The Welfare Implications of Rising Price Dispersion*

Christian Broda (University of Chicago, GSB and NBER) and
John Romalis (University of Chicago, GSB and NBER)

Version: April 24, 2009

Preliminary: please circulate and comment.

Abstract

The conventional belief is that real income growth for most households has been disappointing over the last thirty-five years. We argue that studies of economic progress have mostly abstracted from an important question - what can and what do different consumers purchase with their income? In a world with a stationary relative price distribution this is not an important abstraction. But the distribution of relative prices is anything but stationary, and for similarly defined, though not identical, goods and services there is more dispersion in prices than ever before. The relative prices of some qualities of products is rising - the qualities that are consumed disproportionately by wealthier consumers. Measured against the prices of products that poorer consumers actually buy, their incomes have been rising steadily. We take this to mean that the poor, relative to their past selves, have been experiencing quite rapid progress. As a consequence, we believe that more thought may need to be placed on exactly what we care about when we study inequality, or indeed when we make any conceptually similar comparisons.

* We would like to thank the Economic Research Service at the US Department of Agriculture for access to their Homescan Data, especially Ephraim Leibtag for careful explanation of the data. We would also like to thank Anil Kashyap, Steve Davis, Robert Feenstra, Erik Hurst and David Weinstein for early suggestions and ACNielsen's Ivan Rocabado and Maura Ehlbretch for their careful explanation of the Homescan data. Christian Broda wishes to thank the NSF (grant #0214378) and the Initiative on Global Markets at the University of Chicago GSB for research support. We thank Alejo Costa for excellent research assistance.

I. Introduction

Real income growth for most households has been disappointing over the last thirty-five years. Real income for the bottom half of households grew by just 13 percent from 1973 to 2006, or less than 0.4 percent per annum (Figure 1). Households in the upper reaches of the distribution fared somewhat better, with real household income at the 90th percentile rising by 41 percent in this period or 1.0 percent annually. These and similar facts have led to an active and controversial area of inquiry into the distribution of the fruits of economic progress (see Goldin and Katz (2007) for a recent survey). We argue that some of these basic facts are not wrong, but potentially misleading, and are partly just an artifact of how we measure things. "Real" income or welfare does not only depend on the dollars in consumers' pay packets, it also crucially depends on what they can buy with those dollars. Conventional income and inequality papers implicitly or explicitly assume that the cost of living has been rising at an identical rate for all American consumers. Little attention has been paid to the role played by inflation differentials between the consumption baskets of the rich and poor. In this paper we relax a standard assumption underlying the calculation of conventional price indexes –that rich and poor have identical homothetic preferences so consume a common basket of goods– and re-examine the conventional evidence on income growth and inequality. Using detailed household consumption data between 1994 and 2005, we find that the rise in real income for poorer households has been much faster than implied by conventional statistics, so that the "real" distribution of the gains from economic progress may have been much more equal than we believe.

We argue that an important change in the structure of prices has been occurring for decades, one that has important implications for the relationship between income and welfare. Figure 2 illustrates one of our main points. Given the structure of prices in 1973, income might map into welfare according to the solid line. But relative prices have shifted substantially since 1973. If underlying preferences are nonhomothetic, then the mapping between income and welfare in 2006 does not have to be parallel to that in 1973. It might become flatter, in which case the price index for a person with low permanent income has risen less rapidly than the price index for a person with high permanent income. This has obvious implications for measuring welfare levels. It also potentially impacts how we think about inequality, because the inequality trajectory as measured on the horizontal (nominal income) axis may differ quite substantially from the path of inequality measured on the vertical (welfare) axis.

The paper uses detailed household consumption data on a large set of products sold in grocery, drug, mass merchandise, and other stores - essentially anything that carries a Universal Product Code (UPC or barcode). Moreover, we have detailed information on household characteristics. The relative price of goods consumed disproportionately by poor households has been falling, so that the price index for the poorest quintile of households - including correction for new goods bias - has only risen by 4.5 percent between 1994 and 2005, while the price index for the richest quintile of households has risen by 11.3 percent, and by 13.6 percent for the richest 5 percent of households in the sample. The dominant proximate cause of this difference is a fall in the relative price of existing varieties consumed by the poor, so that a price index that ignores product creation and destruction rose 8.7 percent less for the poorest quintile than for the richest quintile. The introduction of new goods benefits each group substantially but relatively equally, with the rich benefitting a mere 1.9 percent more than the poor from this process.

We cite less formal evidence that strongly suggests that this process is not limited to the products in our sample nor to our time period. Our formal evidence and these additional facts have important implications for the measurement of real income growth and, as a consequence, for inequality. For example, conventionally measured real household income at the 10th percentile rose by 8 percent over our sample period 1994-2005, or 0.7 percent annually. Growth at the 90th percentile was 14 percent, or 1.2 percent annually[1] (Figure 1). But if differences in income-group specific inflation rates in our sample are representative of the broader economy, then real income growth has been much more substantial and equal: 18 percent at the 10th percentile (1.5 percent annually) and 17 percent at the 90th percentile (1.4 percent annually). The higher average growth we find is due to inclusion of the gains from new goods, something mostly omitted from conventional CPI measures.

[1] Census Bureau: Selected Measures of Household Income Dispersion: 1967-2007.

II. Data Description

II. A. Overview

The paper uses detailed household consumption data on a large set of mostly non-durable products sold in grocery, drug, mass merchandise, and other stores. The data is part of the Homescan database, collected by ACNielsen in the United States, that records prices and quantities of the purchases of thousands of households. ACNielsen provides Universal Product Code (UPC or barcode) scanners to a demographically representative sample of households. Households then scan in every purchase they make. We use two extracts of the complete Homescan database that provides us with a vast array of goods with barcodes. Moreover, we have detailed information on household characteristics.

We refer to the first extract of the Homescan data as our “Non-Durable” database. For this extract we have price and quantity data for every UPC purchased by a sample of 41,500 households for every quarter in 1994, and 55,000 households every quarter between 1999:Q1 and 2003:Q4. In addition, we have household-level information on every UPC purchase of a sample of 3500 households in 2003:Q4, together with detailed household characteristics. Table 1A summarizes this database in terms of the number of households, number of UPCs and “modules” (ACNielsen’s classification of different UPCs into broader product categories). Examples of non-food modules included in this database include “cosmetics”, “toys and sporting goods”, “houseware appliances”, “cookware”, and “wrapping materials and bags”.

The second extract we use includes detailed information on the food purchases and demographic characteristics of a large subsample of households included in the Homescan database between 1998 and 2005. We refer to this extract as the “Food” database. In this extract, we have household level data on every purchase in 60 percent of all modules. Examples of food modules are “soft drinks non-carbonated”, “sugar, sweeteners”, “seafood” and “prepared, ready to eat food”. Table 1B provides summary statistics of the number of UPCs, modules and households included in this database. The data is divided into four broad categories: dairy, dry grocery, frozen and processed foods, and random weight products. We obtained detailed household information on approximately 8,000 households from 1998 to 2003, and around 38,000 for 2004 and 2005. In 2005 this extract includes 640 modules and over 380,000 UPCs, most of which are classified under the dry grocery category. As we explain in the next section, we combine the information from both Homescan extracts to compute income-specific price indices over time.

A number of characteristics of the household are included in this database. In particular, household income, the head of household's occupation and education level, and household size are included. The distribution of households by income group and household head education level are provided in Figures 3A and 3B. Since we rely heavily on the information of households that are among the poorest and richest in our data, it is useful to examine how well our data represents the true population. According to the US Census Bureau the cutoffs for the 10th and 20th percentile income distribution are approximately \$12,000 and \$20,000, respectively.[2] Around 8 percent of our sample of households falls below the \$15,000 threshold, and around 14 percent of the households have income less than \$20,000. This implies that in 2005 we have detailed data on over 5,000 households which are in the lowest deciles of the income distribution.[3,4]

These data are ideal for understanding how prices evolve for households in different income groups. First, they include a long time series of price and quantity data for a large sample of non-durable consumption goods consumed by each income group. This is an advantage relative to current studies that do not observe the specific prices that households pay for each item. Our data circumvents these limitations by using data directly collected by a representative set of households. Second, we can identify the different goods purchased by each income group down to the barcode level. While official statistics are based on the basket of a representative agent, these data allow us to measure the differences in consumption baskets across income groups. This information is not observed by the BLS or other statistical agencies. A third crucial characteristic of these data is that along with prices of each product, quantities of the identical products are collected at the same frequency. Therefore we observe expenditure weights by income group. We can be very precise when analyzing income-group specific price indexes.

[2] U.S. Census Bureau, Current Population Survey, 1968 to 2007 Annual Social and Economic Supplements

[3] The U.S. Census Bureau selects a sample of approximately 7,100 households to build the CEX survey.

[4] While we have little information on response rates by different income groups, Nevo et al (2008) suggest that the coverage is good for all income groups, while response rates and measurement error is larger for higher income groups. We take some comfort in that we observe purchases for a large number of households in the upper decile of the income distribution.

II. B. Stylized Facts on Consumption Baskets by Income Group

In this section we use Homescan data to document three key facts that highlight the differences in the pattern of consumption across different income groups. First, the basket of non-durable goods consumed differs systematically by income group – the poor consume lower quality products than the rich. Second, the poor consume fewer varieties of goods (fewer UPCs), and this gap with the rich has been widening over the sample period. Finally, the poor spend relatively more on food versus other bar-coded products.

The Homescan data reveals that poorer households consume lower quality (lower priced) goods. A useful feature of the ACNielsen Food data is that in addition to the price and quantity of each UPC consumed, it provides detailed information on the weight or volume of each product. This allows us to compute unit values for each module – size pair. For instance, within the module "Milk", there are UPCs sold under many different sizes (e.g., 16 oz, 32 oz and 64 oz). A single-person household with income over $100,000 pays 32 percent more per oz of milk than a family of four earning $25,000 - $30,000. In particular, richer households consume a much higher fraction of organic milk. Figure 4A reports this result for all food products. For each module – unit of size pair (e.g., Milk – measured in ounces), we calculate the average unit value (price per ounce) paid by each group relative to the average unit value for this module – unit of size pair. More formally we compute :

$$(1) \qquad rel\_uv_{m,s,I,j} = \ln \frac{\overline{uv}_{m,s,I,j}}{\overline{uv}_{m,s}}$$

where $\overline{uv}_{m,s,I,j}$ is the mean unit value for purchases in module '*m*' of products measured in size units '*s*' by households in income group '*I*' with '*j*' occupants, and $\overline{uv}_{m,s}$ is the average unit price paid by all households for products in module *m* with size units *s*. Figure 4A shows that the milk example is not atypical. Small households with large income routinely pay 30 percent more per unit than larger households with smaller incomes. Since the poor are only paying 5 percent less for the exact same UPC (Figure 4B), most of the lower average price for the poor therefore comes from selecting cheaper brands and more economical sizes.

The raw price facts suggest that we can do better than categorizing households by income alone. Conditional on household income, a larger household is poorer. Henceforth we will categorize households by a per-capita measure: annualized sampled expenditures per capita. Since we know the exact purchase date of each item we can observe for how much of each year each household has been in the sample (overwhelmingly for the whole year), so we know the rate at which households are spending on products with UPCs. For each household, I use their entire sample history to sort them by expenditure-per-capita. Due to the current lack of overlap in the samples, this is done separately for the Food sample and for the Nondurable sample. Food expenditures for each quintile appear in Figure 5A.

A second fact revealed by the Homescan data is that poorer households consume fewer products than richer households, and this gap has been growing. For each household we compute the number of unique UPC's purchased. We then calculate the average number of unique UPC's purchased per household by expenditure quintile (using only the "original" sample and not the extra households in 2004-2005). Figure 5B shows the number of UPCs per household by quintile. The lowest quintile has reduced the number of UPCs they purchase relative to the richest households by an average of 20 percent.

The final fact that we document in this section is that the share of food consumption differs markedly across groups. Using the household-level data in the Non-durable database, we find that food modules account for 73 percent of expenditures for poorer households, but only 57 percent of expenditures for richer households (Figure 5C).

