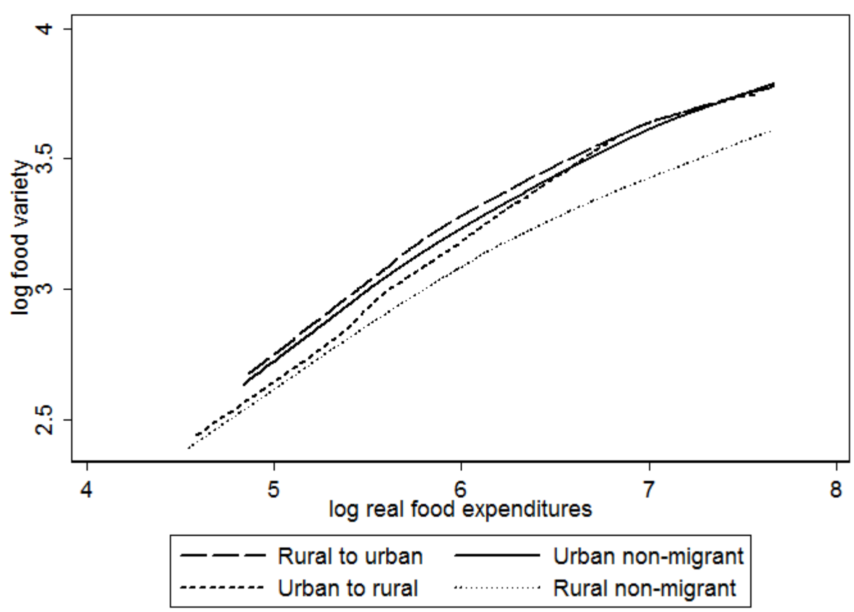
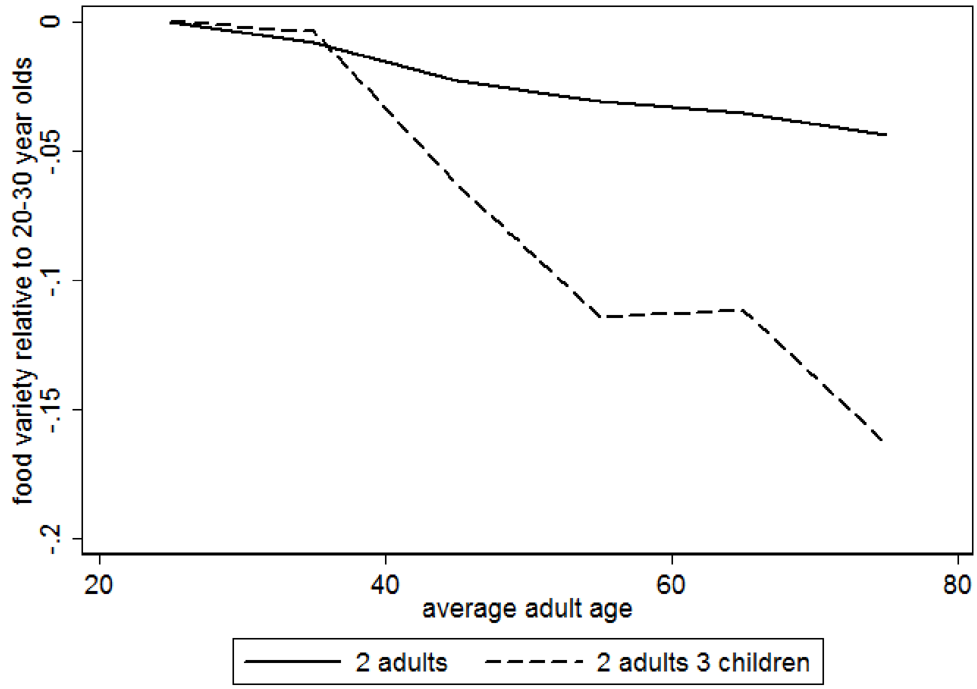
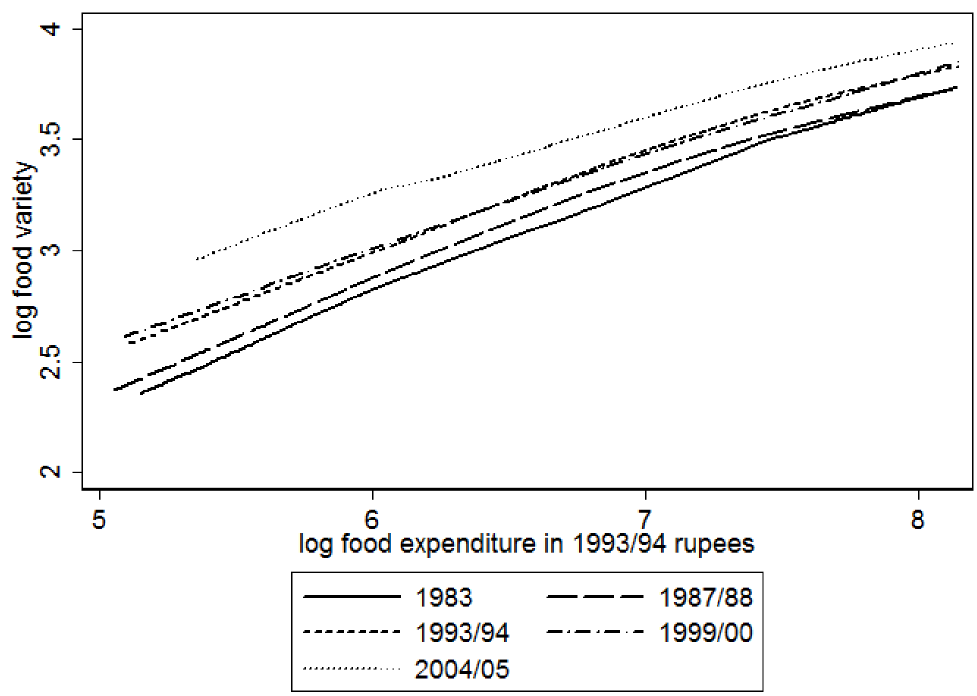



