



WP/13/265

IMF Working Paper

Consumption Based Estimates of Urban Chinese Growth

Marcos Chamon and Irineu de Carvalho Filho

IMF Working Paper

Research Department

Consumption Based Estimates of Urban Chinese Growth

Prepared by Marcos Chamon and Irineu de Carvalho Filho*

Authorized for distribution by Atish R. Ghosh

November 2013

This Working Paper should not be reported as representing the views of the IMF.

The views expressed in this Working Paper are those of the author(s) and do not necessarily represent those of the IMF or IMF policy. Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate.

Abstract

This paper estimates the household income growth rates implied by food demand in a sample of urban Chinese households in 1993–2005. Our estimates, based on Engel curves for food consumption, indicate an average per capita income growth of 6.8 percent per year in 1993–2005. This figure is slightly larger than the 5.9 percent per year obtained by deflating nominal incomes by the CPI. We attribute this discrepancy to a small bias in the CPI, which is of a similar magnitude to the one often associated with the CPI in the United States. Our estimates indicate stronger gains among poorer households, suggesting that urban inflation up to 2005 in China was “pro-poor,” in the sense that the increase in the cost of living for poorer households was smaller than for the average one.

JEL Classification Numbers: D12; E20; I32; O10

Keywords: Household consumption; Income growth; CPI Bias

Author’s E-Mail Addresses: mchamon@imf.org; idecarvalhofilho@imf.org

* de Carvalho Filho: idecarvalhofilho@imf.org; Chamon (corresponding author): mchamon@imf.org. We would like to thank Dora Costa, Angus Deaton, Francisco H. G. Ferreira and Simon Johnson for very helpful discussions on our early work with CPI bias estimation for Latin America, and Eswar Prasad. Any remaining errors are our own.

Contents	Page
I. Introduction	3
II. Empirical Methodology.....	5
III. Data.....	8
IV. Results.....	10
Parametric Model.....	10
Semi-Parametric Model	11
Robustness	12
Distributional Implications	12
V. Conclusion	13
References.....	14
Tables	
1. Summary Statistics.....	17
2. Parametric Regression Results.....	18
3. Implications of Estimated Bias for Consumption and Income Growth in 1993–2005	19
Figures	
1. Non-Parametric Estimates of Relationship between Food Shares and Household Expenditure	20
2. Semi-Parametric Estimates of Relationship between Food Shares and Household Expenditure	20
4. Estimated Bias in 1993–2005 as a Function of CPI-Measured Real Expenditure in 2005	21
4. Estimated Cumulative Bias in China since 1993 across Different Methods and samples.....	22
5. Estimated Cumulative Bias and Confidence Intervals under Different Methodologies....	23

I. INTRODUCTION

China has enjoyed very large economic growth since it began to liberalize its economy. Real GDP per capita has increased almost 10-fold since 1978, growing at an average of 8.5 percent per year. This growth has brought increasing prosperity to Chinese households, lifting tens of millions out of poverty at unprecedented rates. Large growth rates in real per capita income are also observed in urban and rural household surveys: 7 percent per year since 1978 for urban households and 5 percent per year since 1985 for rural households.¹

This paper uses a subset of the Urban Household Survey in 1993–2005 to estimate the real income growth implied by Engel curves for food consumption, following the method developed in Nakamura (1997), Costa (2001) and Hamilton (2001) for the United States; and explored by de Carvalho Filho and Chamon (2012) for Brazil; Gibson, Stillman and Le (2008) for Russia, Barrett and Brzozowski (2010) for Australia, among others.

One of the strongest empirical regularities in economics is that the share of food in total household expenditures declines as (real) income grows (Engel’s law). We estimate a model for household-level budget share of food as a function of real expenditure, relative prices, and household characteristics, using different household expenditure survey cross-sections. Assuming nominal expenditure is measured accurately and preferences are stable, we attribute the difference between the real expenditure growth based on our estimates and the “headline” real expenditure growth obtained by deflating nominal income by the Consumer Price Index (CPI) to measurement error in the latter.

These estimates are particularly interesting in the case of China for a number of reasons. First, it yields real income growth figures that are unrelated to those of standard methods, providing an interesting cross-check. That is particularly relevant in the Chinese context as the reliability of Chinese statistics have generated heated debate among researchers. Some have raised a number of concerns that growth may be overestimated (e.g. Young 2003, and Maddison 2006), while others have argued these concerns are exaggerated and that the statistics are not biased (e.g. Holz 2006a and 2006b, and Chow 2006). Our findings suggest that the urban CPI overstated the true cost of living (so actual real income growth is even higher than indicated by official statistics), but that bias is only about 1 percent per year, which is in line with bias estimates for the U.S.

¹ The discrepancy between GDP growth and income growth in the household surveys can be partly attributed to a declining share of household income in GDP. The household’s share in total GDP has been declining over time, which can help explain the gap between the growth rates in the household surveys and the GDP figures, which can also be caused by other measurement problems (Deaton and Kozel, 2005).

Second, our estimates are not sensitive to errors in price deflators. In fact, the method was developed as a way to infer an implicit price deflator which was then used to gauge the potential error in the standard Consumer Price Index (CPI).

Finally, our estimates contribute to the debate about income inequality in China. Much of that literature has focused on the urban-rural gap (e.g. Ravallion and Chen 2007, Almås and Auglænd Johnsen, 2012, Montalvo and Ravallion 2010). But rising inequality has also been documented even within urban China (e.g. Cai, Chen, and Zhou 2010). While our paper cannot inform urban vs rural comparisons, since it focuses only on urban households, it sheds new light on the inequality debate within urban China. The method can be used to estimate income growth rates at different points of the distribution implicitly using income-specific price deflators. This can control for whether inflation was higher or lower for poor households, yielding insights that could not be obtained just by deflating incomes of the rich and the poor by the same inflation rate.² Our findings confirm the strong growth in income among the Chinese households. In fact, if anything the growth has been slightly stronger than official statistics indicate. Moreover, our estimates indicate stronger growth among the poorer households until 2005.³

Using this method, Costa (2001) and Hamilton (2001) estimate real household income growth in the U.S. since the 1980s to be roughly 1 percent per year higher than implied by nominal income deflated by the CPI. Their estimates are similar to those of the Boskin Commission, which estimated CPI bias at 1.1 percentage points per year in 1995–96 (Boskin and others, 1996). Our estimates for China point to a similar figure, which is remarkable given the challenges of measuring prices in an economy undergoing rapid transformation.⁴

There is an extensive literature arguing that changes in the CPI overestimate the increase in cost of living in the United States.⁵ The main sources of bias include the late introduction of new goods into the CPI basket, failure to (fully) account for improvements in quality, and consumer substitution. Any of these channels could account for the small bias estimated.

² Please note that our sample ends in 2005, and it is possible that relative price changes since then (notably higher food prices) may have disproportionately affected poorer households.

³ The CPI computed by statistical agencies can be interpreted as a weighted average of household price indexes, the weight of each household determined by its total expenditures. Therefore, CPIs tend to track more closely the evolution of the price of more wealthy households (Ley, 2001). The importance for inequality of differences across the income spectrum in the inflation faced by households is highlighted by Goñi, López and Servén (2006) in the Latin American context.

⁴ In a middle income country context, de Carvalho Filho and Chamon (2012) estimated CPI bias of about 3 percentage points per year in Brazil and Mexico.