III. Calculating Inflation Rates by Income Groups

In this section we derive exact price indexes by income group. This differs from conventional or official CPI measures that are based on a representative household in the economy.[5] We build income-group specific price indexes by relaxing two standard assumptions underlying conventional price indexes. First, we allow the expenditure shares on each good consumed to differ between poor and rich. Second, we allow the introduction of new goods to affect the calculation of the cost-of-living index, and we permit the effect to differ across income groups. We adopt a nonparametric approach. We essentially allow for an underlying non-homothetic preference structure, but then approximate this structure with a series of non-symmetrical CES utility functions [... we need to investigate whether we can integrate these back up ...]. Consumers may simply be on different points of the same Engel curve. Since we do not

[5] Statistical offices around the world compute changes in consumer prices for an “average” person in the economy. In the US, the BLS conducts “Point of Purchase Surveys” to assess where people are buying their products. These surveys use demographic and socioeconomic information that allows BLS to monitor how well the selected interviewers represent the overall population.

focus on understanding the reasons behind the differences in consumption behavior across income groups it is simplest to build consumer price indexes based on utility functions where the expenditure shares vary exogenously across income groups.[6]

We now write down these restrictions formally. The first step towards deriving an exact price index is defining a utility function over all goods in our sample. Suppose that the preferences of a particular household with income *I* can be represented by a two-level utility function:

(2) $$\Omega_{It} = F_{It}^{\alpha_I} NF_{It}^{1-\alpha_I}$$

where F_{It} is the sub-utility derived from the consumption of food products and NF_{It} is the sub-utility derived from non-food items in our sample. The Cobb-Douglas assumption between the aggregate food good and the aggregate of other goods is due to the current structure of our sample, and will be relaxed when more extensive data are obtained. We exploit the work of Sato (1976) and Vartia (1976) on ideal price indexes to simply assume that both F_{It} and NF_{It} are multi-level CES functions. Define U_{Ft} as the set of all possible food UPCs in period *t*. For future reference, each group *I* may consume a different set of UPCs, i.e. $U_{FIt} \subset U_{Ft}$, and the set of UPCs consumed in both periods *t* and *t-1* by group *I* is given by $U_{FI} \subset U_F = U_{Ft} \cap U_{Ft-1}$ where U_F is the set of all common UPCs between periods. Non-food UPCs in U_{NF} are similarly arranged. We have no need to impose further structure, at least not yet.

If the set of UPCs available for each group is fixed over time, Sato (1976) and Vartia (1976) have derived the exact price index in the case of any multi-level CES utility function. In the case where expenditure shares of particular UPCs are allowed to vary by income group *I*, the "common goods" exact price index is defined as follows,

(3) $$\pi_{FI} = \prod_{u \in U_{FI}} \left(\frac{p_{uFIt}}{p_{uFIt-1}} \right)^{w_{uFIt}}$$

[6] Of course an alternative way to proceed is to have the same utility function across income groups, but allowing for non-homothetic preferences.

This is the geometric mean of the price changes of individual UPCs that belong to the set U_{FI}, where the weights are ideal log-change weights.[7] These weights are computed using expenditure shares of each income group, s_{uFI}, in the two periods, as follows:

(4) $$s_{uFIt} = \frac{p_{uFIt} c_{uFIt}}{\sum_{u \in U_{FI}} p_{uFIt} c_{uFIt}}$$

(5) $$w_{uFIt} = \frac{\dfrac{s_{uFIt} - s_{uFIt-1}}{\ln s_{uFIt} - \ln s_{uFIt-1}}}{\sum_{u \in U_{FI}} \left(\dfrac{s_{uFIt} - s_{uFIt-1}}{\ln s_{uFIt} - \ln s_{uFIt-1}} \right)}$$

where *c* denotes consumption quantity. The numerator of (5) is the difference in shares over time divided by the difference in logarithmic shares over time. The weights *w* capture all we need to know about how consumers in group *I* value each UPC and how prepared they are to substitute it for other products. Consider what happens in response to a price rise for UPC *u*. If consumers are very prepared to substitute other products for *u*, then the expenditure share on *u* will decline substantially and the weight function in (5) gives a weight much closer to the lower expenditure share, assuming that the denominator in (5) is close to 1. Products that are highly substitutable for other products can receive a much lower weight than their average expenditure share. For products where the expenditure shares barely move in response to a price change, the weight is very close to the simple average expenditure share.

The introduction of new goods implies that a true cost-of-living index will differ from the common-goods exact price index defined in (3). Feenstra (1994) showed how to modify this common-goods exact price index for the case of different, but overlapping, sets of varieties in the two periods. Suppose that there is a set of UPCs $U_{FI} \neq \varnothing$ that are available in both periods, and for which the taste parameters are constant. Extending the work of Feenstra (2004) we can derive different cost-of-living indexes by quintile from the utility structure allowing for product creation and destruction:

[7] As explained in Sato (1976), a price index *P* that is dual to a quantum index, *Q*, in the sense that $PQ = E$ and shares an identical weighting formula with *Q* is defined as "ideal". Fischer (1922) was the first to use the term ideal to characterize a price index. He noted that the geometric mean of the Paasche and Laspayres indices are ideal.

(6) $$COLI_{FI} = \pi_{FI} \times \left(\frac{s_{FIt}^{c}}{s_{FIt-1}^{c}} \right)^{\frac{1}{\sigma_F - 1}},$$

$$\text{where } s_{FI,t}^{c} = \frac{\sum_{u \in U_{FI}} p_{uFIt} c_{uFIt}}{\sum_{u \in U_{FIt}} p_{uFIt} c_{uFIt}}.$$

$COLI_{FI}$ is the cost-of-living index (or exact price index) for food for group *I* adjusted for new-goods bias between periods *t* and *t* – 1, and $s_{FI,t}^{c}$ is the share of common UPCs in food consumed by group *I* to the total food consumption of quintile *I*. We define $COLI_{NFI}$ for non-food items similarly.

Given the Cobb-Douglas aggregator between food and non-food items the “common” goods aggregate exact price index for all items in our sample is:

(7) $$\pi_I = \pi_{FI}{}^{\alpha_I} \pi_{NFI}{}^{1-\alpha_I}$$

where the weights α_I are the simple expenditure shares in Figure 5C. Explicitly allowing product turnover, we obtain the following expression for the relationship between the conventional inflation measures and changes in the cost-of-living index:

(8) $$COLI_I = \pi_{FI}^{\alpha_I} \times \pi_{NFI}^{1-\alpha_I} \times \left(\frac{s_{FIt}^{c}}{s_{FIt-1}^{c}} \right)^{\frac{\alpha_I}{\sigma_F - 1}} \times \left(\frac{s_{NFIt}^{c}}{s_{NFIt-1}^{c}} \right)^{\frac{1-\alpha_I}{\sigma_{NF} - 1}}.$$

Overall inflation adjusted for new-goods bias is comprised of two different components: 1) $\pi_{FI}^{\alpha_I} \times \pi_{NFI}^{1-\alpha_I}$ is the “common-goods” exact price index for group *I* for non-durable goods; and 2) $\left(\frac{s_{FIt}^{c}}{s_{FIt-1}^{c}} \right)^{\frac{\alpha_I}{\sigma_F - 1}} \times \left(\frac{s_{NFIt}^{c}}{s_{NFIt-1}^{c}} \right)^{\frac{1-\alpha_I}{\sigma_{NF} - 1}}$ captures the role that product turnover, or new goods bias, plays for each group.

The geometric average of $s_{FI,t}^{c} / s_{FI,t-1}^{c}$ ratios captures the difference (or bias) between a true cost-of-living index relative to the common-good price indexes like the CPI. Mechanically, when the share of new UPCs consumed by group *I* in period *t* is larger than the share of UPCs that have disappeared from group *I*’s basket in period *t* –1, this $s_{FI,t}^{c}$ ratio is smaller than 1. The smaller is this share ratio, the smaller is the overall inflation rate that takes product turnover into account relative to a conventional (common-goods) price index that does not.

The inflation rate in (8) also depends on the elasticities of substitution σ_F and σ_{NF}. As σ_F grows, the term $\frac{1}{\sigma_F - 1}$ approaches zero, and the bias term $\left(\frac{s^c_{FIt}}{s^c_{FIt-1}}\right)^{\frac{\alpha_I}{\sigma_F - 1}}$ becomes unity. Similarly for σ_{NF} [...we may have to allow for these elasticities to vary by group ...]. When existing varieties are close substitutes to new or disappearing varieties, changes in variety will not have a large effect on the difference between π_I and $COLI_I$. By contrast, when σ_F is small, varieties are not close substitutes, $\frac{1}{\sigma_F - 1}$ is high, and therefore new varieties are very valuable and disappearing varieties are very costly. In this case, the conventional price index is not appropriate.

We can now formally see in (8) the two main assumptions that we relaxed relative to standard official measures of inflation. The first difference with a standard representative agent setup is that the inflation of common goods over time, π_I, has weights that depend on the group *I*. Second, the last two terms in (8) allow for new and disappearing products to impact income groups differently.

IV. Inflation Rates by Income Groups, 1994 – 2005

(i) Food 1994-2005

We use the food database to compute the common goods food inflation rate, π_{FI}, by quintile for 1994 to 2005. For 1998 – 2005 we use precisely the methodology detailed in Section III. Prior to 1998 the lack of household detail prevents us from implementing the Sato (1976) and Vartia (1976) ideal weights for the common-goods price index, so instead we use simple expenditure shares by quintile to weight price changes for 1994-1998. We report π_{FI} in the first three columns of Table 2A. For 1994-2005, π_{FI} ranged from 9.8 percent for the bottom quintile to 14.4 percent for the top quintile and 15.6 percent for the top 5 percent of households.

We calculate the new goods bias for 1998-2005 from the food database, and use the median substitution elasticity of 11.5 from Broda and Weinstein 2007 for σ_F and σ_{NF}. The results in column 5 of Table 2A range from -3.0 percent for the lowest quintile to -3.8 percent for the highest quintile. New goods are therefore contributing substantially to the well-being of all groups, by 0.5 percent annually on average. The lack of household detail prior to 1998 prevents us from calculating the new goods bias by group for 1994-1998. The new goods bias for all consumers for 1994-1999 calculated using food modules in the aggregate nondurable database follows roughly the same annual rate as 1998-2005, at -0.6 percent per annum. For the purpose of our combined inflation rate calculations we will for now simply extrapolate the 1998-2005 bias for each quintile. Our new goods bias estimates for food appear in columns 4 to 6 of Table 2A.

Combining the common goods and bias estimates for food yields the overall food inflation rate for 1994-2005, which ranges from 4.8 percent for the lowest quintile to 7.6 percent for the highest quintile and 8.4 percent for the top 5 percent of households. We report the food results more flexibly in Figure 6 - by percentile. The inflation rate for percentile P is calculated by grouping households between percentiles P-5 to P+5. At the upper reaches of the distribution we have to modify this slightly, so that by the 99th percentile we are simply looking at the top 2 percent of households. The food inflation rate for 1994-2005 increases steadily through the income distribution, from just 3 percent at the bottom of the distribution to 10 percent near the top - ignoring the even higher spike for the 99th percentile.

(ii) Nonfood 1994-2003

We use the nondurable database to compute the nonfood inflation rate, π_{NFI}, by group I for 1994-2003. Since we only have one quarter of household-level data for nonfood items - 3500 households in 2003Q4 - we again employ simple expenditure weights by quintile in 2003 to calculate the nonfood common goods inflation rate for 1994-2003. Table 2B shows that nonfood common goods inflation varies from a low of 10.1 percent for the lowest quintile to 20.7 percent for the highest quintile and 24.1 percent for the top 5 percent of households in our sample. Until we obtain more recent data (coming in Fall 2009), we simply extrapolate those results for 2004 and 2005.

The lack of a second year of household data for nonfood items prevents us from completing calculations of the new goods bias by income group. But we can still say something meaningful. For 1994 to 2003 the "aggregate" data for nonfood items shows the average new goods bias to be -6.3 percent, or -0.7 percent annually. We can use the single quarter of household data to show the propensity of new goods to appear in the expenditures of each income group. The results presented in Figure 7 are a little surprising - new nonfood goods show up disproportionately in the expenditures of poorer households. Almost 70 percent of nonfood expenditures by the poor in 2003 were on products that did not exist in 1994. For the top quintile, this proportion is under 50 percent. Many such new goods are simply inexpensive items sold in mass-merchandise stores. It is therefore likely that nonfood new goods in our sample are as beneficial to the poor as to the rich. We simply set the bias for each group equal to the average bias. Pending further data (coming in Fall 2009) we simply extrapolate the 1994-2003 results to 2005 to match the Food sample. In Table 2C we also present results just for 1994-2003 where we have simply dropped the last two years of food data to align that sample with the nonfood sample - the same picture emerges.

We again present the results more flexibly in Figure 8. The estimated nonfood inflation rate is more volatile for two reasons. The main reason is the smaller amount of expenditures in the household-level nonfood sample - our weights are derived from just 3 months of expenditures for 3500 households. This problem will disappear with the forthcoming data. The second reason for volatility is that nonfood purchases can be more lumpy - contributing to sample variance. Yet the trend seems clear - again the inflation rate is higher for wealthier households. It averages about 5 percent at the bottom of the sample, 10 percent in the middle and 15 percent for households that spend the most.