Figure 16: Variety consumption for migrants and non-migrants




(a) Relative to 21-30 year olds (over the lifecycle)




(b) 1953-1962 cohort (over time)

Figure 17: Variety consumption over the lifecycle and for a single cohort over time

Table 1: Mean household variety and real expenditure by sector/year

| Year | 1983 | | 1987-1988 | | 2004-2005 | | |
|---|---|---|---|---|---|---|---|
| | Rural | Urban | Rural | Urban | Rural | Urban | Categories |
| | | | Mean Varieties | | | | |
| All | 40.5 | 49.1 | 49.9 | 57.8 | 61.6 | 71.4 | 306 |
| Food | 21.4 | 26.5 | 26.4 | 31.5 | 31.4 | 37.3 | 134 |
| Grains | 2.3 | 2.5 | 2.5 | 2.9 | 2.8 | 3.4 | 18 |
| Pulses | 2.2 | 2.9 | 2.5 | 3.2 | 2.9 | 3.6 | 13 |
| Milk | 1.0 | 1.4 | 0.8 | 1.2 | 0.9 | 1.4 | 8 |
| Meat | 0.9 | 1.0 | 1.0 | 1.2 | 1.1 | 1.3 | 7 |
| Oil | 1.1 | 1.3 | 1.1 | 1.2 | 1.1 | 1.2 | 5 |
| Vegetables | 5.1 | 6.9 | 8.0 | 9.5 | 9.9 | 11.6 | 29 |
| Fruit | 0.9 | 1.6 | 1.5 | 2.5 | 1.9 | 3.2 | 24 |
| Sugar/spices | 6.5 | 6.7 | 6.8 | 6.8 | 7.8 | 8.0 | 14 |
| Bev./proc. | 1.4 | 2.2 | 2.2 | 3.1 | 3.0 | 3.7 | 16 |
| | | | Mean Real Expenditures (1993-94 Rural Base) | | | | |
| All | 1301 | 1534 | 1325 | 1616 | 1570 | 2151 | |
| Food | 850 | 972 | 867 | 966 | 844 | 962 | |
| Grains | 375 | 279 | 337 | 245 | 288 | 223 | |
| Pulses | 65 | 72 | 53 | 60 | 49 | 57 | |
| Milk | 90 | 150 | 131 | 154 | 117 | 158 | |
| Meat | 44 | 63 | 44 | 57 | 47 | 60 | |
| Oil | 51 | 86 | 61 | 88 | 78 | 98 | |
| Vegetables | 67 | 78 | 83 | 99 | 89 | 97 | |
| Fruit | 20 | 35 | 24 | 49 | 31 | 53 | |
| Sugar/spices | 88 | 95 | 78 | 84 | 75 | 78 | |
| Bev./proc. | 52 | 134 | 57 | 137 | 72 | 135 | |
| Households | 71132 | 37130 | 61837 | 39112 | 68982 | 39431 | |