⁵ For an overview of this literature, see National Research Council (2002), Hausman (2003) and Lebow and Rudd (2003).

There are a few papers that have used Engel curves to estimate CPI bias in other countries. Among advanced economies, in addition to the Costa (2001) and Hamilton (2001) papers on the U.S., Gibson and Scobie (2010) estimate a bias of one percent per year in New Zealand in 1984–2001, and Beatty and Larsen (2005) estimate a bias of 1.5 percent per year in Canada during 1978–2000. Among developing countries, Gibson, Stillman and Le (2008) estimated CPI bias in Russia during 1994–2001 to be about 1 percent per month.⁶ De Carvalho Filho and Chamon (2012) estimated the bias in Brazil 1987–2002 and Mexico 1984–2004 to be about 3 percent per year (with the large bias being attributed to one-off effects of the trade reforms). Contemporaneous papers on China (both of which cite the present paper) include Cook (2013), who uses the data from the Chinese Household Income Project and estimates a bias similar to ours, and Nakamura, Steinsson and Liu (2013) who use aggregate data and find that inflation was underestimated in the 1990s but overestimated in the 2000s.

The remainder of this paper is organized as follows. Section II describes the methodology. Sections III presents the results and Section V concludes.

II. EMPIRICAL METHODOLOGY

This section uses the same approach from de Carvalho Filho and Chamon (2012), which builds on the methods developed in Nakamura (1997), Costa (2001) and Hamilton (2001).

We start with the demand function for food that emerges from Deaton and Muellbauer's (1980) Almost Ideal Demand System:

$$\begin{aligned}
 w_{i,j,t} = & \phi + \gamma(\ln P_{F,j,t} - \ln P_{N,j,t}) \\
 & + \beta(\ln Y_{i,j,t} - \ln P_{G,j,t}) \\
 & + \sum_x \theta_x \mathbf{X}_{i,j,t} + \mu_{i,j,t},
 \end{aligned} \tag{1}$$

where the subscripts refer to household i , region j , and period t ; w is the share of food in total household expenditures; P_F , P_N and P_G are the true but unobservable price indices of food, nonfood and the general index for all goods; Y is the household's nominal expenditure; \mathbf{X} is a vector of household characteristics; and μ is the residual. A negative β characterizes a necessity good while a positive β characterizes a luxury good. The true price index P_G is measured with CPI error. Let $\Pi_{G,j,t}$ denote the percent cumulative increase in the CPI

⁶ It is possible that challenges related to deflating past expenditures in a high-inflation environment have caused an overestimation of the food budget share during years of high inflation (since typically the recall window to measure non-food expenditures is longer), thereby leading to implausibly large CPI bias estimates in Russia. It is also possible that estimating an expenditure level-specific bias and weighting those estimates by household expenditure as we do would lower their bias estimates for Russia.

measured price and $E_{G,j,t}$ denote the percent cumulative measurement error from period 0 to period t , for food, nonfood or all goods, as indicated by the subscript. Equation (1) can be rewritten as:

$$\begin{aligned}
w_{i,j,t} = & \phi + \gamma(\ln(1 + \Pi_{F,j,t}) - \ln(1 + \Pi_{N,j,t})) \\
& + \beta(\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t})) \\
& + \gamma(\ln P_{F,j,0} - \ln P_{N,j,0}) - \beta \ln P_{G,j,0} \\
& + \gamma(\ln(1 + E_{F,j,t}) - \ln(1 + E_{N,j,t})) - \beta \ln(1 + E_{G,j,t}) \\
& + \sum_x \theta_x \mathbf{X}_{i,j,t} + \mu_{i,j,t},
\end{aligned} \tag{2}$$

We assume that the CPI measurement error does not vary geographically, and rewrite (2) as:

$$\begin{aligned}
w_{i,j,t} = & \phi + \gamma(\ln(1 + \Pi_{F,j,t}) - \ln(1 + \Pi_{N,j,t})) \\
& + \beta(\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t})) \\
& + \sum_j \delta_j D_j + \sum_t \delta_t D_t \\
& + \sum_x \theta_x \mathbf{X}_{i,j,t} + \mu_{i,j,t},
\end{aligned} \tag{3}$$

where D_j and D_t are regional and time dummies and:

$$\delta_j = \gamma(\ln P_{F,j,0} - \ln P_{N,j,0}) - \beta \ln P_{G,j,0} \tag{4}$$

$$\delta_t = \gamma(\ln(1 + E_{F,t}) - \ln(1 + E_{N,t})) - \beta \ln(1 + E_{G,t}) \tag{5}$$

All the terms in Equation (3) are observable and once the equation above has been estimated, we are ready to compute the cumulative CPI bias. If food and nonfood are equally biased, then:

$$\ln(1 + E_{G,t}) = -\delta_t / \beta \tag{6}$$

It seems likely that mismeasurement is less of a problem for food than for nonfood goods. As a result, to the extent that food is a necessity ($\beta < 0$) and food shares increase with the relative price of food ($\gamma > 0$), one can show that equation (6) understates the bias for small positive values of γ as in our estimates.⁷

⁷ Note that this formula yields a multiplicative bias. That is, if the change in the CPI in a given period was 10 percent and its estimated bias is 3 percent, the estimated change in the true cost of living would be 6.7 percent,

(continued...)

The parametric specification discussed above assumes that all households at a given date face the same bias. The estimation of (3) through minimization of squared errors yields an estimate of the bias for the average household. However, the actual CPI index is based on an aggregate consumption bundle that by design disproportionately represents richer households as they account for a disproportionate share of aggregate consumption. Thus, to the extent that the discrepancy between the true cost of living index and the headline CPI varies across the income distribution, there might be substantial differences between the bias facing the average household and the bias for the aggregate consumption bundle, which is the one that maps to the CPI.

The model in equation (3) can be extended to address this concern, by assuming the bias is a linear function of the log of real expenditure:

$$\ln(1 + E_{G,i,t}) = a_t + b_t (\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t}))$$

Maintaining the same assumption that food and non-food are equally biased and that the bias does not vary by region, one can estimate:

$$\begin{aligned} w_{i,j,t} = & \phi + \gamma (\ln(1 + \Pi_{F,j,t}) - \ln(1 + \Pi_{N,j,t})) \\ & + (\beta + \sum_t \lambda_t D_t) (\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t})) \\ & + \sum_j \delta_j D_j + \sum_t \delta_t D_t \\ & + \sum_x \theta_x X_{i,j,t} + \mu_{i,j,t}, \end{aligned}$$

and obtain the following expression for CPI bias at different points in the expenditure distribution.

$$\ln(1 + E_{G,i,t}) = \frac{-\delta_t}{\beta} + \frac{\lambda_t}{\beta} (\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t})) \quad (7)$$

Finally, we can also use a flexible non-parametric approach for the bias by estimating a semi-parametric version of the demand function (1), allowing for estimation of the bias at different

since $(1-0.03) \cdot 1.1 = 1.067$. If the change in the CPI was 100 percent, the estimated change in the true cost of living would be 94 percent (since $1-0.03) \cdot 2 = 1.94$.

levels of expenditure. Still under the assumption that food and non-food are equally biased and that the bias does not vary by region, we can rewrite (3) as:

$$\begin{aligned} w_{i,j,t} &= \phi + \gamma(\ln(1 + \Pi_{F,j,t}) - \ln(1 + \Pi_{N,j,t})) \\ &\quad + \sum_t D_t f_t(\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t}) - \ln(1 + E_{G,i,t})) \\ &\quad + \sum_x \theta_x \mathbf{X}_{i,j,t} + \mu_{i,j,t} \end{aligned}$$