(iii) Combined Cost of Living Index ($COLI_I$)

Our disaggregated household expenditure data for 3500 households in 2003 includes expenditures on both food and nonfood items. We use this sample to construct weights on the share of food versus nonfood items across expenditure groups (Figure 5C and the last column of Table 2C). Not surprisingly, poorer households spend proportionally more on food (73.3 percent) than do richer households (56.7 percent). We use these weights α_I to construct the cost of living index for all items in our sample using Equations (7) and (8). The results for 1994-2005 are in Table 2C and Figure 9. For the poorest quintile, inflation was just 4.5 percent, while for the top quintile it was 11.3 percent, and 13.6 percent for the top 5 percent of households in our sample.

If we focus on the 1994-2003 data to eliminate the small but crude extrapolations for nonfood items, the story is much the same. Inflation for 1994-2003 was 3.9 percent for the poorest quintile, 9.7 percent for the richest quintile and 11.7 percent for the top 5 percent (Table 2C, 6th column). Inflation for rich households is rising roughly 0.6 per cent per annum faster than for poor households. The implications of this for long-term income growth and inequality measures are potentially profound. How profound depends on how extensively this basic fact applies to expenditures outside of our sample.

V. Why Our Results Extend Beyond Our Sample

The proximate force driving our result is that poorer consumers tend to purchase products that have been rising in price less rapidly. Since poorer consumers also systematically purchase less expensive goods (Figure 4A), this suggests that the dispersion of prices of similarly defined goods (such as an ounce of milk) has likely been increasing over time. This turns out to be true in both our food sample and our nonfood sample. We calculate the unit value for each UPC at whatever frequency we observe it, for example, the price of an ounce of milk. For each module - quantity-unit pair (such as Milk, measured in ounces), we rank purchases by unit value. We then calculate percentiles of the distribution of unit prices for each module - quantity-unit pair. The p-th percentile is the price where p-percent of the goods by value (not by quantity) sell at or beneath that price. The use of value rather than quantity to define the percentiles is to reduce problems associated with measurement error in quantity in an international trade database we soon use. Figure 10A shows the unit price of food items at the 50th percentile has been rising relative to the 10th percentile, and that prices at the 90th percentile have been rising relative to the 50th. Figure 10B shows rising price dispersion for nonfood products in our sample.

This rising price dispersion might let the poor experience a lower inflation rate through two mechanisms. Firstly, through purchasing lower unit value items they can expand the range of goods they purchase in an important way. They might now only purchase the cheapest gallon of milk at Walmart, but the savings from doing this allow them to diversify their purchases of other goods - even if they are purchasing fewer UPCs overall. Table 3 shows that poorer households have been diversifying their food expenditures. The ten most important food modules absorbed 28 percent of the lowest quintile's food expenditures in 1998, but only 24 percent in 2005. Secondly, if poorer households are relatively less concerned about the finer qualities of products then they will be more willing to substitute away from products with rapidly rising prices. The "substitution bias" will be higher for poorer households.

Wherever we see a pattern of rising price dispersion for similarly defined goods it is likely that the price index for those goods is rising less rapidly for poor consumers than it is for rich consumers, simply because poor consumers place relatively high value on the "primary" qualities of the good (for example, that it is a glass, that may cost $1 at the local mass-merchandise store) and relatively less weight on the "secondary" qualities of the good - the things we intuitively think of as "quality" (a hand-blown Riedel Sommelier 24% lead crystal Bordeaux stem costing $80 at your fine wine store). Data are always classified by their "primary" physical quality. The simple fact that in any category the poor systematically choose lower unit-value items tells us that they place relatively less value on secondary qualities. The relative price of these secondary qualities is what has been increasing in the ACNielsen Homescan data. All consumers value "quality", but they may value it differently. The beauty of the Homescan data is that we can make explicit adjustments for quality using simple techniques. But other publicly available data exhibits the same price dispersion pattern that is driving our results for a broader range of goods and services. The problem with this data is that quality adjustment is more difficult, but the price patterns are still very suggestive.

International trade data values and quantities have long been collected by detailed product classification. US trade data is available electronically since 1972. From 1990-2007, monthly Census Bureau DVDs "*US Imports of Merchandise*" were used to construct percentiles of the price distribution for each HTSUSA 10-digit product (typically over 3 million observations each year on around 15,000 products). These are summarized in Figure 11A, which shows a steady increase in price dispersion. Similar but more aggregate annual US Import data from 1972-1988 contain about 125,000 observations annually on around 10,000 TSUSA 7-digit products.[8] Figure 11B shows that rising price dispersion has been a feature of this data since 1972, suggesting that a price index for the poor for all goods has likely been declining relative to the price index of the rich at least since 1972.

But what can we say about services? By far the largest service category is "Shelter", absorbing one-quarter of all consumer expenditures (where shelter is defined to exclude items

[8] US trade data moves to a new classification in 1989, but the most detailed Census Bureau data DVDs were first released in 1990.

such as utilities, household supplies, furnishings and equipment). Enrico Moretti (2008) shows that the price of housing has been rising much faster for higher income earners than for lower-income earners, enough to erode about half of the rising return to college between 1980 and 2000.

[... obtain IPUMS data on self-reported house values to show rising dispersion ...]

Another large expenditure category is "Food Away from Home", absorbing 7 percent of consumer expenditures (the same as health care). We use historical Zagat guides to show that the price of a Zagat-rated restaurant rated X relative to the price of a Big Mac has....

.... More anecdotal evidence on services ....

It seems that price dispersion has been increasing for most market-provided goods and services for decades. This makes it highly likely that our formal price index results using the ACNielsen Homescan data are not peculiar to that sample, but may indeed be representative of consumer expenditures generally.

VI. Deeper Causes of Our Results - For Subsequent Research

(i) Production in and trade with developing countries

A previous version of this paper related prices by product module to rising Chinese trade by module, which we interpreted as a supply shock, and found a strong negative correlation. We have removed that from this paper as we undertake something more ambitious on that front. Our best data is the UPC-level Homescan data, but it lacks an important detail - country of origin of the product. Two teams of undergraduates are scanning UPC codes and entering country of origin detail - which can be determined for most products, especially nonfood items. We expect to be able to say much more on how trade with developing countries has contributed to our results. For now we simply present two graphs.

Much of our result appears to be driven by the availability of lower cost alternatives valued by lower-income households. International trade is an increasingly important source of products sold to consumers. Trade with developing countries in particular has two pronounced patterns. Firstly, developing countries supply inexpensive items. Figure 12A shows this for China. Every shipment in every HTSUSA 10-digit commodity in 2005 has been sorted into unit-price deciles - 10 percent by value of each commodity appears in each decile. In 2005 China exported 20 percent of the items in the lowest two deciles, but just over 5 percent in the highest two deciles. The second fact is that developing countries are an increasingly important source of US imports - well over 50 percent of non-energy imports in recent years compared with under 35 percent in 1990 (Figure 12B).[9]

It is highly likely that this rapidly growing source of inexpensive products has something to do with the availability of lower cost alternatives for lower-income households - especially for nonfood items. Until we incorporate country of origin detail into our Homescan data we will say no more.

(ii) Service

It is highly likely that higher income consumers fundamentally consume more service, even when purchasing "goods". A shirt at Neiman Marcus or the $200 pair of jeans at some trendy store is embodied with and sold with more service (such as design, marketing, distribution, and sales) than the shirt or the $30 pair of jeans at Target or Walmart. The productivity of some of this service may have lagged that of goods manufacture or "discount" distribution, causing a long-term rise in its relative price.

.... identify a measure of service from store data ....

(iii) Rising wage inequality

If wealthier consumers consume more skill-intensive products and services then on balance they are purchasing relatively more of the services of the skilled. Some of the rising relative price of skill will be reflected in the prices of the products and services that embody their labor.

(iv) Walmart and other modern outlets

[9] The NICS have been classified as developing for this entire period for consistency. The fact of sharply rising exports from developing countries does not depend on this.

- Figure 13A & 13B

(v) Market Segmentation / Price Discrimination

In 2005 our food database contains 61,119 UPCs (excluding random-weight products) that were sold in either a Walmart or a Wholefoods. Of these UPCs, 53,715 were sold in Walmart and 8,742 were sold in Wholefoods, with an overlap of just 1,338 UPCs sold in both stores. Just 15.3 percent of the UPCs in Wholefoods can be found in Walmart, while just 2.5 percent of Walmart UPCs can be found in Wholefoods. For the exact same UPC, the prices at Walmart are 20 percent less than at Wholefoods. For a similarly-defined food product (an ounce of milk, for example), the unit prices at Walmart are 53 percent less than at Wholefoods.

VII. Implications for Welfare Inequality and Levels

If our results are at all representative of broader consumer expenditures then they have profound implications both for welfare levels and, as a consequence, how we think about inequality. Let us first look to levels. For the bottom half of households, conventionally measured real household income grew by a modest 13 percent since 1973, or about 0.4 percent per annum (Figure 1). Richer households fared much better, with real household income at the 90th percentile increasing by 41 percent or 1.0 percent per annum. If our results are representative beyond our sample in both product space and time, then adjustments for group-specific inflation rates relative to a conventional CPI for all consumers increases the growth rates for the poorest households substantially so that households at all parts of the distribution have seen substantial increases in real income since 1973: 47 percent at the 10th percentile (1.2 percent per annum); 33 percent at the median (0.9 percent per annum); and 51 percent at the 90th percentile (1.3 percent per annum) (Figure 14A). Looking at these figures we see that the fruits of economic progress may have been much more evenly distributed than we currently believe.

The reason for these results is deep yet simple - changes in the price distribution are causing the mapping from (log) nominal income to welfare to become flatter. Imagine that in 1994 the mapping between permanent income and welfare is given by $\ln w = \ln Y$. Using census data on household income as our proxy for the distribution of income, the relationship between income and welfare in 1994 is depicted by the dashed line in Figure 14B. Now look to 2005 (solid line). The dispersion of household income has clearly increased, but the same is not true for welfare. We believe that this has implications for how we should interpret most existing studies of inequality, which focus on the horizontal axis (income distribution) and not on the vertical axis (welfare distribution).

VIII. Conclusion

Studies of economic progress have mostly abstracted from an important question - what can and what do *different* consumers purchase with their income? In a world with a stationary relative price distribution this is not an important abstraction. But the distribution of relative prices is anything but stationary, and for similarly defined, though not identical, goods and services there is more dispersion in prices than ever before. The relative prices of some qualities of products is rising. When consumers value quality differently simply because they are at different points on their Engel curve this can have important implications. It turns out that the qualities that are rising in relative price are the qualities that are consumed disproportionately by wealthier consumers. Measured against the prices of products that poorer consumers actually buy, their incomes have been rising steadily. We take this to mean that the poor, relative to their past selves, have been experiencing quite rapid progress. As a consequence, we believe that more thought may need to be placed on exactly what we care about when we study inequality, or indeed when we make any conceptually similar comparisons.

References (Incomplete)

Baldwin, Robert E. and Cain, Glen G., "Shifts in U.S. Relative Wages: the Role of Trade, Technology and Factor Endowments," NBER Working Paper no. 5934, 1997.

Bernard, Andrew B. and J. Bradford Jensen, "Exporters, Skill Upgrading and the Wage Gap," *Journal of International Economics,* 42(1997):3-32.

Bhagwati, Jagdish and Marvin H. Kosters, eds., *Trade and Wages: Leveling Wages Down?* Washington, D.C.: American Enterprise Institute, 1994.

Boskin, Michael, Ellen Dulberger, Robert Gordon, Zvi Grilliches, and Dale Jorgenson, "Toward a More Accurate Measure of the Cost of Living," Final Report to the Senate Finance Committee, December 4, 1996.

Broda, Christian and David E. Weinstein, "Variety Growth and World Welfare**",** *American Economic Review*, Vol. 94, No. 2, May 2004

Broda, Christian, David E. Weinstein, "Globalization and the Gains from Variety", *Quarterly Journal of Economics*, Vol. 121, Issue 2 - May 2006.

Broda, Christian and David E. Weinstein (2007a) "Product Creation and Destruction: Evidence and Price Implications," NBER Working Paper 13041, April 2007.

Broda, C. and D. Weinstein, "Prices, Poverty and Inequality", forthcoming AEI Well-being Series Monograph, January 2008.

Bureau of Labor Statistics, Handbook of Methods, Chapter 17: The Consumer Price Index, 1997.

Bureau of Labor Statistics, Handbook of Methods, Chapter 16: Consumer Expenditures and Income, http://stats.bls.gov/hom, 1997.

