

For Online Publication

Are Chinese Growth and Inflation Too Smooth?

Evidence from Engel Curves

Emi Nakamura

Jón Steinsson

Miao Liu

Columbia University

Columbia University

University of Chicago

September 22, 2015

A Data Appendix

Our primary source of data for the years 2005-2011 is the “China Urban Life and Price Yearbook,” published by the Chinese National Bureau of Statistics (NBS). For 1995-2004, we use the previous version of this yearbook, known as the “Chinese Price and Urban Household Survey Yearbook.”¹ These yearbooks compile data from the annual Urban Household Income and Expenditure Survey (UHIES) run by the NBS. From these yearbooks, we obtain the following variables at both the regional and income-group levels: commodity price indexes (CPI and RPI, at the regional level only), household expenditures by commodity, household size, the number of people earning income (including non-wage income such as retirement earnings) per household, and the number of people employed per household.

Our second source of data is the “China Statistical Yearbook,” also published by the NBS, from which we obtain the following variables at the regional level: sex ratio, elderly dependency ratio, child dependency ratio, and unemployment rate.² The household survey also has data on sex ratio. However, 2006 data on sex ratio are missing in the survey. We use instead general sex ratio from the main statistical yearbook as a proxy.

Our third source of data is the “China Sub-County Population Statistics,” published by China’s

¹Most of these data are available in Chinese in digital form from the China Knowledge Resource Integrated Database (CNKI) database.

²These data are available in digital form from the website of NBS at <http://www.stats.gov.cn/english/statisticaldata/yearlydata/>.

Ministry of Public Security. The Chinese are required to register with the local public security office whenever they move. We obtain from these data a measure of net migration ratio defined as the ratio of population moving into the urban areas of each province minus the ratio of population moving out these areas.³ From these data, we construct a variable that is zero up until 2002 and then equal to the cumulative net in-migration into each region’s urban area. We include this variable as a control for the potential role of migration on our analysis (we have also investigated this issue in several other ways, as we discuss in section 5).⁴

We have made a number of adjustments and modifications to the raw expenditure data to eliminate inconsistencies and errors, and to harmonize the expenditure categories across years:

1. The NBS revised the survey yearbook in 2002 and 2007, resulting in some inconsistencies in the definitions of commodities. First, the categories “Fruits, Melons and their Products” and a “Nuts and Kernels” are separate categories in the 1995-2001 data; however, “Nuts and Kernels” becomes a subcategory of “Fruits, Melons and their Products” in 2002. Thus, the definition of “Fruits, Melons and their Products” is inconsistent before versus after 2002. To address this issue, we combine the categories “Nuts and Kernels” and “Fruits, Melons and their Products” for the period before 2002, which harmonizes the definitions over time. This newly defined “Fruits, Melons and their Products” is the “Fruit” category we use in our pooling regressions.

Second, we combine the “Meat” and “Poultry” categories in the 2002-2011 data to create a larger category consistent with the “Meat and Poultry” category before 2002. This newly defined “Meat and Poultry” category is the “Meat” category we use in our pooling regressions.

2. We corrected a number of errors in the raw data. In 1995 and 1996, the sub-categories of “Food” do not add up to “Food.” Moreover, the “Other Foods” category has a value of zero for all regions in these two years only. We therefore believe that the “Other Foods” expenditures are misreported in these years. We define “Other Foods” in these two years as equal to the gap between “Food” and the sum of all its subcategories.

3. In 2006, the price observations from the household yearbook for Beijing, Tianjin, Shanghai,

³These data is available in digital form from China InfoBank at <http://www.bjinfobank.com/>.

⁴The NBS collects and publishes an alternative dataset on migration between regions. These data are based on the Chinese census in 2010, the 1% mini-Census in 2005, and on a smaller survey in non-Census years. The migration series from this dataset have large discontinuities in 2005 and 2010. This arises because annual NBS yearbook population data are based on hukou counts (only people with local hukou are included), not resident counts (all people with or without a local hukou), whereas the census counts are for residents. See Li and Gibson (2014).

and Chongqing are all zero for all goods. We are able to partially fill in these missing data by using the fact that, when available, observations of the urban and national CPI are the same for these regions, and the national CPI is also available from the main Statistical Yearbook. We therefore substitute national CPI data from the main Statistical Yearbook for these missing values whenever possible. Unfortunately, some missing values remain since the main Statistical Yearbook reports the CPI for a less detailed array of categories than the household survey. In particular, price observations for Starch, Bean, Condiment, Sugar, Milk, Cake are missing in 2006. This means that our pooling estimation for 2006 does not incorporate information from these subcategories of food.