We estimate $f_t(\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t}) - \ln(1 + E_{G,i,t}))$ non-parametrically using the differencing method proposed in Yatchew (1997). In a nutshell, that method consists of ordering households by their (CPI-measured) real income, and (higher-order) differencing the equation above so as, to an approximation, eliminate the terms involving $f()$. We are then able to estimate the parametric terms $\hat{\phi}, \hat{\gamma}, \hat{\delta}_j$ and $\hat{\theta}_x$, obtain the non-parametric term:

$$f_t(\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t}) - \ln(1 + E_{G,i,t})) = w_{i,j,t} - \hat{\phi} - \hat{\gamma}(\ln(1 + \Pi_{F,j,t}) - \ln(1 + \Pi_{N,j,t})) - \sum_x \hat{\theta}_x \mathbf{X}_{i,j,t},$$

and estimate $f_t()$ using a locally-weighted linear regression with quartic kernel weights. The bias at a given CPI-measured real income level at time t is then estimated as the increase in CPI-measured real income that would have prevented the Engel curves from shifting. That is, we solve at each expenditure level for the value of $E_{G,i,t}$ that satisfies:

$$\hat{f}_t(\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t}) - \ln(1 + E_{G,i,t})) = \hat{f}_0(\ln Y_{i,j,t} - \ln(1 + \Pi_{G,j,t}))$$

III. DATA

A subset of the Urban Household Survey (UHS) conducted by the National Bureau of Statistics (NBS) is available through the Databank for China Studies at the Chinese University of Hong Kong. That databank covers the entire UHS for 1986–1992 and a subset of 10 provinces/municipalities for 1993–1997.⁸ Data from that same subset of provinces/municipalities in 1998–2005 was made available through a special collaboration agreement with China's National Bureau of Statistics. Provincial-level CPI coverage broken down by expenditure group starts in 1993 for most of the provinces in our sample. As a result, we limit our sample to those 10 provinces/municipalities in 1993–2005.

⁸ Those provinces/municipalities are: Anhui, Beijing, Chongqin, Ganshu, Guangdong, Hubei, Jiangsu, Liaoning, Shanxi and Sichuan.

This sub-sample is fairly representative of the entire urban sample.⁹ According to national accounts data, urban households accounted for 58 percent of final household consumption expenditures in 1993, with that share rising to 73 percent by 2005.

The UHS is based on a probabilistic sample and stratified design. It provides household-level information for a number of variables, including detailed information on income and consumption expenditures. It also provides demographic and employment information of household members, living conditions and a number of other household characteristics. The data are collected over the course of the year. Households are asked to keep a record of their income and expenditures, which is collected every month by a surveyor. Table 1 reports summary statistics for our sample. The sample size increases significantly starting in 2002, when the survey instrument was refined to obtain more detailed responses to some questions and the sample size was increased.

Our measure of food consumption covers food, beverages and tobacco (which were aggregated by the NBS in a same consumption category). It includes expenditures inside and outside of the house. The food budget share is computed as a ratio of total consumption expenditures.¹⁰ Alternatively, we could have computed it as a share of a total expenditures, which in addition to consumption would include transfers made by the household, different forms of savings, and housing construction and purchase expenditures (which are quite large due to the housing reform and privatization of the housing stock). Not only were these non-consumption expenditures large (consumption expenditures were on average only about 80 percent of income), but they also increased over time (consumption as a share of income declines 7 percentage points in our sample). In our view, the share of food in consumption expenditures provided a better measure of the share of resources required to meet those basic necessities, and more reliable Engel curves. The results are similar if the food share is computed relative to a broader measure of expenditures or income.

The income variable we focus on includes labor income, property income, transfers (both social and private, including gifts), and income from household sideline production. Neither our income nor consumption measures capture the value of owner-occupied housing.¹¹ In

⁹ These provinces/municipalities are slightly richer than the others, with incomes about 10 percent higher than the average.

¹⁰ The categories covered under consumption include: food; clothing and footwear; household appliances, goods and services; medical care and health; transportation and communications; recreational educational and cultural services; housing; and sundries.

¹¹ Households report their estimate for the rental value of owner-occupied housing from 2002 onwards. It is very rare in our sample for households to live in a rented private house (most are rented public units), so we cannot meaningfully estimate the rental value of owner-occupied houses at market prices in previous years.

order to control for the effects of home ownership on the food budget share, we use dummies for the ownership/rental status of the home.

All variables are expressed on an annual basis, with nominal values deflated to 2005 prices using the respective province CPI, which is also produced by the NBS.

IV. RESULTS

We first estimate “true real income growth” (i.e., growth in income deflated by the estimated true cost of living index) based on equation (3). Our specification allows for regional variation in relative prices. The controls for household characteristics include dummies for home ownership, rental of a private home, public housing, or other condition; gender of the head of the household; presence of a spouse; whether the head of the household, the spouse, or both have labor income; the log of the household size; and the share of household members in age groups: 0 to 4; 5 to 9; 10 to 14; 15 to 19; and 20+ years old.

Table 1 provides summary statistics for our sample. The average budget share of food in consumption gradually declines from 54 percent in 1993 to 44 percent in 2005. The relative price of food fluctuates during the sample, but remains within a 15 percent band centered at its sample average. CPI-measured real consumption per capita increases 78 percent during our sample (4.9 percent per year), while income per capita increases 99 percent (5.9 percent per year). Note the substantial increase in real per capita food expenditures (even though the food share in the budget has declined), which confirms the increasing affluence of Chinese households. There has been a slight decline in household size over time. Perhaps the most striking trend in Table 1 is the rapid increase in home ownership rates, from 24 percent in 1993 to 86 percent in 2005. Much of that increase was due to the housing reform (65 percent of home owners in 2005 bought their house through the reform).

Parametric Model

Table 2 reports the linear regression results. The first two columns estimate the bias assuming it is constant across households in a same year (as in Costa 2001 and Hamilton 2001). The third and fourth columns allow the bias to vary linearly on the log of real expenditure. The coefficients on the time dummies are negative and statistically significant, suggesting that food shares declined by more than would be predicted based on relative price and household characteristics. The estimated coefficients on log consumption expenditure range from -0.13 to -0.18, which are reasonable slopes for the Engel curve. The coefficients on the log of the relative price of food are positive, as expected, and statistically significant, ranging from 0.30 to 0.37. These parameters suggest that a 1 percent increase in consumption will lower the food budget share by a similar amount as a ½ percent decline in the relative price of food. Table 2 also reports the resulting estimate for the cumulative CPI bias, $E_{G,t}$, which is negative as expected (the values reported correspond to its absolute value). Given

that negative bias, the implied gross change in the true cost of living is $(1-|E_{G,t}|)$ times the gross change in the CPI, and the resulting gross true real income growth is $1/(1-|E_{G,t}|)$ times the gross real income growth obtained by deflating nominal income by the CPI.

Since consumption expenditures are potentially measured with error, we also present estimates where income is used as an instrument. This is particularly important because attenuation bias would tend to drive down the absolute value of the coefficient on the log of expenditure, hence increasing the estimate of the CPI bias in equation (6).