## Table 2: List of goods in the survey and categories

| **GRAINS** | **OILS** | **VEGETABLES** | **SUGAR/SPICES** |
|---|---|---|---|
| rice | vanaspati, margarine | potato | sugar |
| chira | mustard oil | onion | gur |
| khoi, lawa | groundnut oil | radish | candy (misri) |
| muri | coconut oil | carrot | honey |
| other rice products | edible oil (others) | turnip | salt |
| wheat/atta | **MEATS** | beet | turmeric |
| maida | eggs | sweet potato | black pepper |
| suji, rawa | fish, prawn | arum | dry chillies |
| sewai, noodles | goat/mutton | pumpkin | garlic |
| bread (bakery) | beef/buffalo | gourd | tamarind |
| other wheat products | pork | bitter gourd | ginger |
| jowar & products | chicken | cucumber | curry powder |
| bajra & products | others | parwal / patal | oilseeds |
| maize & products | **FRUITS** | jhinga / torai | other spices |
| barley & products | banana | snake gourd | **BEVERAGES/PROCESSED** |
| small millets & products | jackfruit | cauliflower | tea |
| ragi & products | watermelon | cabbage | coffee |
| cereal substitutes | pineapple | brinjal | ice |
| **PULSES** | coconut | lady's finger | cold beverages |
| arhar (tur) | guava | palak/other leafy vegetables | fruit juice |
| gram (split) | singara | french beans and barbati | coconut green |
| gram (whole) | orange,mausami | tomato | other beverages |
| moong | mango | peas | biscuits |
| masur | kharbooza | chillis (green) | salted refreshments |
| urd | pears (naspati) | capsicum | prepared sweets |
| peas | berries | plantain (green) | cooked meals |
| soyabean | leechi | jackfruit (green) | cake, pastry |
| khesari | apple | lemon | pickles |
| other pulses | grapes | other vegetables | sauce |
| gram products | other fresh fruits | | jam jelly |
| besan | coconut (copra) | | other processed foods |
| other pulse products | groundnut | | |
| **MILK** | dates | | |
| milk: liquid (litre) | cashewnut | | |
| baby food | walnut | | |
| milk : condensed/ powder | other nuts | | |
| curd | raisin (kishmish, monacca etc.) | | |
| ghee | other dry fruits | | |
| butter | | | |
| ice-cream | | | |
| other milk products | | | |

Table 3: Number of food varieties consumed at different levels of aggregation

| | Units | Mean | Median | 10th pct | 90th pct |
|---|---|---|---|---|---|
| | 43rd NSS round 1987-88 | | | | |
| State | 18 | 133 | 134 | 129 | 134 |
| Region | 62 | 125 | 126 | 116 | 133 |
| District | 491 | 83 | 94 | 29 | 115 |
| Rural Village | 7467 | 41 | 42 | 29 | 55 |
| Urban Block | 4206 | 51 | 51 | 37 | 66 |
| Rural Household | 74373 | 24 | 23 | 15 | 34 |
| Urban Household | 41608 | 29 | 29 | 18 | 42 |
| | 61st NSS round 2004-05 | | | | |
| State | 18 | 131 | 132 | 126 | 134 |
| Region | 62 | 125 | 127 | 116 | 133 |
| District | 532 | 100 | 100 | 84 | 115 |
| Rural Village | 6910 | 53 | 52 | 40 | 66 |
| Urban Block | 3965 | 61 | 61 | 47 | 76 |
| Rural Household | 68982 | 33 | 32 | 22 | 44 |
| Urban Household | 39431 | 36 | 37 | 24 | 51 |