B Additional Robustness

B.1 Allowing for a Break in 2002

One potential concern regarding our baseline results is that the shape of the Engel curve may have changed over time. Perhaps, for example, Engel curves have become flatter as China has developed, leading our model to be misspecified. Figure A.3 compares the results of our baseline analysis to results based on estimating separate Engel curves for the pre-2002 and post-2002 periods. The figure shows that this modification has almost no impact on our results.

B.2 Generalized Least Squares Estimation

In Figure A.4, we divide through equation (3) in the main text by the average expenditure share for each product category. The logic for doing this is to reduce potential heteroskedasticity in the baseline specification. In the baseline specification, the dependent variable is the level of the expenditure share. The average level of the expenditure share varies quite a bit across products—e.g., it is much higher for food/total than for fruit/food. This much larger level of the expenditure share for some products than others may be associated with a larger variance of the error term for those products. Dividing through by the average expenditure share will then put less weight on the product categories with large product shares.⁵ It is clear from Figure A.4 that once the baseline specification is scaled in this way, it yields results that are closer to the results for the specification that pools only the subcategories of food. However, since both food/total and the subcategories of

⁵This specification is a simple Generalized Least Squares regression under the assumption that the variance of the error term varies by product category and is proportional to the average level of the product category's expenditure share.

food yield similar results, this scaling affects our results in only a modest way and, in particular, does not change the results qualitatively.

B.3 Log-log versus Linear-log Engel curves

An additional question is whether our results are sensitive to the functional form of the Engel curve. To address this question, in Figure A.5, we change the dependent variable in equation (3) in the main text from the level of the expenditure share to the logarithm of the expenditure share. We refer to this as the “log-log” as opposed to the “linear-log” specification. From an economic standpoint, this means that we are assuming a different functional form for the underlying demand system. An advantage of this specification from a theoretical point of view is that the expenditure share is constrained to be positive. From an econometric standpoint, the effect of this change is similar to the rescaling in Figure A.4 in that it effectively puts more weight on the product categories with smaller average absolute levels of expenditure shares. Again the results are not much affected.

References

LI, C., AND J. GIBSON (2014): “Agglomeration Economies in China: Locations and Effects,” Working Paper.