In the first two columns of Table 2 the bias is constant across households in a same survey (first columns). This yields a cumulative bias estimate for the CPI deflator of 34.3 percent in 1993–2005, or 3.4 percent per year (with the CPI overstating the increase in the estimated true cost of living, and as a result understating the growth in real income). When instrumental variables are used, that bias estimate increases to 4.0 percent per year. In columns (3) and (4) the bias is allowed to vary linearly with the log of (CPI-measured) real expenditures. The bias for the average household (population weighted average) in column (3) is 2.0 percent per year, which increases to 2.3 percent per year when income is used as an instrument (column 4). The expenditure weighted averages, in columns (3) and (4) are much lower: 0.9 and 1.1 percent per year respectively.

These results indicate that the typical Chinese urban household has enjoyed a substantially higher increase in its purchasing power than the standard deflation of its expenditure by the CPI suggests. But once the estimates are weighted by household expenditure, which is the relevant weighting for comparison with aggregate figures, this effect mostly disappears.

Semi-Parametric Model

We now turn to the semi-parametric estimation of the model. Figure 1 shows the non-parametric estimates of the relationship between the food budget share and the log of real expenditure. As expected, the food budget share declines with real expenditure and the curves tend to shift downward over time (occasional upward shifts can be explained by increases in the relative price of food). We proceed to estimate this relationship semi-parametrically, using the same controls as in Table 2. Based on these semi-parametric Engel curves (Figure 2), we can estimate the CPI bias by computing the necessary change in real expenditure, at each real expenditure level that would maintain the Engel curves in the same position, as discussed in Section II.

Figure 3 shows the estimated annual bias from 1993 to 2005 as a function of headline real expenditure, and for illustration purposes, the estimated density function of the log of CPI-deflated real expenditure in 2005. The bias is higher for the poorest households, and gradually declines as real expenditure increases, becoming negative at the very upper tail of the distribution. The annualized bias for the average population-weighted household is 1.6

percent per year, whereas the expenditure weighted aggregate bias is only 0.9 percent per year. The pattern in Figure 3 confirms the results from the parametric estimation, whereby poorer households have enjoyed larger gains than those indicated by the deflation of their nominal income by the CPI. These “unmeasured” gains were smaller for richer households, and even negative for the very rich.

Robustness

Figure 4 reports the estimated biases under different methods. We present estimates using: (i) Our baseline sample and (ii) A Winsorized sample. The Winsorization sets the value of food, consumption expenditures and income for observations below the 5th and above the 95th percentile to the level of that percentile. The thick line corresponds to our preferred estimates, based on the semi-parametric estimation. Across all methods and samples, the population weighted estimates yield a much larger bias than the expenditure weighted ones. The parametric specifications that assume a constant bias across all households in a same year yield much larger estimates. The parametric specifications that allow the bias to vary linearly with the log of (CPI-measured) real expenditure yield results that are fairly comparable to the semi-parametric estimates. Since the estimated bias is notably different across levels of income when allowed to vary, we prefer to focus on those more flexible specifications.

Figure 4 plots the point estimates and the 95 percent confidence interval for the parametric specifications where the bias varies linearly on the log of (CPI-measured) real income, and for the semi-parametric specification. It also plots, for illustration purposes, a line corresponding to the cumulative effects of a bias equal to 1 percent per year, a figure often associated with the bias in the U.S. CPI. In the case of our expenditure weighted measures, the 1 percent per year line is within the confidence interval most of the time, and coincides with the end-sample point-estimates in the 3 methods.

Distributional Implications

Table 3 summarizes the implications of our bias estimates for the evolution of consumption and income for different groups. If we deflate nominal consumption and income by the CPI, their average value grows $\frac{1}{2}$ percentage point faster per year than their median. Growth rates are higher as we move from poorer to richer quintiles. Consumption growth for the richest quintile is 6.1 percent per year, whereas for the poorest one is only 2.9 percent per year. For income growth, those figures are 6.7 and 4.2 percent per year respectively. The picture reverses if we use our alternative deflator. Once that adjustment is made, consumption and income growth is stronger for the poorer quintiles. The consumption growth in the richest quintile is now 5.5 percent per year, which is lower than the 6.4 percent per year for the poorest one. For income growth those figures are 6.2 and 7.7 percent per year respectively.

Average and median consumption grow at a similar rate, at around 5.7 percent per year. The same is true for income, with average and median growth at 6.9 percent per year.

V. CONCLUSION

This paper has used household-level consumption data to estimate the path of real income growth compatible with the shrinking participation of food expenditure in total household expenditures in urban China. Our estimates suggest that the urban CPI provides a fairly accurate measure of the cost of living for the urban economy as a whole. Our expenditure weighted estimates, which is the relevant measure for comparison with the CPI, points to a discrepancy of 1 percentage point or less per year, comparable to the bias often associated with the U.S. CPI. This is a remarkable precision considering the challenges related to measuring the cost of living in a rapidly changing economy such as China. But even a fairly accurate CPI may not be appropriate for different groups of households, as consumption patterns tend to vary substantially with income. Our results suggest that the cost of living for poorer households has increased slower than the CPI, thus contributing to reduce real expenditure inequality.

Some caution is warranted when interpreting these distributional results. The survey sample only covers migrants from rural (which are among the poorest groups in urban areas) starting in 2002. Coverage of very rich households is limited (the 99th percentile of income in the 2005 sample was 120,000 Yuan), and since they represent a small share of the population our estimates are less precise at the upper tail of the distribution. Moreover, the increasing shift from public to private provision of health and education services could have affected the evolution of the food budget shares. But with these caveats in mind, this paper shows that the evolution of consumption patterns suggests that real income growth in urban China has been more equitable than it is perceived to be.

REFERENCES

- Almås. Ingvild and Åshild Auglænd Johnsen, 2012, “The Cost of Living in China: Implications for Inequality and Poverty,” Norwegian School of Economics, mimeo.
- Barrett, Garry F., Brzozowski, Matthew, 2010. “Using Engel Curves to Estimate the Bias in the Australian CPI,” *Economic Record* 86 (272), 1–14.
- Beatty, Timothy, and Erling R. Larsen, 2005, “Using Engel Curves to Estimate Bias in the Canadian CPI as a Cost of Living Index,” *Canadian Journal of Economics*, Vol. 38, No. 2, pp. 482–99.
- Boskin, Michael, Ellen Dulberger, Robert Gordon, Zvi Griliches, and Dale Jorgenson, 1996, “Toward a More Accurate Measure of the Cost of Living,” Final Report to the U.S. Senate Finance Committee (December).
- Cai, Hongbin, Yuyu Chen, and Li-An Zhou, 2010, “Income and Consumption Inequality in Urban China: 1992–2003” *Economic Development and Cultural Change*, Vol. 58(3), pp. 385-413.
- de Carvalho Filho, Irineu and Marcos Chamon, 2012, “The Myth of Post-Reform Income Stagnation: Evidence from Brazil and Mexico,” *Journal of Development Economics*, Vol. 97, No. 1, pp. 368–386.
- Chow, Gregory, 2006, “Are Chinese Official Statistics Reliable?,” *CESifo Economic Studies* 52(2): 396–414.
- Cook, Jonathan A., 2013, “Does Chinese Inflation Understate Cost of Living?,” mimeo.
- Costa, Dora, 2001, “Estimating Real Income in the United States from 1888 to 1994: Correcting CPI Bias Using Engel Curves,” *Journal of Political Economy*, Vol. 109, No. 6, pp. 1288–1310.
- Deaton, Angus, and Valerie Kozel, 2005, “Data and Dogma: The Great Indian Poverty Debate,” *The World Bank Research Observer* Vol. 20, pp.177–99.
- Deaton, Angus, and John Muellbauer, 1980, “An Almost Ideal Demand System,” *American Economic Review*, Vol. 70, No. 3, pp. 312–26.
- Gibson, John, Steven Stillman and Trinh Le, 2008, “CPI Bias and Real Living Standards in Russia During the Transition.” *Journal of Development Economics*, Vol. 87, No. 1, pp. 140–160.
- Gibson, John and Grant Scobie, 2010, “Using Engel curves to estimate CPI bias in a small, open, inflation-targeting economy,” *Applied Financial Economics*, 20(17): 1327–1335.