Table 4: Parameter estimates

| Group | $\sigma_g$ | 95% CI | $\psi_g$ | $\epsilon_g$ | $F_g$ |
|---|---|---|---|---|---|
| | Feenstra(1994) | | State/sector mean for 1983 | | |
| Grains | 1.85 | (1.02,2.68) | 0.33 | 9.97 | 0.06 |
| Pulses | 3.53 | (-1.15,8.22) | 0.23 | 3.74 | 0.09 |
| Milk | 17.96 | (-28.84,64.76) | 0.03 | 5.12 | 0.08 |
| Meat | 3.54 | (-3.51,10.59) | 0.07 | 10.01 | 0.06 |
| Oil | 8.77 | (-20.39,37.93) | 0.23 | 4.05 | 0.30 |
| Vegetables | 5.45 | (4.39,6.51) | 0.14 | 3.27 | 0.01 |
| Fruit | 6.57 | (-3.04,16.17) | 0.11 | 3.12 | 0.07 |
| Sugar/spices | 6.41 | (1.84,10.98) | 0.06 | 5.52 | 0.00 |
| Bev/proc. | 1.22 | (0.95,1.50) | 2.22 | 5.95 | 1.39 |
| | Deaton(1988) | | State/sector mean for 2004-05 | | |
| Grains | 2.04 | | 0.27 | 7.58 | 0.08 |
| Pulses | 2.21 | | 0.24 | 3.65 | 0.08 |
| Milk | 1.94 | | 0.02 | 16.92 | 0.03 |
| Meat | 1.93 | | 0.06 | 16.67 | 0.08 |
| Oil | 1.19 | | 0.24 | 3.91 | 0.25 |
| Vegetables | 1.20 | | 0.15 | 3.38 | 0.00 |
| Fruit | 5.74 | | 0.12 | 2.33 | 0.10 |
| Sugar/spices | 0.74 | | 0.11 | 6.19 | 0.00 |
| Bev./proc. | 1.32 | | 2.34 | 5.13 | 0.67 |

Confidence intervals for $\sigma$ use the Delta Method.

Other parameters evaluated using $\sigma$ from Feenstra.

Table 5: Welfare gains from variety over 1983-2005

| | CES model | | Variety Engel Curve | | | |
|---|---|---|---|---|---|---|
| | Aggregate | Average | Level | Distribution | | |
| | | | | 10th | 50th | 90th |
| | | | Group Means | | | |
| Grains | -0.012 | -0.006 | 0.063 | 0.053 | 0.067 | 0.079 |
| Pulses | 0.024 | 0.053 | 0.054 | 0.038 | 0.044 | 0.049 |
| Milk | 0.000 | -0.008 | -0.005 | 0.001 | -0.001 | -0.002 |
| Meat | 0.000 | -0.011 | -0.006 | -0.004 | -0.004 | -0.004 |
| Oil | 0.000 | 0.024 | 0.033 | 0.014 | 0.019 | 0.024 |
| Vegetables | 0.000 | 0.078 | 0.060 | 0.059 | 0.058 | 0.057 |
| Fruit | 0.001 | 0.056 | 0.029 | 0.000 | 0.010 | 0.018 |
| Sugar/spices | 0.001 | -0.063 | 0.014 | 0.019 | 0.017 | 0.015 |
| Bev./proc. | 0.023 | 0.638 | 0.468 | 0.338 | 0.379 | 0.410 |
| | | | Aggregated across groups | | | |
| Mean | 0.000 | 0.103 | 0.094 | 0.057 | 0.068 | 0.076 |
| Median | -0.001 | 0.100 | 0.097 | 0.061 | 0.067 | 0.075 |
| 90th | 0.005 | 0.192 | 0.129 | 0.087 | 0.090 | 0.099 |
| 10th | -0.003 | 0.008 | 0.055 | 0.026 | 0.038 | 0.049 |
| Rural mean | 0.000 | 0.058 | 0.078 | 0.048 | 0.061 | 0.072 |
| Urban mean | 0.000 | 0.146 | 0.108 | 0.066 | 0.074 | 0.080 |
| Mean no bev./proc. | -0.002 | 0.003 | 0.036 | 0.029 | 0.038 | 0.044 |
| Pctile group shares | | | | | | |
| Rural | | | | 0.056 | 0.063 | 0.067 |
| Urban | | | | 0.113 | 0.081 | 0.055 |