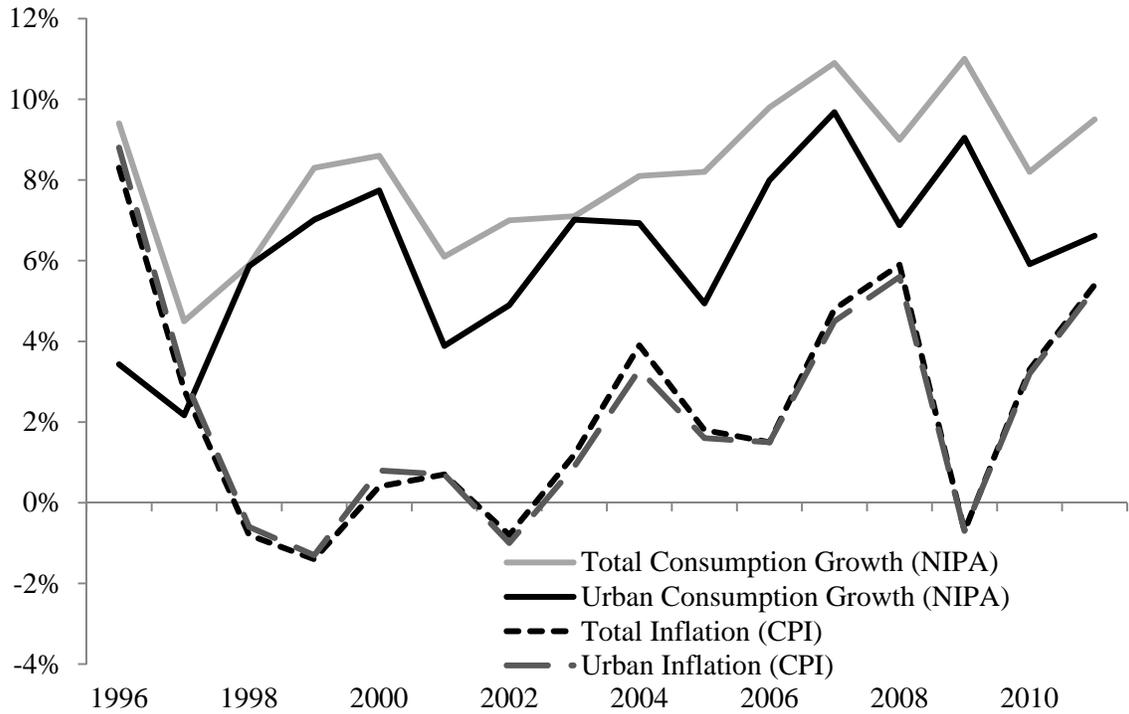


Figure A.1: Official Inflation and Real Consumption: Urban vs. Total

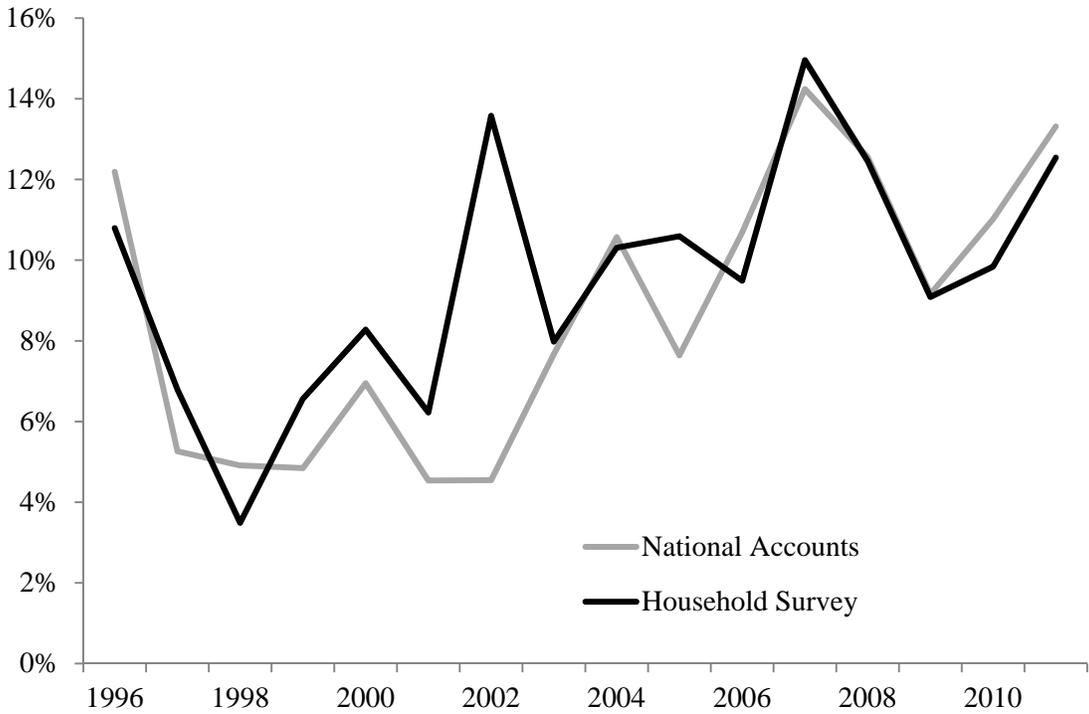


Figure A.2: Nominal Consumption: National Accounts vs. Household Survey

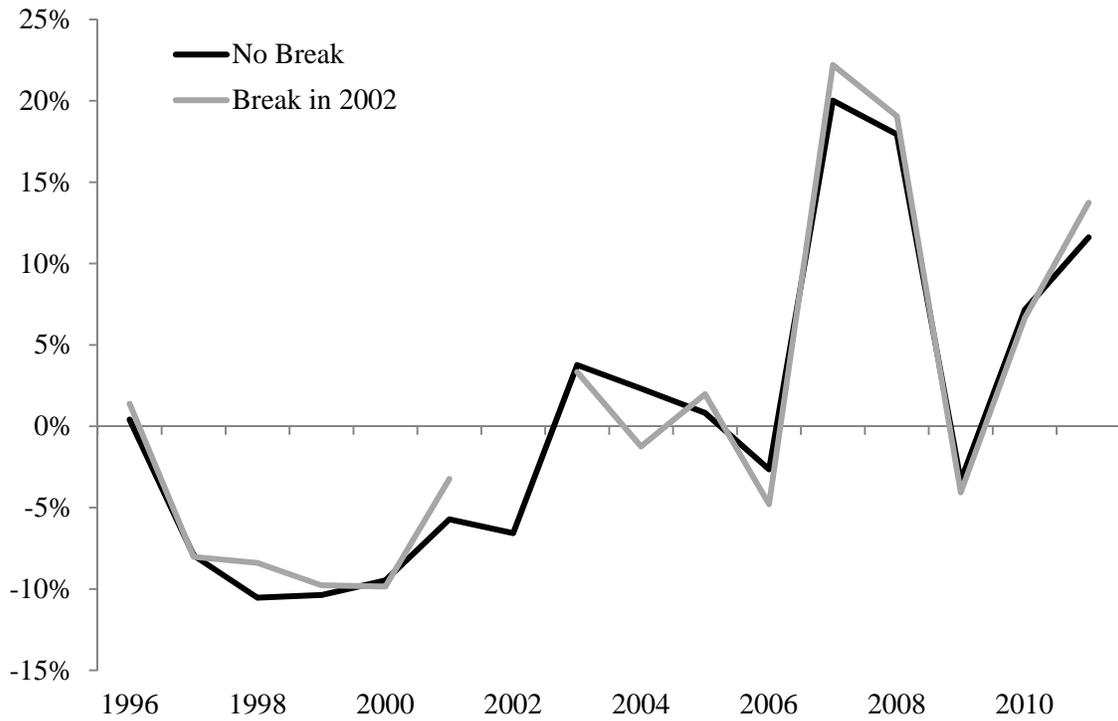


Figure A.3: Separate Engel Curves Pre and Post 2002

The figure plots results based on our baseline specification and an alternative version of this specification where we allow the Engel curve coefficients to be different before and after 2002. In both cases the results are based on a specification that pools the Engel curve for food expenditures as a fraction of total expenditures and Engel curves for the expenditures on 14 major subcategories of food as a fraction of food expenditures.