- Goñi, Edwin, Humberto López and Luis Servén, 2006, “Getting Real About Inequality: Evidence from Brazil, Colombia, Mexico and Peru,” World Bank Policy Research Working Paper ; 3815
- Hamilton, Bruce, 2001, “Using Engel’s Law to Estimate CPI Bias,” *American Economic Review*, Vol. 91, No. 3, pp. 619–30.
- Hausman, Jerry, 2003, “Sources of Bias and Solutions to Bias in the Consumer Price Index,” *Journal of Economic Perspectives*, Vol. 17, No. 1, pp. 23–44.
- He, Xinhua, 2010, “Noteworthy Discrepancies in China’s GDP Accounting,” *China & World Economy* 18(4): 88–102.
- Holz, Carsten A. 2006a., “China’s Reform Period Economic Growth: How Reliable Are Angus Maddison’s Estimates?,” *Review of Income and Wealth* 52(1): 85–119.
- Holz, Carsten A. 2006b., “China’s Reform Period Economic Growth: How Reliable Are Angus Maddison’s Estimates? Response to Angus Maddison’s Reply,” *Review of Income and Wealth* 52(3): 471–475.
- Lebow, David and Jeremy Rudd, 2003, “Measurement Error in the Consumer Price Index: Where do We Stand?” *Journal of Economic Literature* 41(1): 159–201.
- Ley, Eduardo, 2001, “Whose Inflation? A Characterization of the CPI Plutocratic Bias,” IMF Working Paper 01/59.
- Maddison, Angus, 2006, “Do Official Statistics Exaggerate China’s GDP Growth? A Reply to Carsten Holz,” *Review of Income and Wealth* 52(1): 121–126.
- Montalvo, Jose and Martin Ravallion, 2010, “The Pattern of Growth and Poverty Reduction in China” *Journal of Comparative Economics*, vol. 38(1), pp. 2–16.
- Nakamura, Emi, Jón Steinsson and Miao Liu, 2013, “Are Chinese Growth and Inflation Too Smooth? Evidence from Engel Curves,” mimeo.
- Nakamura, Leonard, 1997, “Is the U.S. Economy Really Growing Too Slowly? Maybe We’re Measuring Growth Wrong,” *Federal Reserve Bank of Philadelphia Business Review* (March–April), pp. 3–14.
- National Research Council, 2002, *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indices*, Charles L. Schultze and Christopher D. Mackie, eds. (Washington, DC: National Academy Press).
- Ravallion, Martin and Shaohua Chen, 2007, “China’s (Uneven) Progress against Poverty” *Journal of Development Economics*, Vol. 82(1), pp. 1–42.

Yatchew, A., 1997, "An Elementary Estimator of the Partial Linear Model," *Economics Letters*, Elsevier, vol. 57 (2), pages 135–43 (December).

Young, Alwyn, 2003. "Gold into Base Metals: Productivity Growth in the People's Republic of China during the Reform Period," *Journal of Political Economy* 111(6): 1220–1261.

Table 1. Summary Statistics

	1993	1995	1997	1999	2001	2003	2005
Share of Food in Consumption	0.54 [0.151]	0.55 [0.151]	0.52 [0.146]	0.47 [0.144]	0.43 [0.142]	0.44 [0.141]	0.44 [0.144]
Relative price of food	94.96 [4.21]	105.08 [3.328]	101.17 [3.339]	95.44 [2.43]	91.77 [1.359]	93.79 [0.81]	100 [0]
Real Expenditure on food	7380 [3687]	7582 [3949]	7817 [4353]	8113 [4404]	8533 [4902]	8190 [4714]	8841 [5001]
Real per capita expenditure on food	2374 [1172]	2458 [1241]	2515 [1307]	2672 [1398]	2875 [1603]	2796 [1621]	3091 [1773]
Real Consumption Expenditure	14073 [9302]	15808 [10889]	16724 [11724]	18189 [13105]	20365 [15932]	19975 [16198]	23348 [20147]
Real Per Capita Consumption	4487 [2801]	5089 [3410]	5360 [3573]	5987 [4242]	6835 [5173]	6800 [5585]	8139 [7374]
Ln (Consumption/CPI)	9.41 [0.519]	9.51 [0.538]	9.56 [0.554]	9.63 [0.573]	9.72 [0.61]	9.69 [0.643]	9.84 [0.65]
Ln (Current Monetary Income/CPI)	9.58 [0.504]	9.68 [0.541]	9.74 [0.559]	9.85 [0.551]	9.96 [0.607]	9.94 [0.663]	10.12 [0.662]
Ln (Household size)	1.13 [0.269]	1.12 [0.259]	1.13 [0.251]	1.11 [0.241]	1.09 [0.258]	1.08 [0.278]	1.06 [0.291]
Share ages 0 to 4	0.03 [0.09]	0.03 [0.089]	0.03 [0.084]	0.02 [0.076]	0.02 [0.074]	0.02 [0.073]	0.02 [0.069]
Share ages 5 to 9	0.07 [0.133]	0.06 [0.129]	0.05 [0.119]	0.05 [0.113]	0.04 [0.107]	0.04 [0.105]	0.04 [0.104]
Share ages 10 to 14	0.09 [0.144]	0.08 [0.141]	0.08 [0.142]	0.07 [0.138]	0.07 [0.132]	0.06 [0.124]	0.05 [0.119]
Share ages 15 to 19	0.07 [0.137]	0.07 [0.137]	0.08 [0.139]	0.08 [0.143]	0.07 [0.139]	0.07 [0.137]	0.06 [0.133]
Share ages 20 and up	0.75 [0.165]	0.76 [0.162]	0.77 [0.157]	0.78 [0.161]	0.8 [0.167]	0.82 [0.172]	0.83 [0.171]
Male head	0.69 [0.462]	0.66 [0.473]	0.65 [0.477]	0.62 [0.485]	0.67 [0.47]	0.74 [0.441]	0.73 [0.442]
Spouse present	0.67 [0.471]	0.64 [0.479]	0.63 [0.482]	0.61 [0.487]	0.66 [0.475]	0.72 [0.45]	0.71 [0.452]
Head has labor income	0.84 [0.368]	0.82 [0.381]	0.83 [0.377]	0.81 [0.393]	0.76 [0.429]	0.72 [0.447]	0.71 [0.454]
Spouse has labor income	0.74 [0.438]	0.73 [0.444]	0.73 [0.444]	0.71 [0.455]	0.63 [0.482]	0.56 [0.496]	0.54 [0.499]
Head and spouse have labor income	0.71 [0.453]	0.7 [0.46]	0.69 [0.461]	0.67 [0.471]	0.58 [0.493]	0.51 [0.5]	0.49 [0.5]
Owner occupied house	0.24 [0.429]	0.34 [0.474]	0.49 [0.5]	0.65 [0.477]	0.78 [0.417]	0.79 [0.408]	0.86 [0.344]
Rental Unit	0.02 [0.123]	0.01 [0.117]	0.01 [0.098]	0.01 [0.096]	0.01 [0.104]	0.02 [0.134]	0.02 [0.131]
Other living arrangement	0 [0.039]	0 [0.016]	0 [0.025]	0 [0.016]	0 [0.036]	0.07 [0.249]	0.02 [0.138]
Public Housing	0.74 [0.439]	0.64 [0.479]	0.5 [0.5]	0.34 [0.474]	0.21 [0.409]	0.13 [0.332]	0.1 [0.3]
N	6109	6297	6242	6295	6300	19351	21849