Cage, R., T. Garner, and Javier Ruiz-Castillo, "Constructing Household-Specific Consumer Price Indexes: An Analysis of Different Techniques and Methods," BLS Working Paper Number 354, March 2002.

Feenstra, Robert C., "Integration and Disintegration in the Global Economy," *Journal of Economic Perspectives,* 12(1998):31-50.

Feenstra, Robert C. and Gordon H. Hanson, "Productivity Measurement and the Impact of Trade and Technology on Wages: Estimates for the U.S., 1972-1990," *Quarterly Journal of Economics*, August , 114(1999):907-940.

Feenstra, Robert C. and Gordon H. Hanson, ""Global Production and Inequality: A Survey of Trade and Wages," in James Harrigan, ed., *Handbook of International Economics*, Basil Blackwell, 2003.

Freeman, Richard B., “Are Your Wages Set in Beijing?” *Journal of Economic Perspectives,* 9(1995):15-32.

Garner, Thesia, David Johnson, and Mary Kokoski, “An Experimental Consumer Price Index for the Poor,” *Monthly Labor Review*, September, 32–42, 1996.

Goldin, C. and L. Katz, “Long Run Changes in the US Wages Structure: Narrowing, Widening and Polarizing” NBER working paper 13568, September 2007.

Hobijn, B. and D. Lagakos, "Inflation Inequality in the United States" *Review of Income and Wealth*, Dec. 2005

Krugman, Paul, “Growing World Trade: Causes and Consequences,” *Brooking Paper on Economic Activity,* 1(1995):327-362.

Lebow, David and Jeremy Rudd, “Measurement Error in the Consumer Price Index: Where Do We Stand?” *Journal of Economic Literature*, 41, 159–201, 2003.

Leamer, Edward E., “In Search of Stolper-Samuelson Linkages between International Trade and Lower Wages,” in Susan M. Collins, ed., *Imports, Exports, and the American Worker,* Washington, D.C.: Brookings Institution Press, 1998, 141-203.

Pollak, Robert, “Group Cost of Living Indexes,” American Economic Review: Papers and Proceedings, 70, 273–8, 1980.

Prais, S. J. (1959), “Whose Cost of Living?” Review of Economic Studies, 26, 126–34

Slaughter, Matthew J. “What Are the Results of Product-Price Studies and What Can We Learn from Their Differences?” in Robert C. Feenstra, ed., *The Effects of International Trade on Wages*, Chicago: University of Chicago Press, 2000a, 129-165.

Slaughter, Matthew J., “Production Transfer within Multinational Enterprises and American Wages,” *Journal of International Economics*, 50(2000b):449-472.

Table 1A: ACNielsen "Non-Durable" Homescan Database 1994, 1999-2003

		1994	1999	2000	2001	2002	2003
# Households	Aggregate	41500	55000	55000	55000	55000	55000
# UPCs	Aggregate	532789	651343	666445	658055	690036	697272
# Modules	Aggregate	1097	1096	1095	1093	1092	1093
# Households	Household						3500
# UPCs	Household						366324
# Modules	Household						1060

Table 1B: ACNielsen "Food" Homescan Database 1994, 1999-2003

		1994	1998	1999	2000	2001	2002	2003	2004*	2005
# Households	Sample 1		7624	7124	7523	8216	8684	8833	8533	8225
# UPCs	Sample 1		237766	237302	244556	248875	257935	259802	256825	255805
# Modules	Sample 1		616	636	636	635	640	639	640	640
# Households	Samples 1+2								39572	38857
# UPCs	Samples 1+2								379202	381938
# Modules	Samples 1+2								640	640

*2004 data file for Frozen; Meat; and Produce items was corrupted. 2005 data substituted here for summary statistics.

Table 2A: Food Inflation 1994-2005

Quintile	Common 94-98	Common 98-05	Common 94-05	Bias 94-98*	Bias 98-05	Bias 94-05	Food 94-05
1	0.053	0.043	0.098	-0.017	-0.030	-0.046	0.048
2	0.057	0.042	0.102	-0.018	-0.031	-0.049	0.048
3	0.063	0.047	0.113	-0.019	-0.032	-0.050	0.057
4	0.064	0.054	0.122	-0.020	-0.034	-0.053	0.062
5	0.074	0.065	0.144	-0.022	-0.038	-0.059	0.076
Top5%	0.076	0.074	0.156	-0.023	-0.040	-0.062	0.084
ALL	0.064	0.056	0.123	-0.020	-0.034	-0.053	0.064

* Extrapolated at 1998-2005 rate.

Table 2B: Nonfood Inflation 1994-2005

Quintile	Common 94-03	Common 03-05*	Common 94-05	Bias 94-03	Bias 03-05*	Bias 94-05	Nonfood 94-05
1	0.101	0.022	0.125	-0.063	-0.014	-0.077	0.038
2	0.117	0.025	0.145	-0.063	-0.014	-0.077	0.057
3	0.157	0.033	0.195	-0.063	-0.014	-0.077	0.103
4	0.137	0.029	0.170	-0.063	-0.014	-0.077	0.080
5	0.207	0.043	0.259	-0.063	-0.014	-0.077	0.162
Top5%	0.241	0.049	0.302	-0.063	-0.014	-0.077	0.202
ALL	0.157	0.033	0.196	-0.063	-0.014	-0.077	0.104

* Extrapolated at 1994-2003 rate.

Table 2C: Combined Inflation 1994-2005

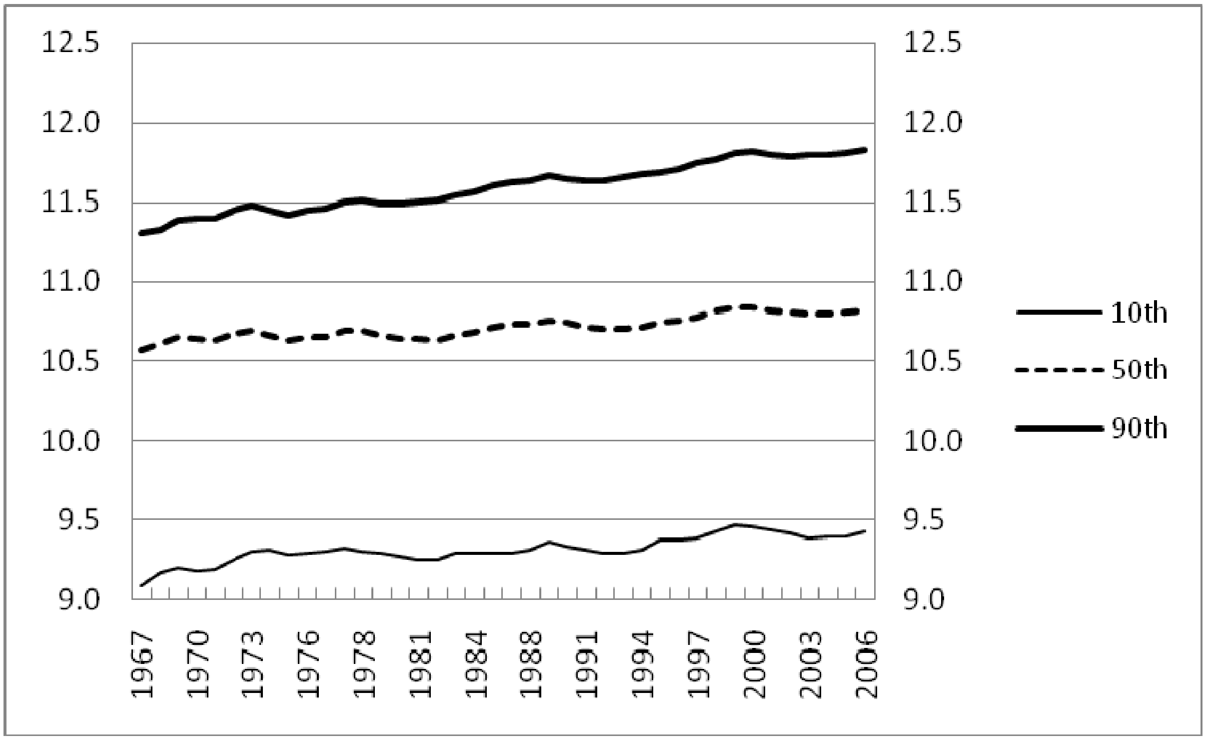
Quintile	Food 94-05	Nonfood 94-05	Combined 94-05	Food 94-03	Nonfood 94-03	Combined 94-03	Food Weight
1	0.048	0.038	0.045	0.042	0.031	0.039	0.733
2	0.048	0.057	0.051	0.039	0.046	0.041	0.693
3	0.057	0.103	0.072	0.052	0.084	0.063	0.666
4	0.062	0.080	0.069	0.057	0.065	0.060	0.629
5	0.076	0.162	0.113	0.072	0.131	0.097	0.567
Top5%	0.084	0.202	0.136	0.081	0.163	0.117	0.551
ALL	0.064	0.104	0.078	0.043	0.084	0.057	0.645

Table 3: Most Important Food Modules by Quintile, 1998 & 2005

1998 Lowest Quintile		2005 Lowest Quintile	
Share	**Module**	**Share**	**Module**
4.2%	DAIRY-MILK-REFRIGERATED	3.5%	DAIRY-MILK-REFRIGERATED
4.1%	RANDOM WEIGHT BEEF	3.4%	RANDOM WGHT BEEF
4.0%	SOFT DRINKS - CARBONATED	3.0%	SOFT DRINKS - CARBONATED
3.9%	CEREAL - READY TO EAT	2.9%	CEREAL - READY TO EAT
2.9%	RANDOM WEIGHT FRUIT	2.3%	RANDOM WGHT PREPARED FOODS
2.0%	BAKERY - BREAD - FRESH	2.2%	RANDOM WEIGHT FRUIT
1.9%	RANDOM WEIGHT CHICKEN	1.8%	BAKERY - BREAD - FRESH
1.9%	RANDOM WEIGHT VEGETABLES	1.6%	FRUIT DRINKS-OTHER CONTAINER
1.6%	COOKIES	1.6%	RANDOM WEIGHT CHICKEN
1.5%	FRUIT DRINKS-OTHER CONTAINER	1.4%	COOKIES
27.8%		**23.7%**	

1998 Highest Quintile		2005 Highest Quintile	
4.7%	RANDOM WEIGHT BEEF	4.3%	RANDOM WGHT BEEF
4.3%	RANDOM WEIGHT FRUIT	4.1%	RANDOM WGHT PREPARED FOODS
3.4%	RANDOM WEIGHT VEGETABLES	3.4%	RANDOM WEIGHT FRUIT
2.9%	DAIRY-MILK-REFRIGERATED	2.6%	RANDOM WEIGHT VEGETABLES
2.7%	SOFT DRINKS - CARBONATED	2.4%	DAIRY-MILK-REFRIGERATED
2.4%	SOFT DRINKS - LOW CALORIE	2.0%	SOFT DRINKS - LOW CALORIE
2.4%	CEREAL - READY TO EAT	2.0%	SOFT DRINKS - CARBONATED
2.1%	RANDOM WEIGHT CHICKEN	1.8%	BAKERY - BREAD - FRESH
2.0%	BAKERY - BREAD - FRESH	1.8%	CEREAL - READY TO EAT
1.6%	RANDOM WEIGHT COLD CUTS	1.6%	RANDOM WEIGHT CHICKEN
28.4%		**26.0%**	