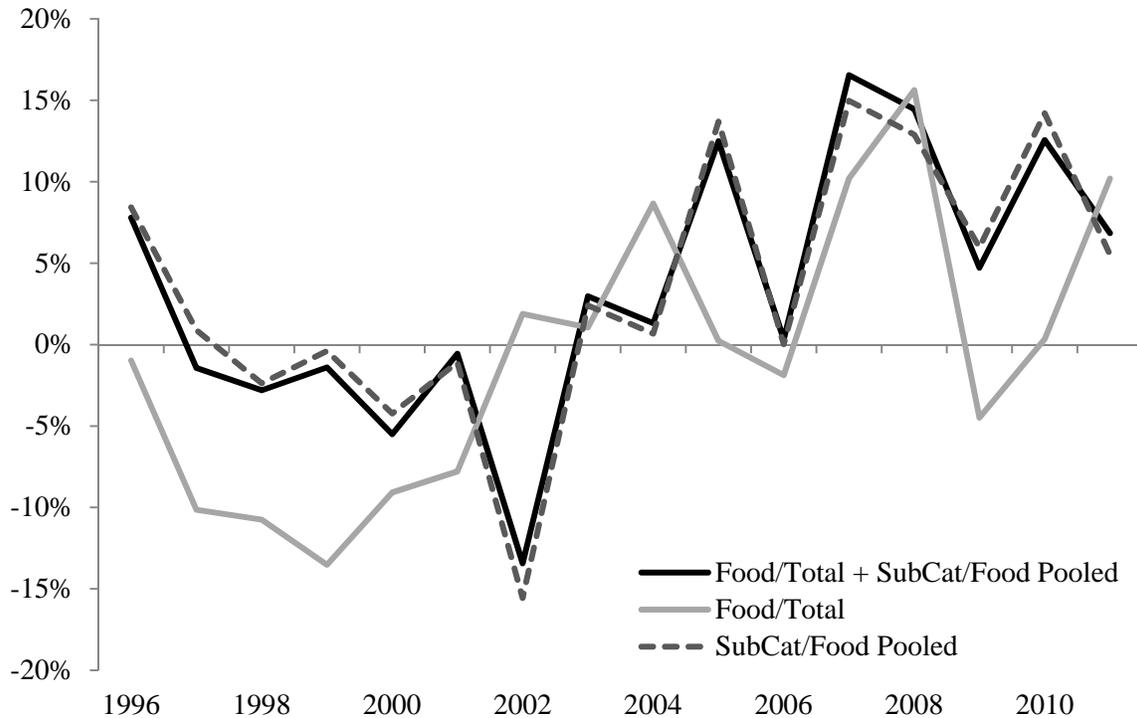


Figure A.4: Food versus Subcategories of Food Scaled

The figure plots results based on three specifications. The first is the baseline specification that pools the Engel curve for food expenditures as a fraction of total expenditures and Engel curves for the expenditures on 14 major subcategories of food as a fraction of food expenditures. The second specification is for the Engel curve for food expenditures only. The third specification pools the Engel curves for the expenditures on the 14 major subcategories of food as a fraction of food expenditures only. In the pooled specifications, the Engel curves are scaled by the average expenditure share for the product in question to reduce heteroskedasticity.

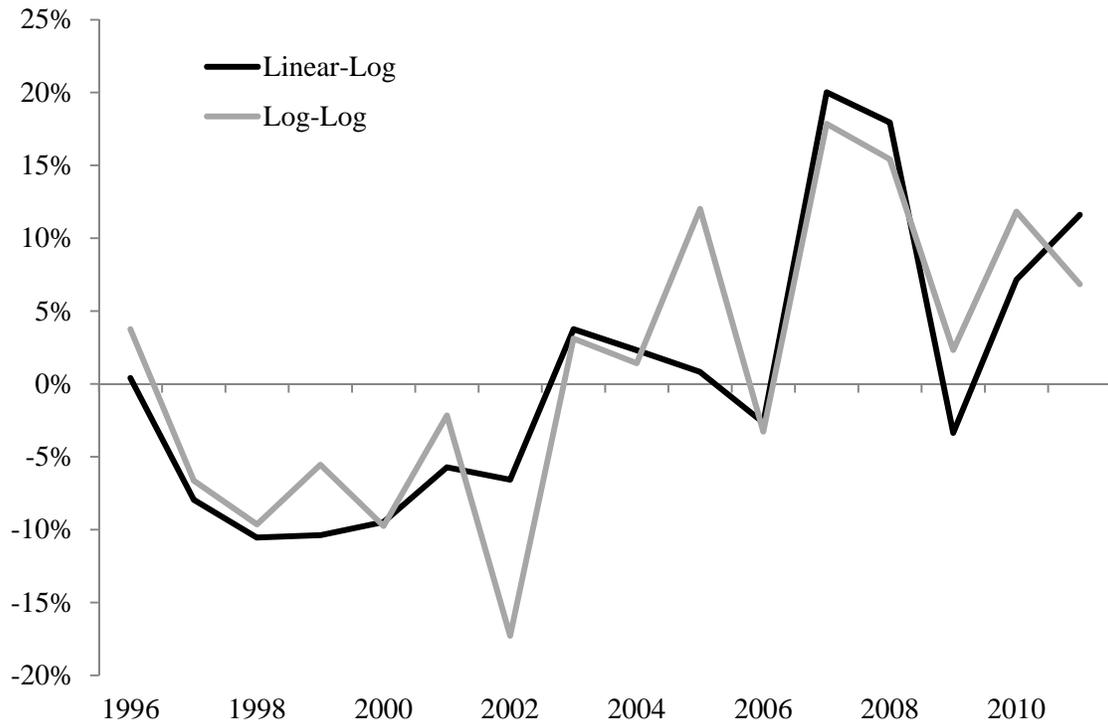


Figure A5: Linear-Log versus Log-Log Specification

The figure plots results based on our baseline specification and an alternative version of this specification where the dependent variable is the log of the expenditure share as opposed to the level of the expenditure share. In both cases the results are based on a specification that pools the Engel curve for food expenditures as a fraction of total expenditures and Engel curves for the expenditures on 14 major subcategories of food as a fraction of food expenditures.