Notes: For presentation purposes only odd years displayed. Share of food measured relative

Table 2. Parametric Regression Results

Dependent Variable: Share of Food				
	Bias invariant on income		Bias linear on income	
	(1)	(2)	(3)	(4)
	Full-OLS	Full-IV	Full-OLS	Linear-IV
Dummy for 1995	-0.006 [0.003]	-0.007 [0.003]	-0.019 [0.042]	0.02 [0.048]
Dummy for 1997	-0.019 [0.003]	-0.021 [0.003]	-0.115 [0.042]	-0.094 [0.048]
Dummy for 1999	-0.034 [0.002]	-0.036 [0.002]	-0.25 [0.041]	-0.233 [0.047]
Dummy for 2001	-0.05 [0.002]	-0.053 [0.002]	-0.358 [0.039]	-0.321 [0.046]
Dummy for 2003	-0.061 [0.002]	-0.065 [0.002]	-0.541 [0.034]	-0.579 [0.040]
Dummy for 2005	-0.061 [0.002]	-0.066 [0.002]	-0.562 [0.034]	-0.592 [0.040]
Ln(Relative price of food)	0.301 [0.019]	0.3 [0.019]	0.362 [0.019]	0.366 [0.019]
Ln(Real Consumption)	-0.145 [0.001]	-0.134 [0.001]	-0.184 [0.003]	-0.173 [0.004]
Ln(Household size)	0.07 [0.001]	0.065 [0.001]	0.072 [0.001]	0.068 [0.001]
Number ages 0 to 4	-0.014 [0.005]	-0.014 [0.005]	-0.013 [0.005]	-0.013 [0.005]
Number ages 5 to 9	0 [0.000]	0 [0.000]	0 [0.000]	0 [0.000]
Number ages 10 to 14	0.028 [0.003]	0.028 [0.003]	0.032 [0.003]	0.032 [0.003]
Number ages 15 to 19	0.007 [0.003]	0.006 [0.003]	0.013 [0.003]	0.013 [0.003]
Number ages 20 and up	0.11 [0.003]	0.105 [0.003]	0.117 [0.003]	0.113 [0.003]
Male head	0.008 [0.003]	0.009 [0.003]	0.007 [0.003]	0.008 [0.003]
Spouse present	-0.002 [0.003]	-0.002 [0.003]	-0.001 [0.003]	-0.001 [0.003]
Head of household has labor income	-0.038 [0.001]	-0.038 [0.001]	-0.037 [0.001]	-0.037 [0.001]
Spouse has labor income	-0.027 [0.002]	-0.028 [0.002]	-0.027 [0.002]	-0.027 [0.002]
Head and spouse have labor income	0.011 [0.002]	0.009 [0.002]	0.01 [0.002]	0.008 [0.002]
Owner occupied unit	0.05 [0.003]	0.045 [0.003]	0.045 [0.003]	0.039 [0.003]
Rental unit	0.04 [0.004]	0.038 [0.004]	0.037 [0.004]	0.033 [0.004]
Public Housing	0.063 [0.003]	0.059 [0.003]	0.058 [0.003]	0.053 [0.003]
Constant	1.743 [0.008]	1.647 [0.010]	2.102 [0.031]	2.015 [0.036]
Observations	134385	134374	134385	134374
R-squared	0.42	0.42	0.43	0.43
Cumulative bias 1993-95 (%)	3.99 [1.97]	5.05 [2.12]	4.34 [1.55]	5.35 [1.63]
Cumulative bias 1993-97 (%)	12.4 [1.6]	14.24 [1.71]	9 [1.34]	10.31 [1.43]
Cumulative bias 1993-99 (%)	20.7 [1.24]	23.47 [1.32]	12.67 [1.2]	14.4 [1.33]
Cumulative bias 1993-2001 (%)	29.14 [1.18]	32.81 [1.24]	17.21 [1.34]	19.59 [1.53]
Cumulative bias 1993-2003 (%)	34.47 [0.92]	38.31 [0.98]	22.74 [1.18]	25.05 [1.41]
Cumulative bias 1993-2005 (%)	34.29 [1.04]	38.8 [1.1]	21.58 [1.37]	24.56 [1.64]
Expenditure Weighted Bias				
Cumulative bias 1993-95 (%)			4.15 [1.76]	5.91 [1.94]
Cumulative bias 1993-97 (%)			7.19 [1.67]	8.86 [1.89]
Cumulative bias 1993-99 (%)			8.46 [1.65]	10.4 [1.93]
Cumulative bias 1993-2001 (%)			10.68 [1.93]	13.74 [2.27]
Cumulative bias 1993-2003 (%)			12.69 [1.78]	13.88 [2.2]
Cumulative bias 1993-2005 (%)			10.28 [2.06]	12.3 [2.51]

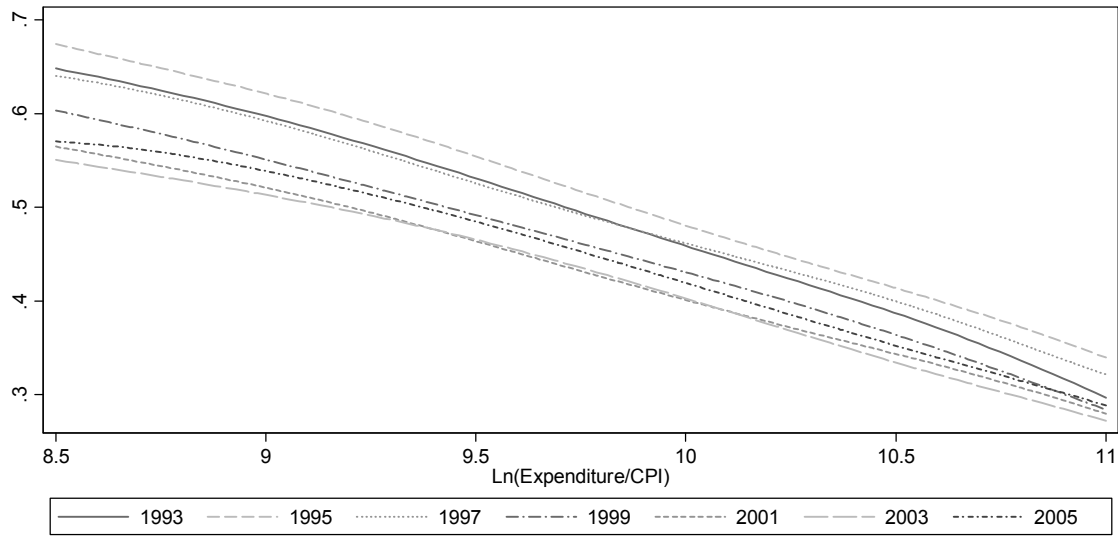
Notes: Robust standard errors for the regression coefficients and bootstrapped standard errors for bias estimates in brackets. Controls also include regional dummies. Income is used as an instrument to consumption in the IV regressions. Cumulative bias reported corresponds to $|E_{G,t}|$. The implied gross change in the true cost of living is $(1-|E_{G,t}|)$ times the gross change in the CPI, and the resulting gross true real income growth is $1/(1-|E_{G,t}|)$ times the gross real income growth obtained by deflating nominal income by the CPI.