Gains expressed as percent reduction in cost-of-living due to variety/fixed costs.
Across group aggregates use group share weights except the last row,
which uses food exp. percentile specific group share weights.
Statistics by sector for 17 major states and Delhi.

Table 6: Welfare gains from variety for Urban vs. Rural sector 2004-2005

| | CES model | | Variety Engel Curve | | | |
|---|---|---|---|---|---|---|
| | Aggregate | Average | Level | Distribution | | |
| | | | | 10th | 50th | 90th |
| | | | Group Means | | | |
| Grains | -0.001 | 0.050 | 0.044 | 0.036 | 0.043 | 0.049 |
| Pulses | 0.000 | 0.035 | 0.020 | 0.015 | 0.018 | 0.020 |
| Milk | 0.000 | 0.004 | 0.003 | 0.000 | 0.002 | 0.003 |
| Meat | 0.000 | 0.004 | 0.002 | -0.001 | 0.000 | 0.001 |
| Oil | 0.000 | 0.038 | 0.009 | 0.002 | 0.004 | 0.006 |
| Vegetables | 0.000 | 0.023 | 0.015 | 0.012 | 0.013 | 0.014 |
| Fruit | 0.000 | 0.033 | 0.008 | 0.005 | 0.007 | 0.008 |
| Sugar/spices | 0.000 | 0.005 | 0.003 | 0.038 | 0.039 | 0.039 |
| Bev./proc. | 0.000 | 0.269 | 0.054 | 0.114 | 0.098 | 0.083 |
| | | | Aggregated across groups | | | |
| Mean | 0.000 | 0.060 | 0.023 | 0.028 | 0.030 | 0.031 |
| Median | 0.000 | 0.056 | 0.024 | 0.028 | 0.029 | 0.029 |
| 90th pct. | 0.000 | 0.101 | 0.035 | 0.036 | 0.044 | 0.050 |
| 10th pct. | -0.001 | 0.031 | 0.007 | 0.013 | 0.016 | 0.017 |
| Mean no bev./proc. | 0.000 | 0.027 | 0.019 | 0.019 | 0.023 | 0.026 |
| Pctile group shares | | | | 0.030 | 0.030 | 0.030 |

Gains expressed as percent reduction in cost-of-living due to variety/fixed costs.
Across-group aggregates use group share weights except the last row,
which uses food exp. percentile specific group share weights.
Statistics for 17 major states.

## Table 7: Welfare gains with other specifications

| | Gains over 1983-2005 | | | Gains over subperiods | | | |
|---|---|---|---|---|---|---|---|
| | Deaton $\sigma$ | High $\sigma$ | 4 Groups | 1983-1988 | 1988-1994 | 1994-2000 | 2000-2005 |
| Mean | 0.129 | 0.054 | 0.068 | 0.022 | 0.031 | -0.011 | 0.051 |
| Rural mean | 0.113 | 0.044 | 0.070 | 0.017 | 0.023 | -0.010 | 0.044 |
| Urban mean | 0.143 | 0.062 | 0.066 | 0.027 | 0.039 | -0.012 | 0.057 |
| Agg. share weights | | | | | | | |
| Rural 90/10 diff. | 0.021 | 0.014 | 0.014 | 0.014 | 0.009 | -0.003 | 0.004 |
| Urban 90/10 diff. | -0.003 | 0.008 | 0.008 | 0.003 | 0.010 | 0.000 | 0.000 |
| Pctile share weights | | | | | | | |
| Rural 90/10 diff. | 0.000 | 0.012 | -0.001 | 0.010 | 0.004 | -0.003 | -0.003 |
| Urban 90/10 diff. | -0.078 | 0.006 | -0.023 | -0.011 | -0.025 | 0.015 | -0.030 |
| | Gains for urban vs. rural 2004-05 | | | Urban vs. rural gains other years | | | |
| | Deaton $\sigma$ | High $\sigma$ | 4 Groups | 1983 | 1987-88 | 1993-94 | 1999-00 |
| Mean | 0.058 | 0.013 | 0.021 | 0.020 | 0.020 | 0.021 | 0.025 |
| Agg weighted 90/10 diff. | 0.017 | 0.001 | -0.002 | 0.014 | 0.010 | 0.004 | 0.005 |
| Pctile weighted 90/10 diff. | 0.015 | -0.001 | -0.009 | 0.010 | 0.006 | 0.001 | 0.003 |