Figure 1: Real Household Income Levels by Percentile



Source: Census; Selected Measures of Household Income Dispersion: 1967-2007

Figure 2: Hypothetical Mappings From Income To Welfare

Figure 3A: Household Income Distribution in ACNielsen Sample

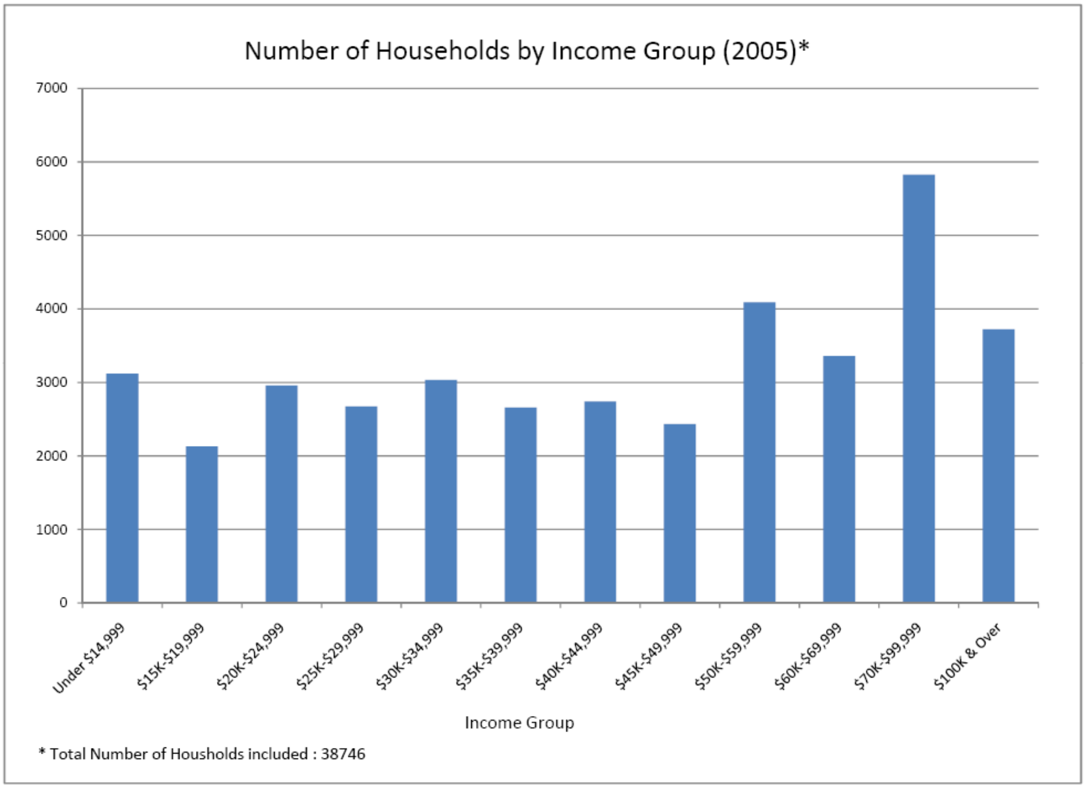


Figure 3B: Household Head Education Distribution in ACNielsen Sample

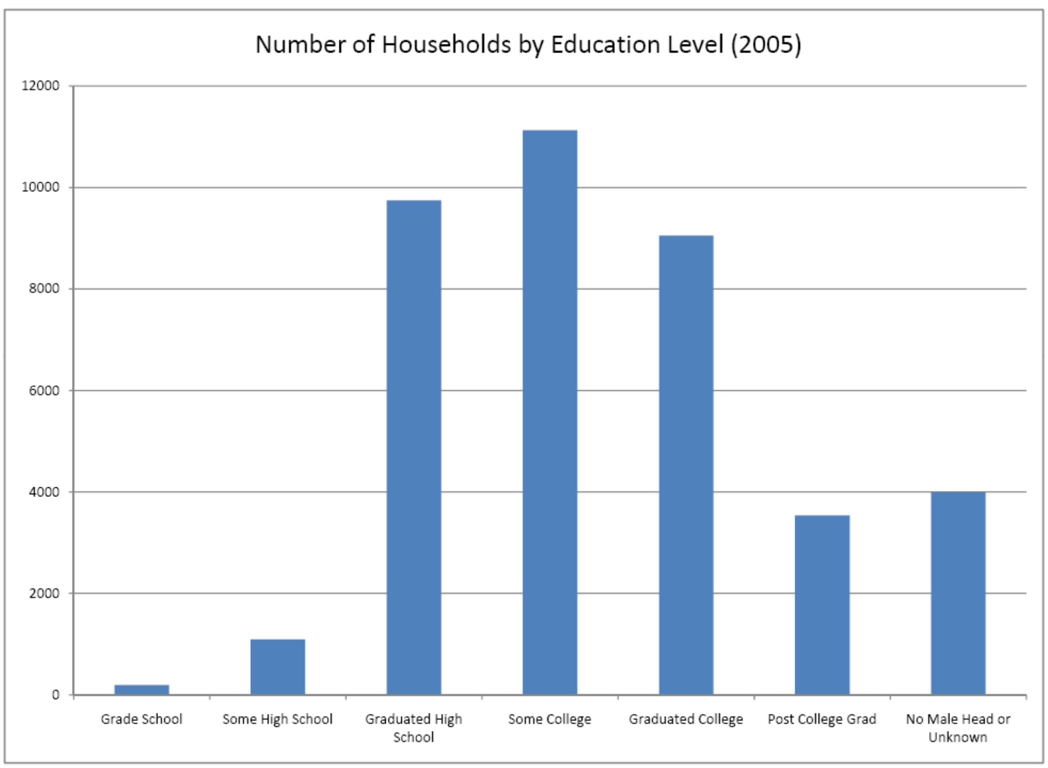


Figure 4A: Relative Prices Paid for Similarly Defined Goods

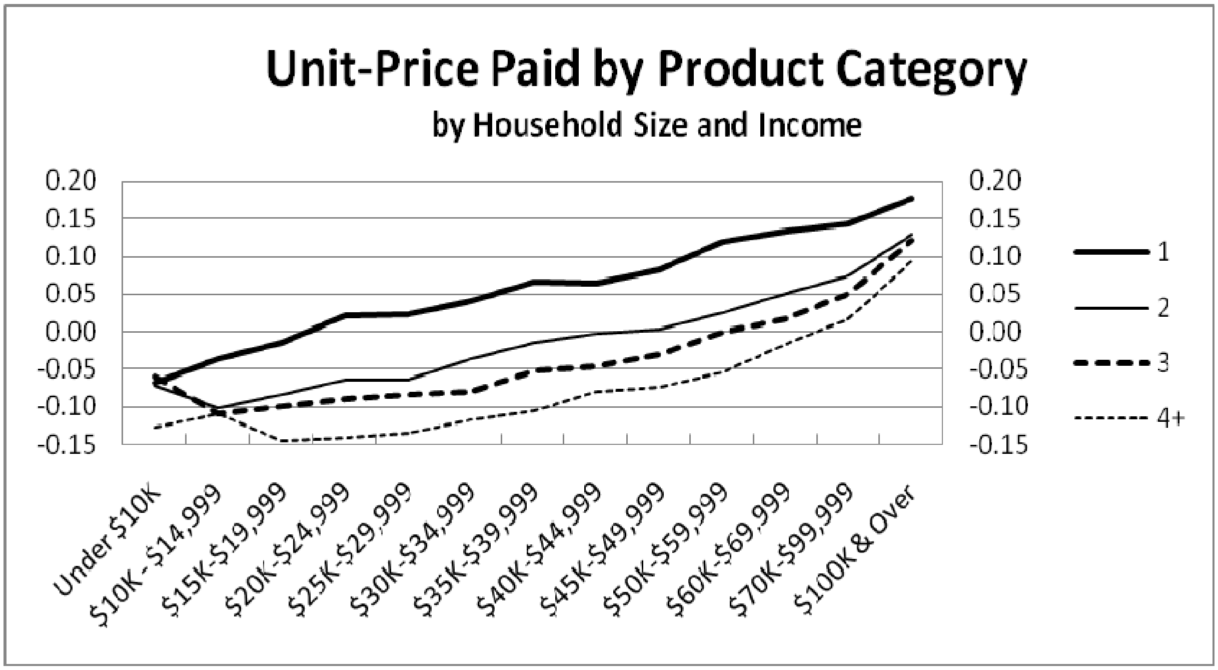


Figure 4B: Relative Prices Paid for Identical UPC

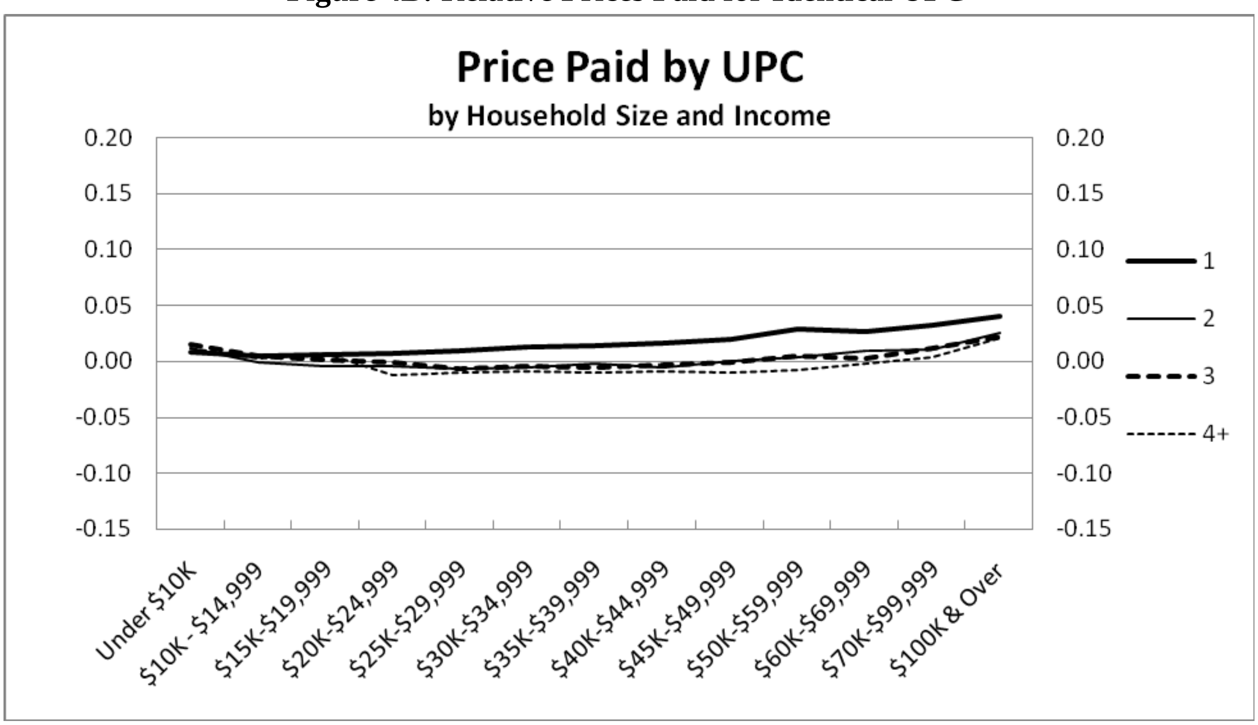


