Harvard-China Project on Energy, Economy and Environment

John A. Paulson School of Engineering and Applied Sciences,

Harvard University.

Working Paper

Urban Household Consumption in China

Jing Cao, Mun S. Ho, Wenhao Hu, Dale Jorgenson

December 31, 2017

HARVARD-CHINA PROJECT
on Energy, Economy and Environment

https://chinaproject.harvard.edu

Urban Household Consumption in China

Jing Cao, Mun S. Ho, Wenhao Hu, Dale Jorgenson

December 31, 2017

Abstract

Many studies have noted substantial differences between income and consumption measures of well-being and inequality in many countries. A national welfare measure based on consumption requires an econometric approach that can account for differences among household types and require data on regional prices. We estimate a flexible consumption function using surveys of consumption in urban households in China and using detailed regional price data. This function distinguishes households by demographic characteristics and provide price and income elasticities. We use the estimated consumption function to project trends in aggregate food, nondurable goods, services and housing consumption by incorporating projections of population changes, income and prices. Such a function allows a future calculation of household equivalent scales to construct indices of national welfare that takes into consideration the size and structure of households.

Jing Cao is Professor of Economics in Tsinghua University, School of Economics and Management; Wenhao Hu is a Ph.D. candidate at Tsinghua University; Ho is a Visiting Scholar at the Harvard China Project on Energy, Economy and Environment based in the School of Engineering and Applied Sciences; Jorgenson is Professor of Economics, Harvard University. Cao and Ho are supported by the Harvard Global Institute, Hu is supported by China Scholarship Council. Ho is also a Visiting Scholar at Resources for the Future, Washington. This paper is also supported by National Science Foundation of China (project codes: 71422013 and 71173130), Hang Lung Center for Real Estate, Tsinghua University Research Center for Green Economy and Sustainable Development, Tsinghua University Initiative Scientific Research Program.

1. Introduction

In the face of a sharp slowdown in economic growth and rising inequality, the government of China pressed ahead with economic reforms, including a shift away from a reliance on exports and investment towards consumption, and away from manufacturing towards services. Understanding consumption at a micro level is a key to designing economic rebalancing and consumption policies in a way that recognizes the heterogeneity of households. Furthermore, since distinct trends in consumption inequality versus income inequality are observed in many countries, understanding China's consumption inequality is also important for designing poverty alleviation policies.

The economics literature has long discussed the substantial differences between income and consumption expenditures, and this also holds for China. For example, in 2012 the disposable income per capita of the highest 10% households was about 7.8 times that of the lowest 10% households, while the expenditure per capita was only about 5.9 times[1]. This implies a typical pattern of higher saving rates for higher-income households.

There are numerous studies contrasting consumption and income inequality in the U.S., for example, Meyer and Sullivan (2013) examined the changes between 2000 and 2011 and found that after the Great Recession, although the lower-income households suffered a greater fall in incomes, their consumption fell by less. The lowest decile had a growth rate of income that is 10% lower than the highest decile, but their growth rate of consumption is only 2% lower than that of the highest decile.

Consumption is a better measure of material well-being than income; it reflects long-run resources in the life-cycle model and permanent income model. A particular year's income may not reflect the level of assets, access to credit or life-time income[2]. The quality of household consumption data in many countries has been lagging that of

[1] China Statistical Yearbook 2013, Tables 11-6

[2] Jorgenson (1998) discusses the sharp difference between income and consumption measures of U.S. welfare.

income data, so there have been fewer studies of consumption inequality, but a substantial number of studies for the U.S. have emerged (see the survey by Attanasio and Pistaferri 2016). Unfortunately, studies in developing countries are still lagging due to the lack of adequate data.

The difference between consumption and income highlights how poverty alleviating policies should be evaluated. The government sets the poverty line according to disposable income instead of expenditure measures. However, Meng, Gregory and Wang (2005) showed that such a poverty line might significantly underestimate the effective poverty rate with the rising saving rates. During China's high income growth era, the national savings rate rose from 17.4% in 1995 to 28.8 in 2008.[3]

A consumption model that recognizes the different characteristics of households, including incomes and demographics, would also help in making projections of national consumption. As we show later, the household data in China show typical patterns of falling budget shares for food and other necessities as income rises. The rapid growth of incomes in China is accompanied by a rapid change in consumption patterns. The changes in household structures – smaller families, more elderly members, better education – have also contributed to changes in aggregate consumption demand.

In this paper, we estimate a translog model of consumption that allows for heterogeneity among household types, and allow a flexible degree of substitution and income elasticities. Using repeated cross-sections of household data we are able to identify both price and income elasticities for four major categories of consumption. Our formulation also allows for a direct aggregation over households to deliver an aggregate demand function for commodities. Instead of having distinct functions for different households, an aggregate demand that is a function only of prices, aggregate income and distribution measures is very useful for empirical general equilibrium models. For

[3] Meng (2003), Chamon and Prasad (2007), and Wei and Zhang (2011) discuss various reasons for the rising savings rate including income uncertainty, housing prices, and higher private responsibility for health and education expenses.

example, a similar model of U.S. consumption in Jorgenson and Slesnick (2008) has been applied in the computable general equilibrium model of the U.S. economy in Jorgenson et al. (2013) to analyze tax policies.

Much of the research so far about household demand estimation in China has mainly focused on narrowly defined goods. Ortega, Wang and Eales (2009) estimate meat demand, while Hovhannisyan and Gould (2014) use the AIDS model to estimate the food demand in urban China. Fan, Wailes and Cramer (1995) who estimated a 2-stage LES-AIDS model using rural provincial data from 1982 to 1990. All of these papers use time series of macro consumption data, not micro-level household data. Chen and Xing (2011) estimated the demand for cigarettes in urban China using household data from 1999 to 2001. Cao, Ho and Liang (2016) estimates the energy demand in urban China using a two-stage AIDS model with household data from 2002 to 2009. So far, very few studies include a complete set of consumption items using micro-level data.