Percent reduction in cost-of-living due to variety/fixed costs. Means for 35 state/sectors (over time) or 17 states (urban versus rural).
90/10 difference is extra welfare gain for 90th percentile over 10th.
High $\sigma$ uses upper end of 95% CI of Feenstra estimates.

## Table 8: OLS regression of food variety on district level covariates for 1987-1988.

| Dep. var. | Mean variety | | Mean resid. variety | | Aggregate variety | |
|---|---|---|---|---|---|---|
| Variable | Coef. | s.e. | Coef. | s.e. | Coef. | s.e. |
| Real food exp. | 0.307*** | (0.070) | -0.058 | (0.069) | 0.288** | (0.122) |
| Pop. density | 0.032* | (0.018) | 0.032* | (0.017) | -0.012 | (0.023) |
| Road density | 0.039*** | (0.012) | 0.039*** | (0.011) | 0.009 | (0.009) |
| Dist. to coast | -0.028*** | (0.009) | -0.027*** | (0.008) | 0.000 | (0.007) |
| Share electric. | 0.325*** | (0.050) | 0.312*** | (0.049) | 0.013 | (0.049) |
| Sample hh. in district | 0.020 | (0.027) | 0.017 | (0.027) | 0.250*** | (0.051) |
| Constant | 0.802 | (0.518) | 0.657 | (0.514) | 1.335 | (0.989) |
| $R^2$ | 0.503 | | 0.377 | | 0.718 | |

$*$, $**$ and $***$ denote significance at the 10%, 5% and 1% levels. Robust standard errors.
N=271 Indian districts. All variables in logs except share of households with electricity.

Table 9: OLS regression of weekly household minutes spent on shopping and other travel

| Dep. var. | Shopping | | Other travel | |
|---|---|---|---|---|
| Variable | Coef. | s.e. | Coef. | s.e. |
| A: Village/block fixed effects | | | | |
| log expenditure | 26.5*** | (8.8) | 140.8*** | (22.1) |
| log household size | 58.4*** | (8.8) | 495.6*** | (22.2) |
| B: Urban vs. rural (omitted) | | | | |
| log expenditure | 19.1 | (12.7) | 71.8** | (33.1) |
| log household size | 76.3*** | (15.4) | 612.2*** | (49.1) |
| urban | 61.6*** | (17.8) | −168.8*** | (35.7) |
| C: Village/town size (small village omitted) | | | | |
| log expenditure | 16.8 | (11.8) | 74.4** | (32.1) |
| log household size | 78.3*** | (14.9) | 608.5*** | (48.1) |
| medium village | 1.0 | (17.2) | -25.6 | (52.3) |
| big village | 6.6 | (18.7) | −78.1* | (47.6) |
| small town | 44.3* | (25.1) | −229.2*** | (69.0) |
| medium town | 48.1** | (20.2) | −244.8*** | (62.2) |
| big town | 90.6*** | (29.2) | −208.2*** | (59.6) |

N=18,563 households. Standard errors (clustered by district) in parentheses.

*, ** and * * * denote significance at the 10%, 5% and 1% levels.

Sample means are 200 (shopping) and 883 (other travel) minutes per week.