Figure 5A: Per Capita Food Expenditure in ACNielsen Sample

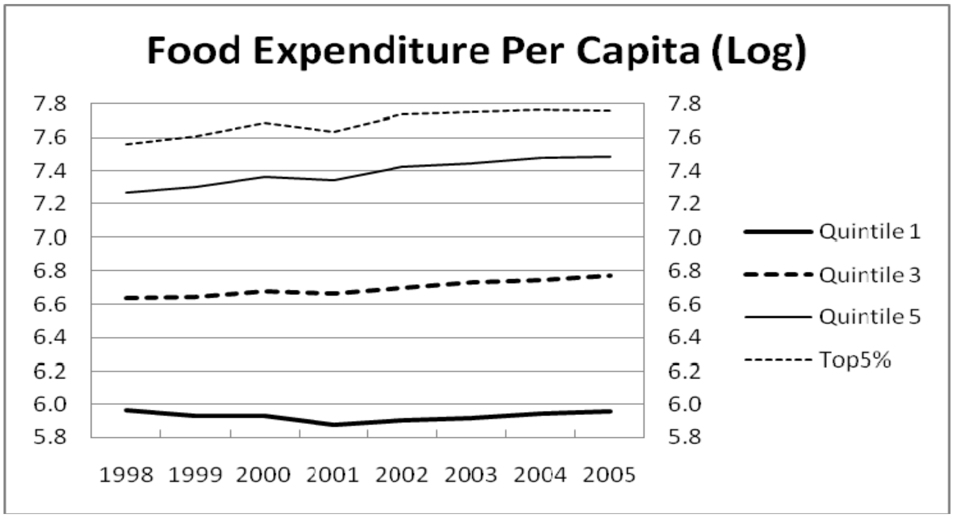


Figure 5B: UPCs per Household in ACNielsen Sample

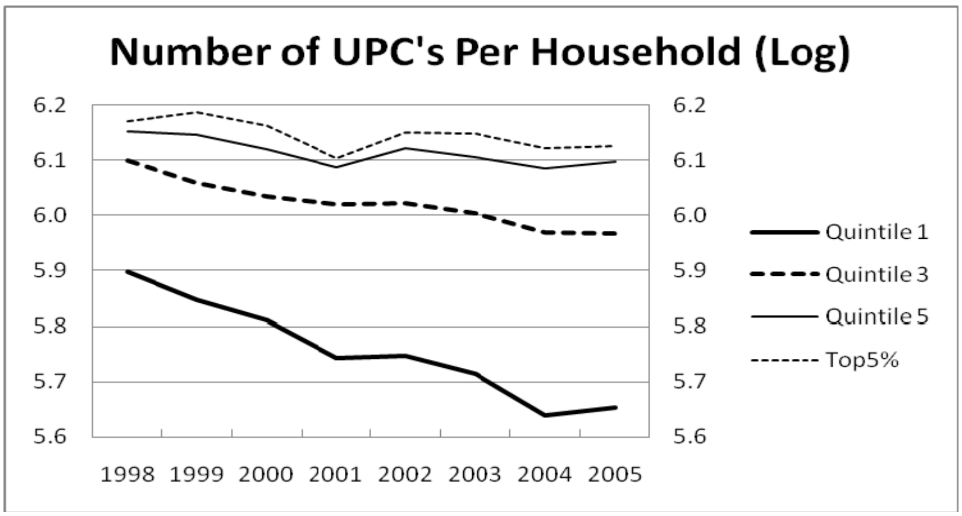


Figure 5C: Food Share in ACNielsen Sample

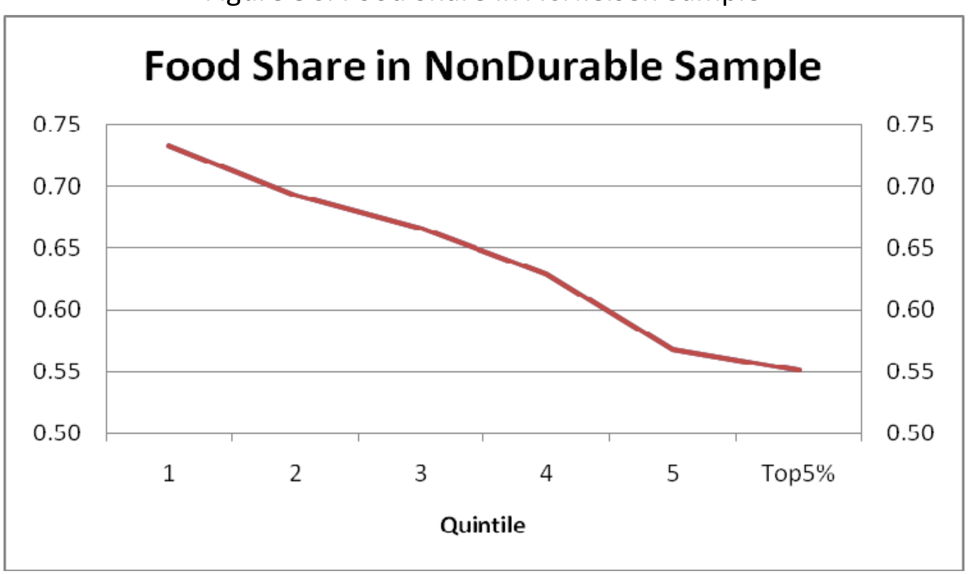


Figure 6: Food Inflation Rate by Percentile 1994-2005

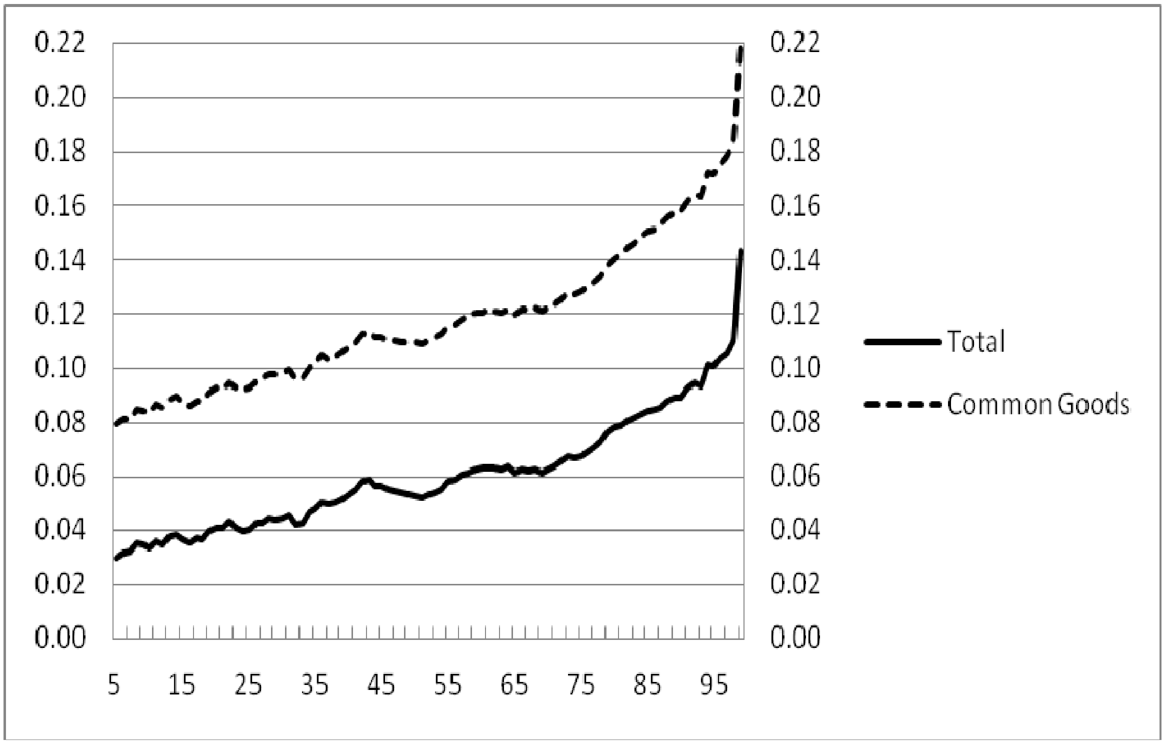


Figure 7: Nonfood Expenditure Shares on New Goods by Quintile 2003

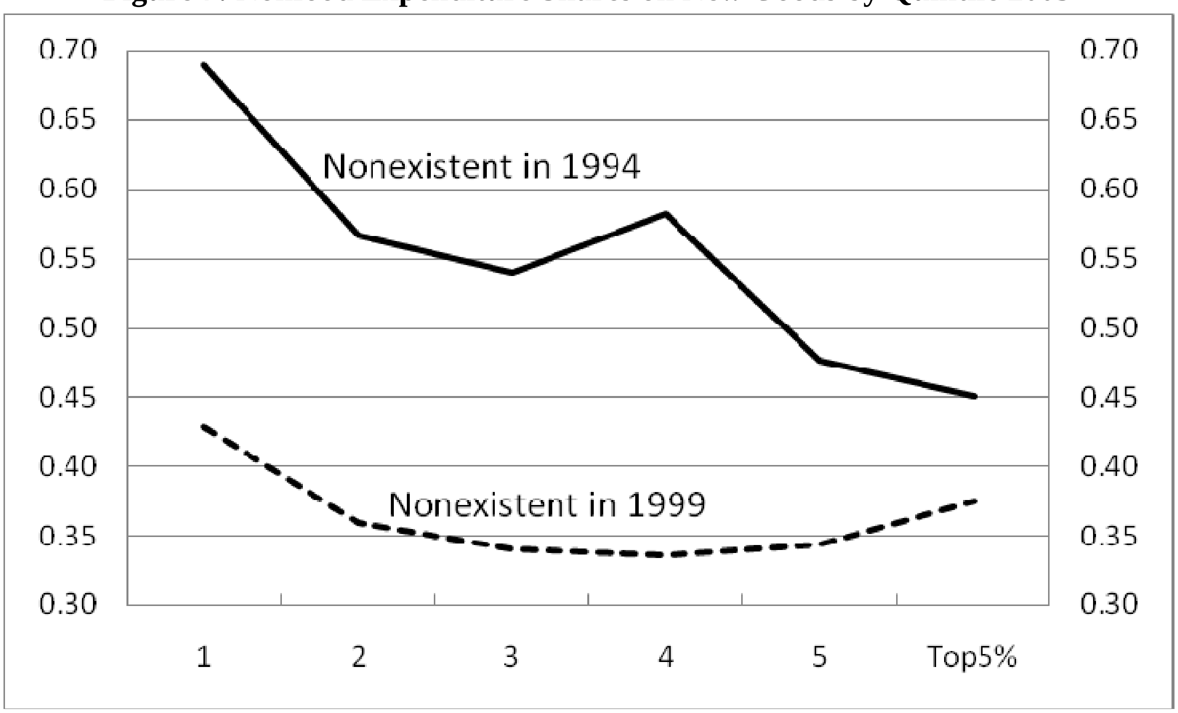


Figure 8: Non-food Inflation Rate by Percentile 1994-2005

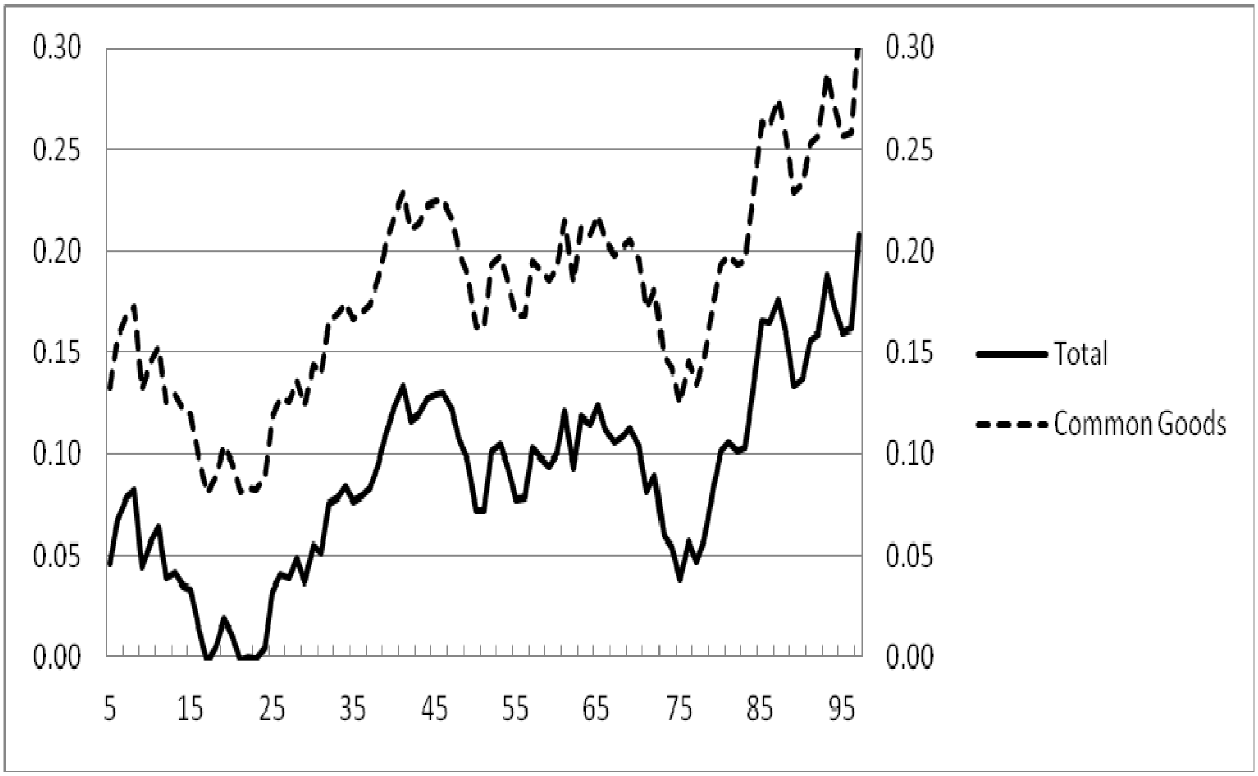