Table 3. Implications of Estimated Bias for Consumption and Income Growth in 1993–2005.

Annual Growth in 1993-2005 in Percent Per Year							
	Mean	Median	Quintile (from poorer to richer)				
			1	2	3	4	5
Consumption							
Deflated by CPI	5.1	4.5	2.9	4.0	4.5	5.1	6.1
Correcting for Estimated Bias	5.7	5.8	6.4	6.0	5.8	5.7	5.5
Income							
Deflated by CPI	6.1	5.7	4.2	5.5	5.9	6.3	6.7
Correcting for Estimated Bias	6.9	6.9	7.7	7.5	7.2	6.9	6.2

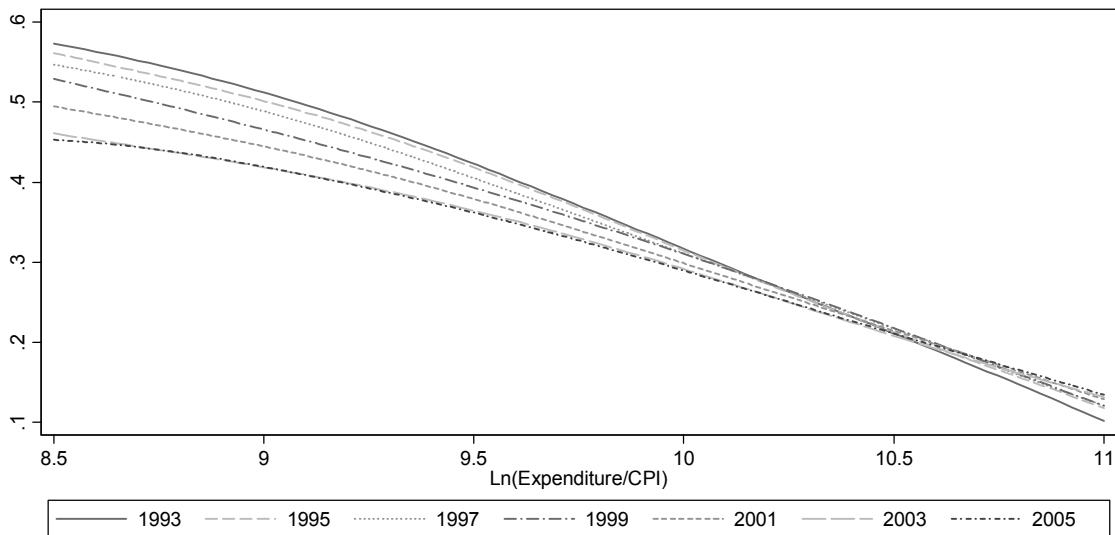
Notes: Corrected estimates based semi-parametric results. Quintiles based on (CPI-deflated) real expenditures.

Figure 1. Non-Parametric Estimates of Relationship between Food Shares and Household Expenditure



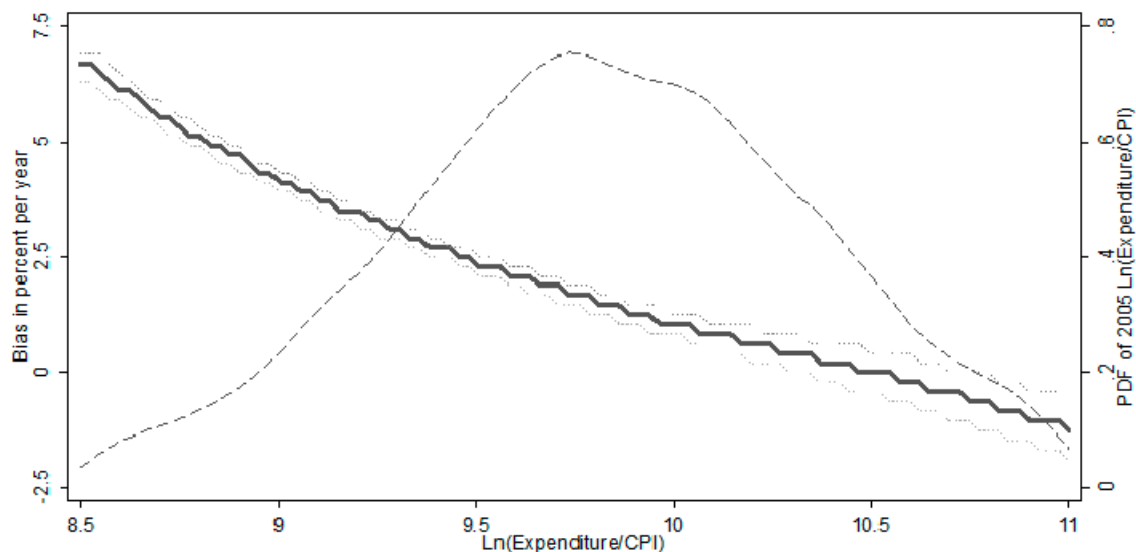
Note: Curves obtained from locally weighted linear regressions using quartic kernel weights.

Figure 2. Semi-Parametric Estimates of Relationship between Food Shares and Household Expenditure