# A Appendix: Alternative demand systems

In this appendix I describe several other parameterizations of demand that generate variety Engel curves and their positive and normative implications.

## A1. Fixed budget costs

Instead of utility fixed costs, monetary fixed costs could limit variety. These may be indirect complements to variety purchase (e.g. costs of transport and storage) or enter the variety budget directly (e.g. indivisibilities or bulk discounting). Suppose then we write the budget constraint as $\int_0^n q_i p_i di + n^\nu F_b \leq X$ to reflect these fixed costs (so we assume that the costs we are interested in are part of expenditure X). We can take n as given and solve for quantities as a function of $X - n^\nu F_b$, and then write the utility maximization problem as:

$$\max_n \left( \frac{X - n^\nu F_b}{pn^{-\psi}} \right) \tag{23}$$

We can re-write the expression as $\left(\frac{X}{pn^{-\psi}}\right) - \frac{n^\nu F_b}{pn^{-\psi}}$ and by defining $F = F_b/p$ and $\epsilon = \nu + \psi$ the variety choice problem is identical to the one from equation 5. The two models thus have identical implications for welfare and variety choice.

The only differences concern empirical predictions. One difference between the two models is that fixed utility costs imply that an estimate of marginal benefit $\psi$ has some predictive power for the slope of the variety Engel curve, and there is some evidence to support this. The other difference is related to the division of the budget fixed cost across expenditure categories. When the fixed cost is measured as part of expenditure on the good, it is still true that holding n constant the quantity demand function for a variety takes the usual CES form (in $X - n^\nu F_b$). However, the expenditures on that variety will depend on how/whether the fixed costs are allocated across expenditure shares, so expenditure shares would not take a CES form. If the fixed costs are separate from expenditures on the variety then expenditure shares would still be CES in $X - n^\nu F_b$, but we would have to add them to expenditures on variety in a group to get $X$.

## A2. No fixed costs - quadratic utility

Suppose the consumer problem is:

$$U = \alpha \int_0^n q_i di - \frac{1}{2}\gamma \int_0^n q_i^2 di \tag{24}$$

s.t. $\int_0^n q_i p_i \leq X$. Demand curves are linear with $q_i = \frac{\alpha}{\gamma} - \frac{\lambda p_i}{\gamma}$. The marginal utility of expenditure ($\lambda$) falls in expenditures, shifting demand curves up and lowering reservation prices given by $p_i = \frac{\alpha}{\lambda}$.

With symmetric prices $p_i = p$ and associated budget constraint $nqp = X$, indirect utility is strictly increasing in variety with $V = \alpha\frac{X}{p} - \frac{\gamma X^2 p^2}{2n}$. We can bound variety demand by adding $-Fn^{\epsilon}$ to the utility function, generating upward sloping variety Engel curves that depend only on fixed costs, given by $n = \left[\frac{\gamma X^2 p^2}{2\epsilon F}\right]^{\frac{1}{1+\epsilon}}$. The budget constraint ensures constant quantity and an inverse relationship between variety and quantity per variety.

With $p_i = pi^{\frac{1}{\theta}}$ there is a well-defined variety Engel curve even with no cost to variety, given by $n = \left(\frac{\gamma I}{\alpha p}[\frac{\theta}{1+\theta} - \frac{\theta}{2+\theta}]^{-1}\right)^{\frac{\theta}{1+\theta}}$. The variety Engel curve slope depends only on relative prices $\theta$ so should be closely related to relative quantities/prices/budget shares. Increasing the product differentiation parameter $\gamma$ would raise variety but lower welfare (similar to a fall in $\sigma$). The expenditure function is

$$X = \left[U\frac{\gamma(1+\theta)}{\alpha^2}\right]^{\frac{1+\theta}{\theta}} \left(\frac{\alpha p}{\gamma}\left[\frac{\theta}{1+\theta} - \frac{\theta}{2+\theta}\right]\right) \tag{25}$$

Variety and relative quantities/expenditures are uniquely determined by a single parameter, and variety can only increase through relative prices (or relaxation of supply-side constraints for constrained households). Adding fixed costs to this model can alleviate this issue but does not yield a tractable close-form solution.