Figure 9: Cost of Living Index for Entire Sample by Quintile, 1994-2005

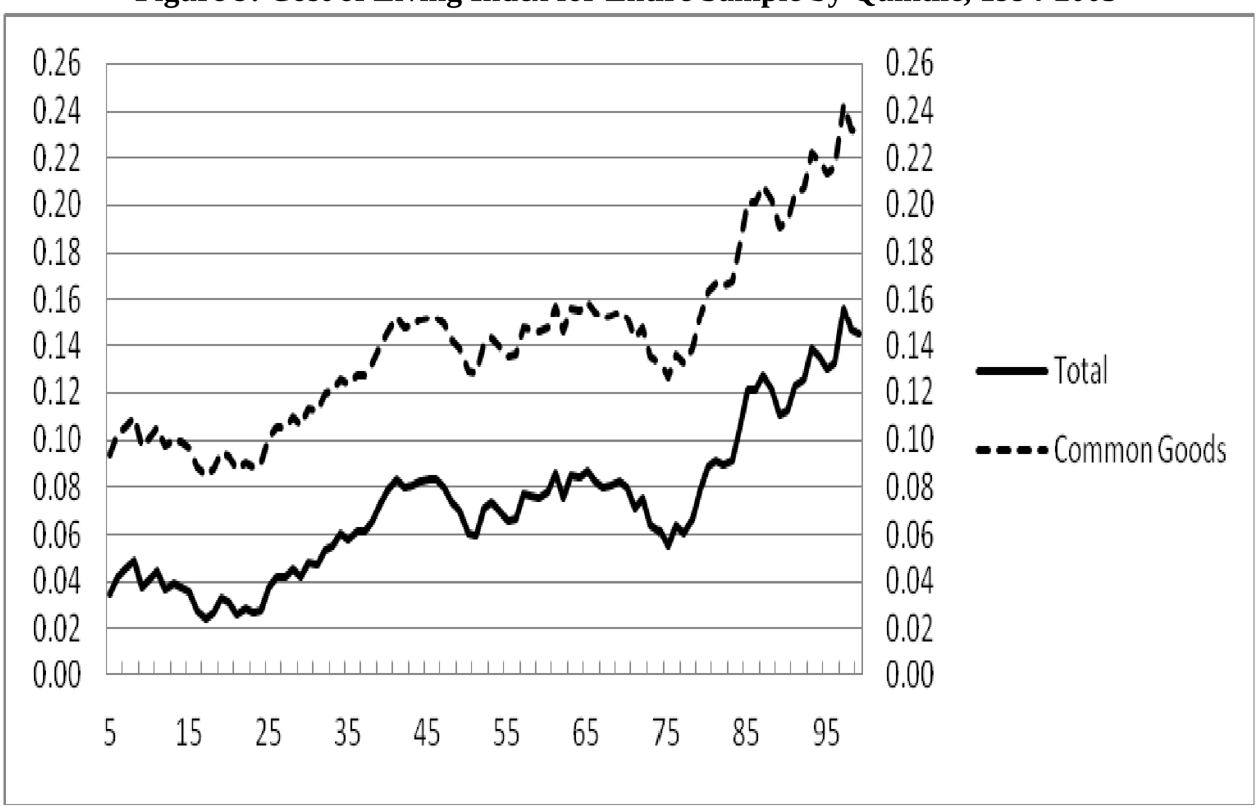


Figure 10A: Rising Price Dispersion for Food UPCs

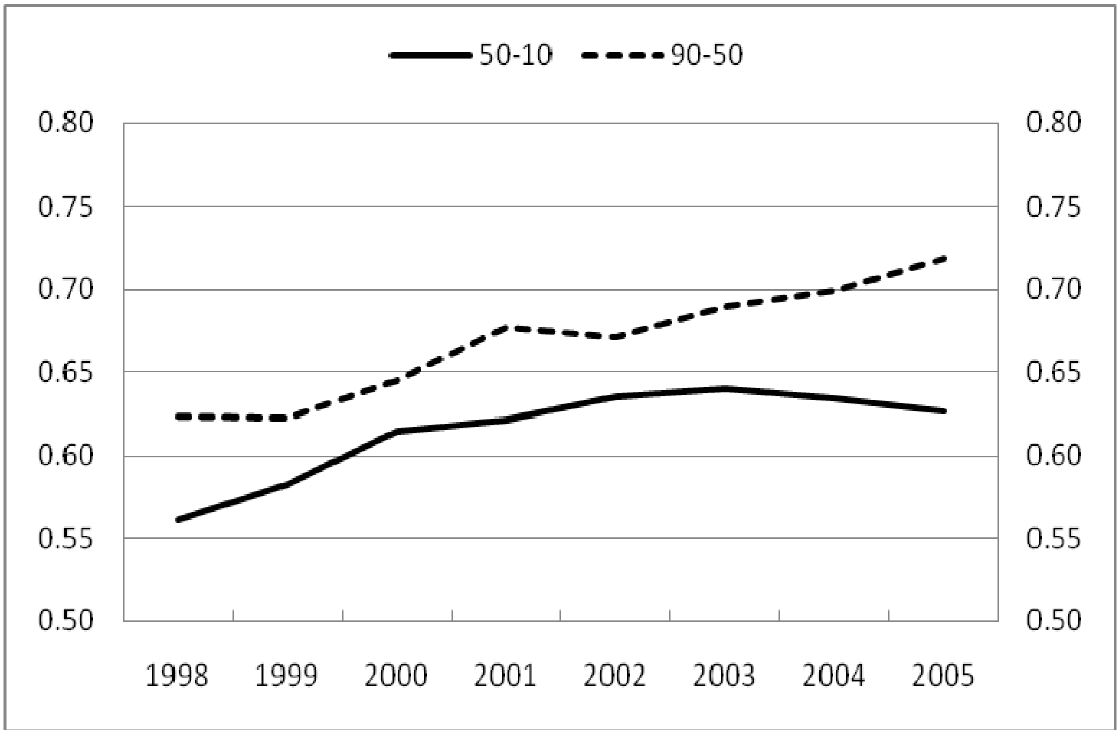


Figure 10B: Rising Price Dispersion for Non-food UPCs

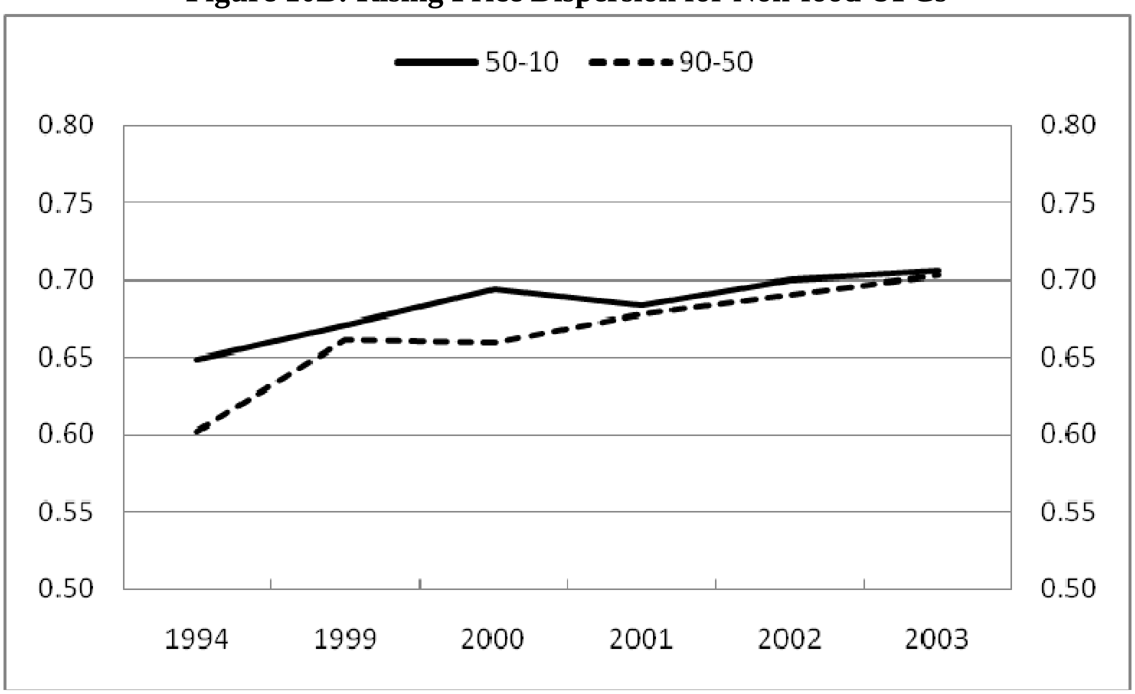


Figure 11A: Rising Price Dispersion for All US Imports 1990-2007

Figure 11B: Rising Price Dispersion for All US Imports 1972-1988

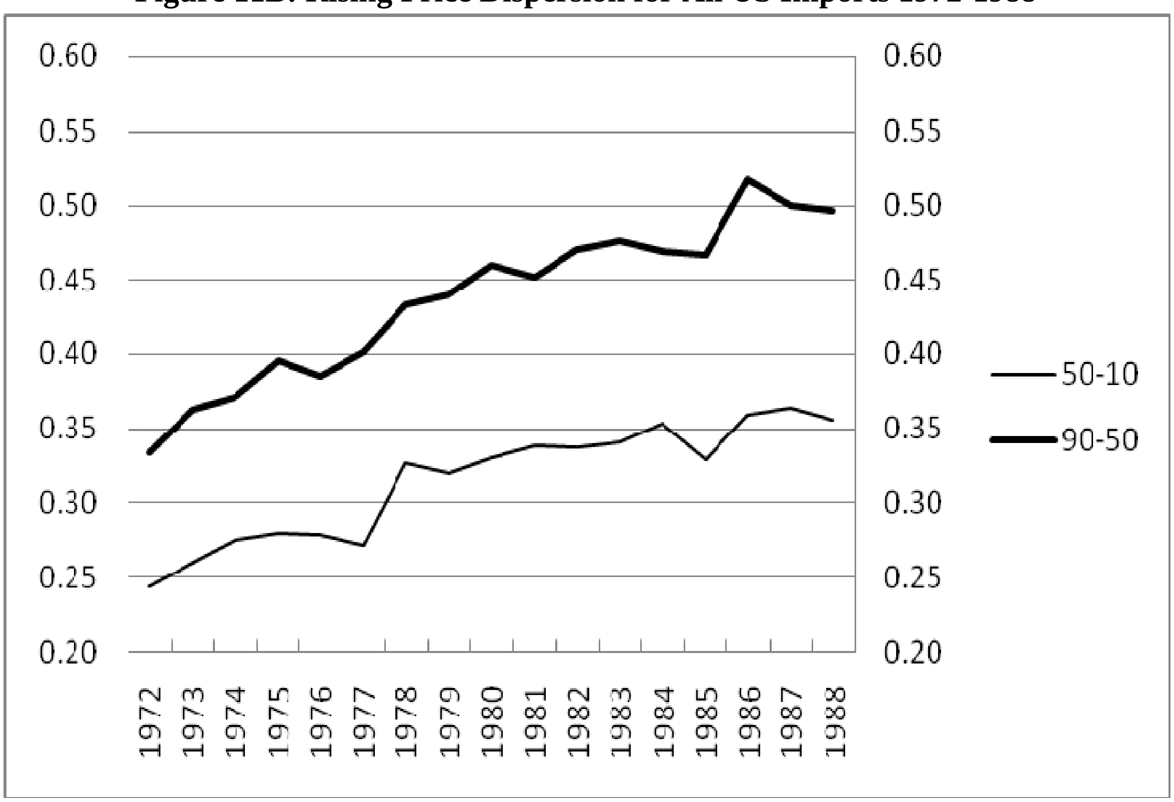


Figure 12A: Developing Countries Produce Lower-Priced Items: China's Share of US Imports by Relative Price, 2005