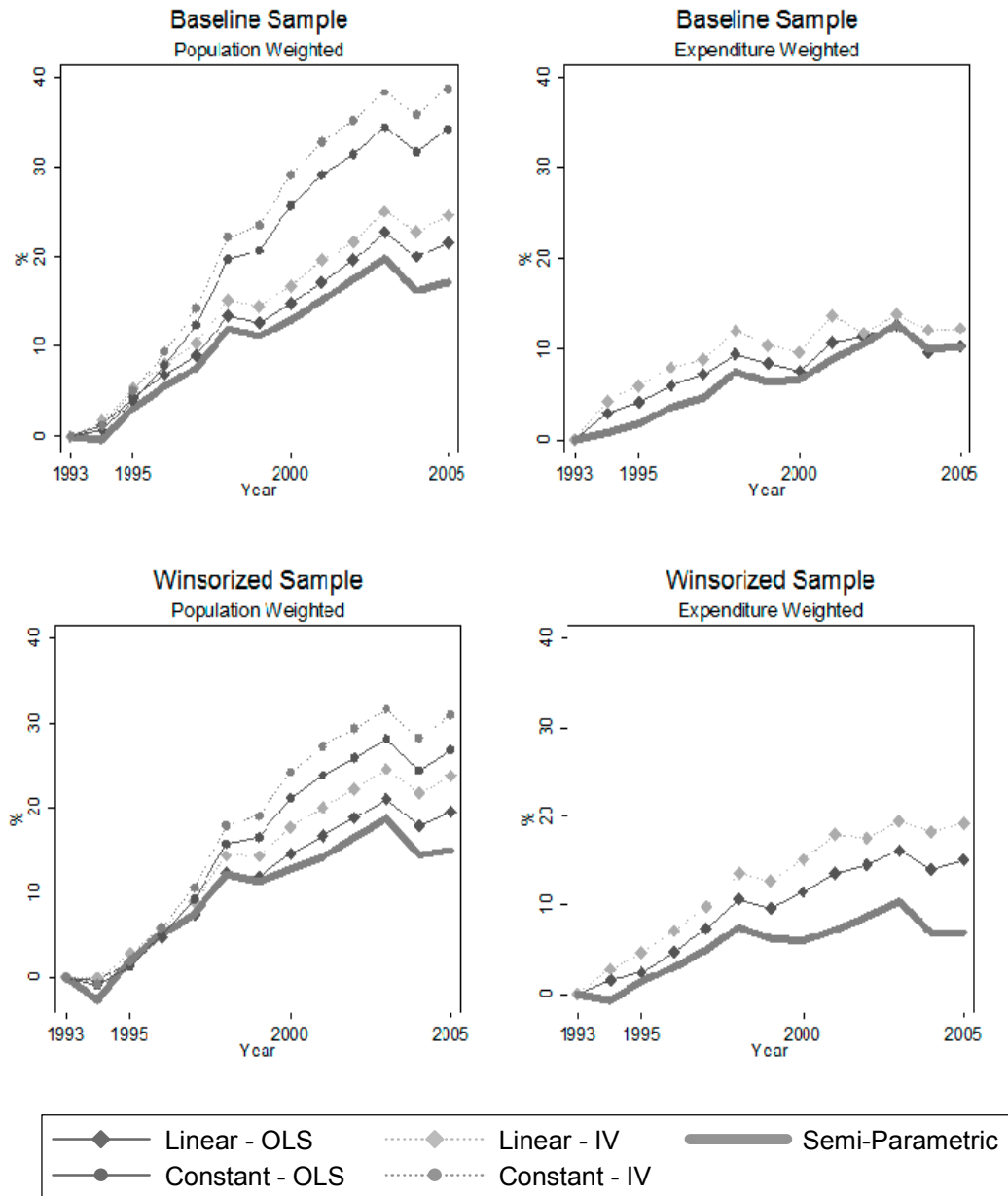
Note: Curves obtained from locally weighted linear regressions using quartic kernel weights. Parametric controls are the same as in the regressions in Table 2.

Figure 3. Estimated Bias in 1993–2005 as a Function of CPI-Measured Real Expenditure in 2005



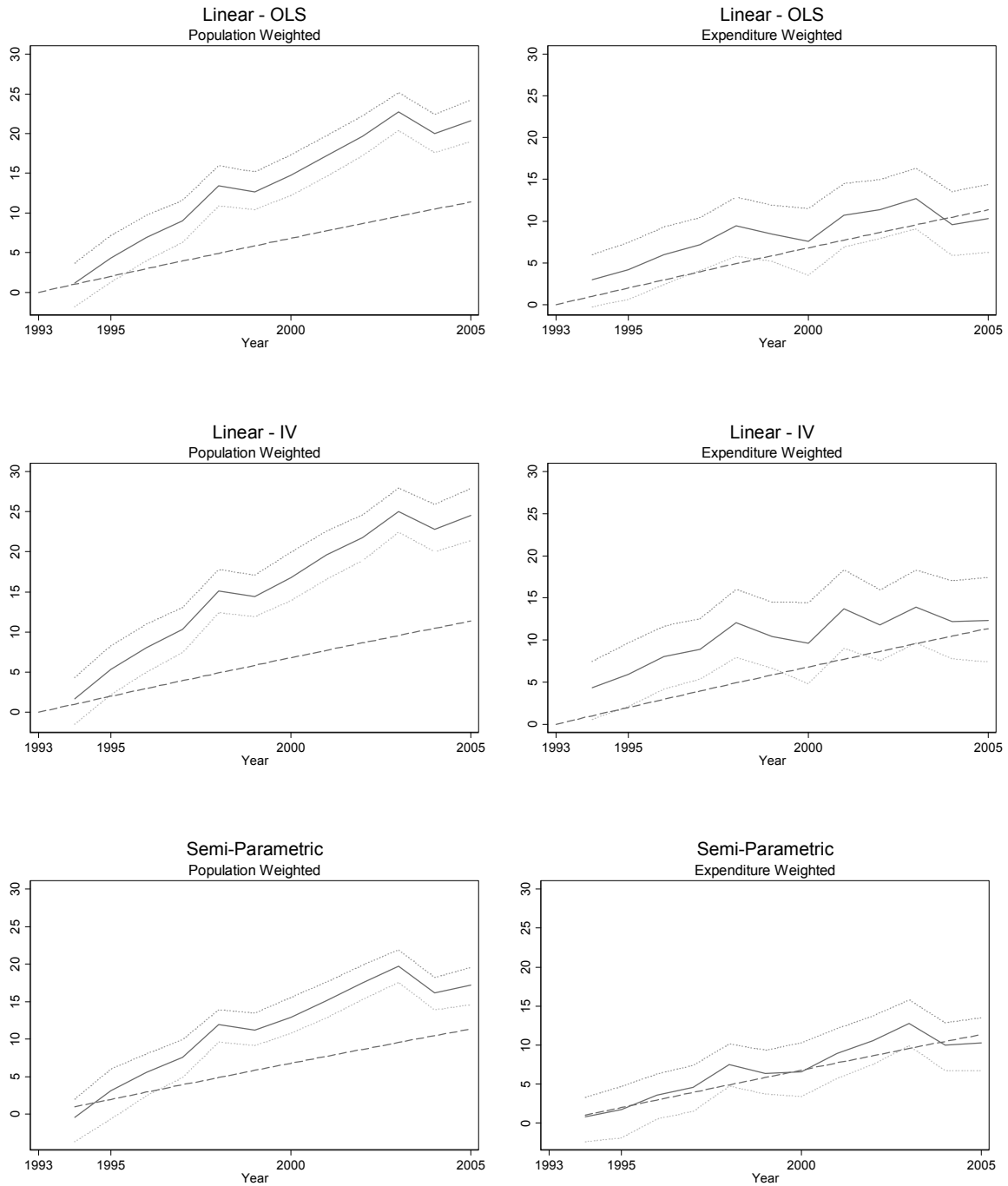
Note: Based on semi-parametric bias estimates from the full sample. Solid line corresponds to bias estimates and dotted lines to 95 percent confidence interval. Dashed line corresponds to the distribution of real expenditures. Bias estimates based on shift of semi-parametrically estimated Engel curves.

Figure 4. Estimated Cumulative Bias in China since 1984 Across Different Methods and Samples.



Notes: “Constant” Refers to specification where the bias is constant across all households in a given year; “Linear” to the one where it is linear on the log of (CPI-measured) real expenditure; and “Semi-Parametric” where it is a non-parametric function of (CPI-measured) real expenditure.

Figure 5. Estimated Cumulative Bias and Confidence Intervals under Different Methodologies.



Notes: Solid line corresponds to the point estimates and dotted lines to the 95 percent confidence interval. Dashed line indicates the cumulative bias under an annual bias of 1 percent per year, and is drawn only for illustration purposes.