## A3. Welfare measurement

CES utility typically yields higher welfare from variety than other demand systems. The welfare gain for a new good n under constant elasticity demand curves is $CS_n(CES) = \frac{p_n q_n}{\sigma_n - 1}$ with $\sigma_n$ the own-price elasticity of demand. Hausman (2003) observes that a lower bound on welfare gains for demand curves convex to the origin is given by linear demand curves, that have $CS_n(linear) = \frac{p_n q_n}{2\sigma_n}$. The translog falls somewhere in-between. This suggests that a conversion factor of $\frac{\sigma_n - 1}{2\sigma_n}$ should be applied to the welfare gains under CES to get a lower bound, which reduces welfare gains by between 50% (for high elasticities) to 75% (for an elasticity of 2).

In terms of the variety Engel curve model we could write:

$$\max_n \frac{X}{P(n)} - F(n) \tag{26}$$

replacing $P(n)$ with any expenditure function, such as the homothetic translog index in Feenstra (2009) or quadratic utility above.[36]. The complication of these other specifications is that the elasticity $\frac{\partial P(n)}{\partial n}\frac{n}{P(n)}$ will not be constant, leading to non-linear (log) variety Engel curves, but locally we could calculate the lower marginal benefit from variety under these demand specifications.

# B Appendix: Feenstra index

Let k denote the time period/area, g is the grouping within which we expect the elasticity of substitution to be constant (there are G groupings in total) and i indexes varieties within each grouping. The across and within group CES utility functions are:

$$U_k = \left(\sum_{g=1}^{G} u_{kg}^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \tag{27}$$

$$u_{kg} = \left(\sum_{i \in I_{gk}} d_{kgi}^{1/\sigma_g} q_{kgi}^{\frac{\sigma_g - 1}{\sigma_g}}\right)^{\frac{\sigma_g}{\sigma_g - 1}} \tag{28}$$

where $I_{kg}$ is the set of group g varieties consumed in area k, $q_{kgi}$ is the quantity of variety i of group g consumed in area k and $d_{kgi}$ is a quality or taste parameter. Denoting the set of group g varieties common to area k and 0 as $I_g$, the exact price index for area k relative to area 0 is given by:

$$P_k = \prod_{g=1}^{G} \left(\frac{P_{kg}}{P_{0g}}\right)^{w_{kg}} \tag{29}$$

[36]We could use $P(n) = \frac{1}{1 + \frac{\lambda_{new}}{2\sigma}}$ as the index where $\lambda_{new} = \frac{\sum_{i \in new} q_i p_i}{X}$

$$P_{kg} = \prod_{i \in I_g} \left( \frac{p_{kgi}}{p_{0gi}} \right)^{w_{kgi}} \left( \frac{\lambda_{kg}}{\lambda_{0g}} \right)^{\frac{1}{\sigma_g - 1}} \tag{30}$$

$$\lambda_{kg} = \frac{\sum_{i \in I_g} p_{kgi} q_{kgi}}{\sum_{i \in I_{gk}} p_{kgi} q_{kgi}} \tag{31}$$

Sato-Vartia share weights are given by $w_{kg} = \frac{\frac{s_{kg} - s_{0g}}{\ln s_{kg} - \ln s_{0g}}}{\sum_{g \in G} \frac{s_{kg} - s_{0g}}{\ln s_{kg} - \ln s_{0g}}}$, $w_{kgi} = \frac{\frac{s_{kgi} - s_{0gi}}{\ln s_{kgi} - \ln s_{0gi}}}{\sum_{i \in I_g} \frac{s_{kgi} - s_{0gi}}{\ln s_{kgi} - \ln s_{0gi}}}$, $s_{kg} = \frac{p_{kg} q_{kg}}{\sum_{g \in G} p_{kg} q_{kg}}$, and $s_{kgi} = \frac{p_{kgi} q_{kgi}}{\sum_{i \in I_g} p_{kgi} q_{kgi}}$. The necessary and sufficient condition for exactness with CES demand is that $d_{kgi} = d_{0gi}$ for $i \in I_g$ and $I_g \neq \emptyset$.