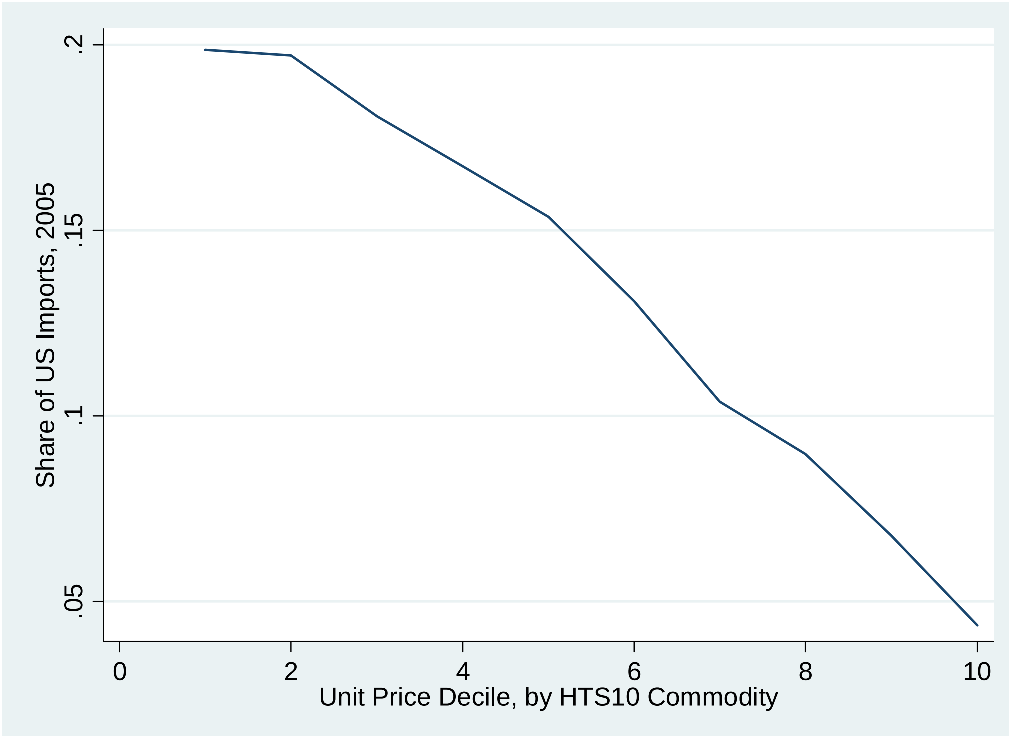


Figure 12B: Developing Countries' Share of US Non-Energy Imports, 1990-2007

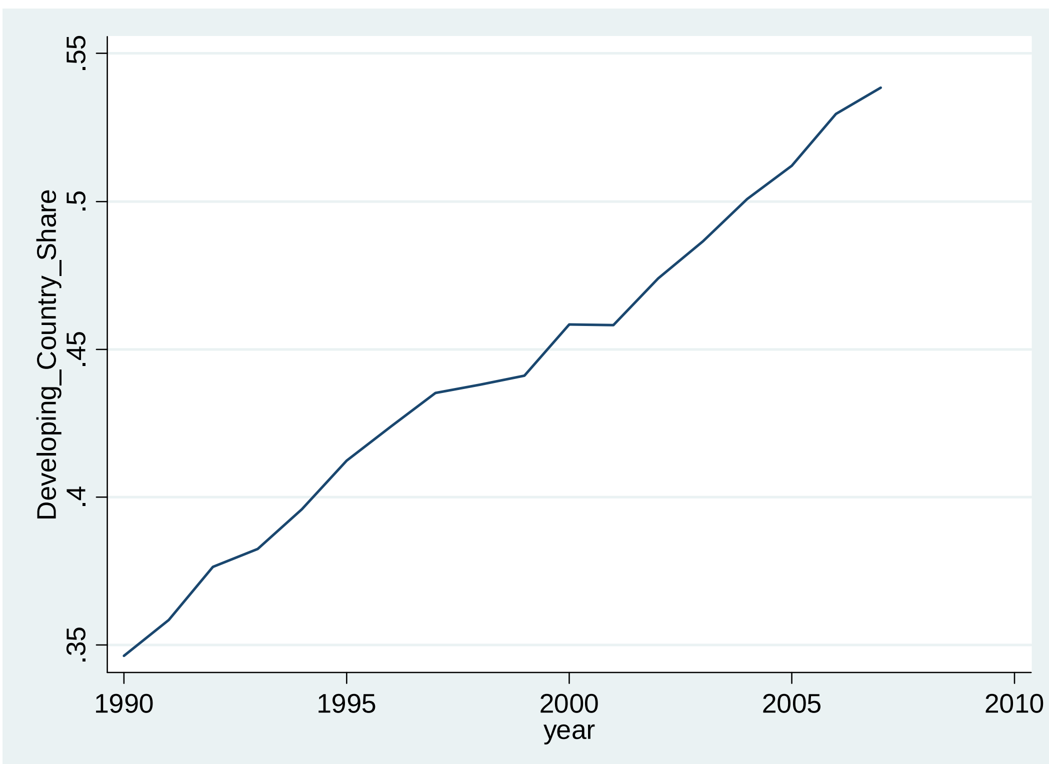


Figure 13A: Shopping for Food at Walmart by Quintile
Share of Food Expenditure

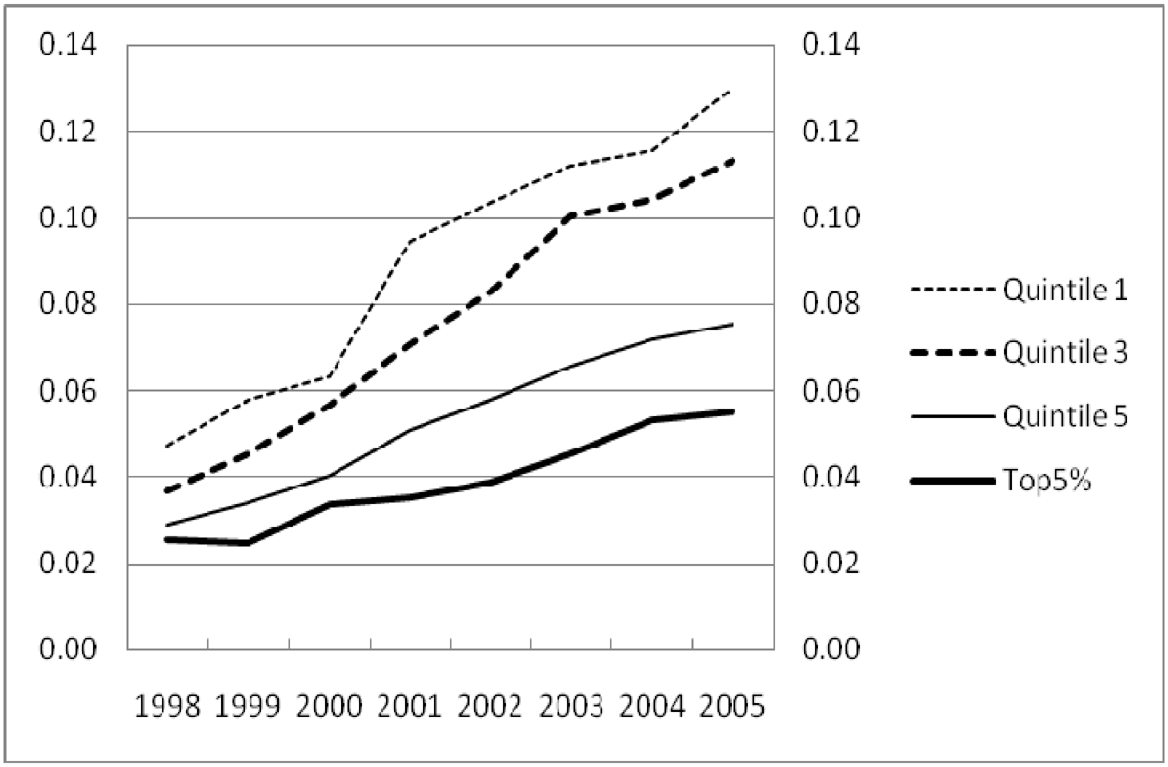


Figure 13B: Shopping for Food at Wholefoods by Quintile
Share of Food Expenditure

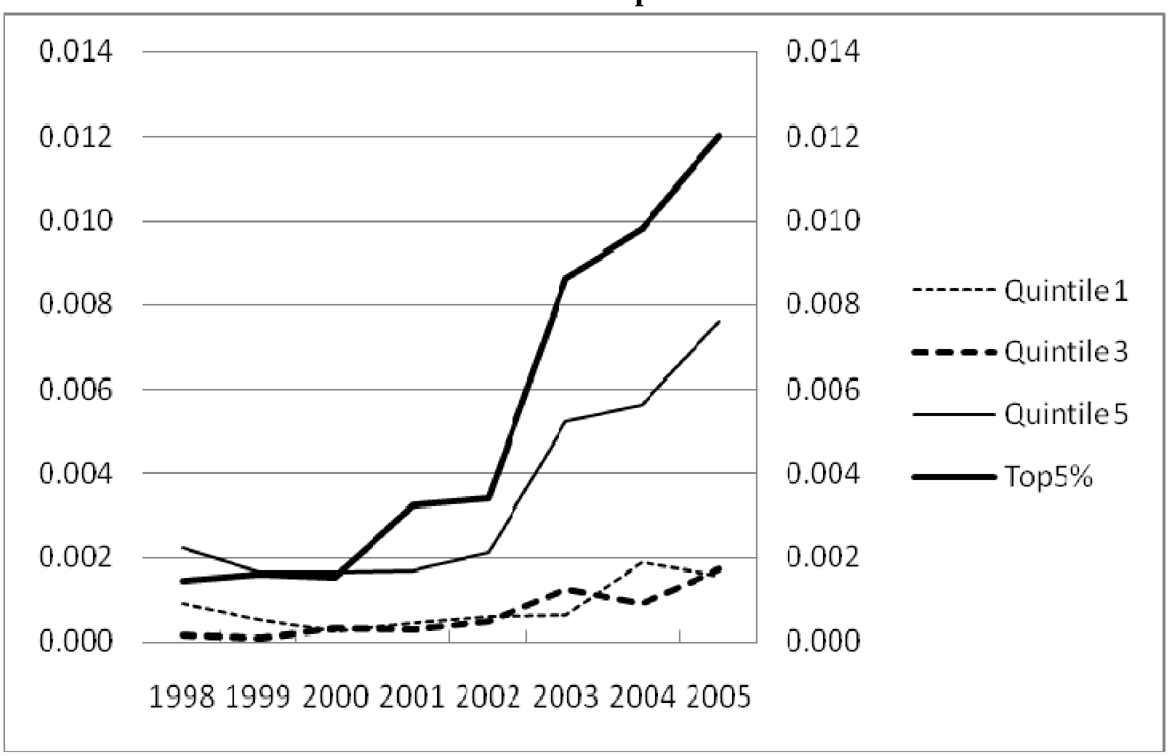


Figure 14A: Real Household Income After Extrapolating Differences In Group-Specific Inflation Rates From "Conventional" CPI

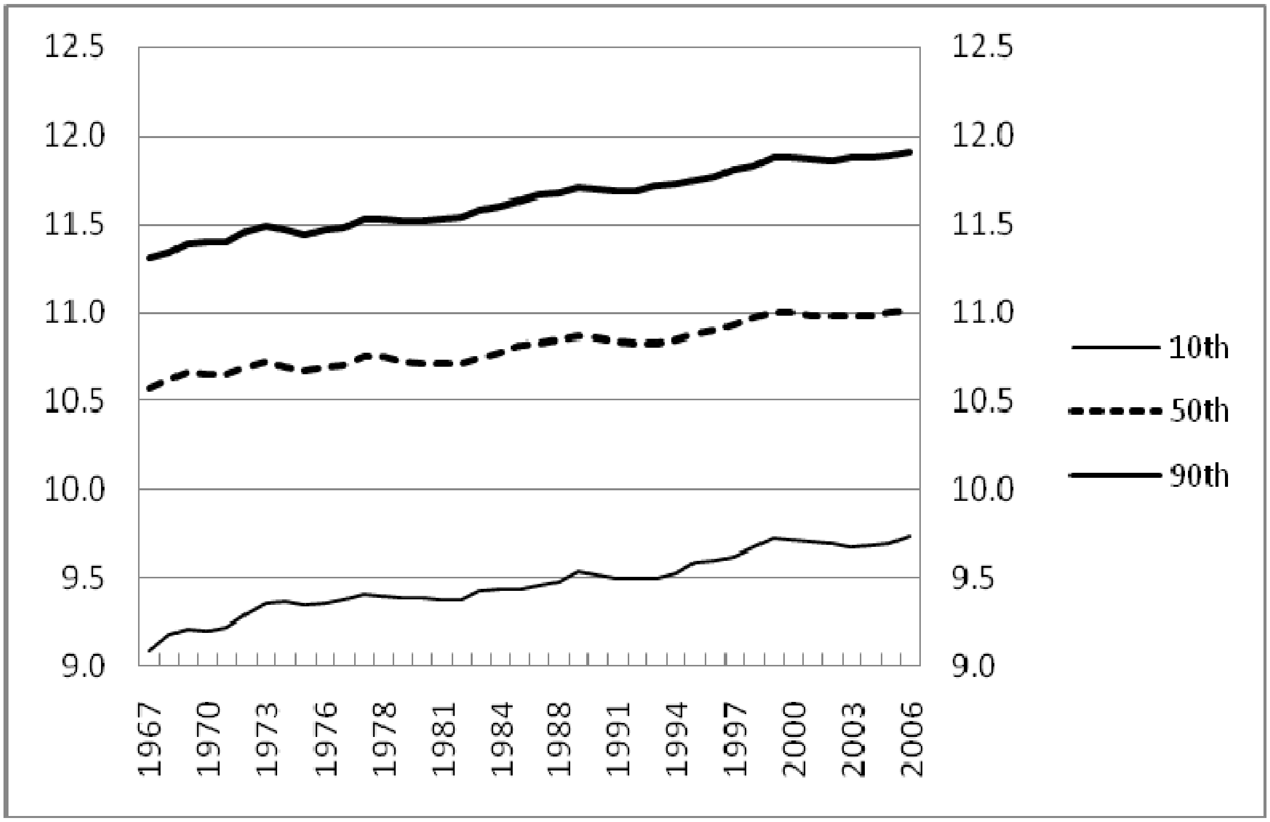


Figure 14B: Household Income and Welfare Distribution: 1994 and 2005