Our paper is the first effort at estimating a comprehensive demand function using household survey data from urban Chinese households covering the period 1992 to 2009. We distinguish households by size, location, education and employment of household head, and presence of children or elderly members. Our model allocates total expenditures into food, consumer goods, services and housing. We also show how aggregate consumption demand may be projected taking into account future price and income effects, in addition to households' demographic changes.

We combine expenditure data for 180,000 households from the Urban Household Income and Expenditure Survey (UHIES) and price information from the Consumer Price Index (CPI) from 1992 to 2009. China has an unusually large share of owner-occupied housing, so we make an extra effort to impute rental equivalents using estimates of rent-to-price ratios and regional housing prices. We also use cross-sectional prices of detailed commodities in each region to identify regional differences in the benchmark year 2009.

We find that food is less price elastic than the other three kinds of goods. Consumer goods and services are income elastic, while food and housing are income inelastic. We then project aggregate consumption shares based on projections of prices and incomes from the China CGE model in Cao et al. (2017). The income effects dominate gradually with time. Therefore, the food share and residence share keep falling, while shares for consumer goods and service go up.

The paper is constructed as follows. We introduce the translog model of consumer behavior in Section 2. In Section 3, we discuss the data including rental equivalent imputations and cross-sectional prices that vary over regions and time. Section 4 presents aggregate consumer behavior, and we show how the use of a longer time series of prices affect the estimates in short repeated cross-sections. In Section 5, we project consumption patterns using a China CGE model. We conclude and summarize the results in Section 6.

2. Methodology

We follow the econometric model of consumer behavior described in Jorgenson and Slesnick (1987) and briefly summarize the main equations here. We assume a translog indirect utility function for household k:

$$\ln V_k = \ln(\frac{p}{M_k})' \cdot \alpha_p + \frac{1}{2}\ln\left(\frac{p}{M_k}\right)' \cdot B \cdot \ln\left(\frac{p}{M_k}\right) + \ln\left(\frac{p}{M_k}\right)' \cdot B_A \cdot A_k \qquad (2.1)$$

in which:

$\mathrm{p} = (p_1, p_2, \dots, p_N) -$ vector of prices of consumption bundles,

$x_k = (x_{1k}, x_{2k}, \dots, x_{Nk}) -$ vector of quantities consumed by household k,

$M_k = \sum_{n=1}^{N} p_n \cdot x_{nk} -$ total expenditures of household k,

$\omega_{nk} = p_n \cdot x_{nk}/M_k -$ expenditure share of the n-th commodity,

$\omega_k = (\omega_{1k}, \omega_{2k}, \dots, \omega_{Nk}) -$ vector of expenditure shares,

$A_k -$ vector of (0,1) attribute indicators.

In this form, the preference differences among households are introduced through the attribute vector A_k. The matrices α, B and B_A are constant parameters that are the same for all the households. We consider a model with N=4 consumption bundles – food, consumer goods, services and housing.

Lau (1982) discusses the conditions required for exact aggregation of the translog function, that is, the restrictions needed so that an aggregate demand function is obtained by explicit aggregation over households. These conditions are:

$$\mathrm{i}' \cdot \mathrm{B} \cdot i = 0, \ \mathrm{i}' \cdot B_A = 0 \qquad (2.2)$$

where i is a vector of 1's. In addition, homogeneity of the demand function allows us to choose a normalization:

$$\mathrm{i}' \cdot \alpha = -1$$

The vector of expenditure shares derived by Roy's identity is:

$$\omega_k = \frac{1}{D(p)}[\alpha + B \ln p - B_M \ln M_k + B_A A_k] \qquad (2.3)$$

where the denominator takes the following form under the aggregation conditions:

$$D(p) = -1 + B_M{}' \cdot \ln p \qquad (2.4)$$

$$B_M = B\iota$$

Integrability of the demand system also requires that the matrix of price substitution effects be symmetric and nonpositive definite:

$$B' = B$$

Two methods are used to estimate this demand system. The first method combines one cross section of household data with time series of aggregate prices and shares. The second method follows Jorgenson and Slesnick (2008) in using repeated cross sections; pooling all years (1992-2009) with household observations, where prices vary across region and time. Since we have a longer time series of national prices (1981-2011) we estimate the demand function using both methods and compare the results. In both methods, the parameters α_p, B_M, B_A are identified by the cross-section household data.

The price coefficients are identified from aggregate time series of prices in the first method, and regional prices in the second method.

For the first method, we derive the aggregate expenditure share vector for year *t* by summing over all households:

$$\omega_t = \frac{1}{D(p_t)} \cdot \left(\alpha_p + B \cdot \ln p_t - B_M \frac{\sum M_{kt} \cdot \ln M_{kt}}{\sum M_{kt}} + B_A \cdot \frac{\sum M_{kt} \cdot A_k}{\sum M_{kt}}\right) + \mu_t \qquad (2.5)$$

where we have added a vector of unobservable random disturbances, μ_t ..

The first step is to consider a random sample of observations on individual expenditures on the four bundles in a given year with cross-sectional data. If there is no cross-sectional price data, we have to assume households face the same price and the model for household expenditures (2.3) takes the form:

$$-\omega_k = \alpha_p - B_M \cdot \ln M_k + B_A \cdot A_k + \mu_k, \ (\mathrm{k} = 1,2,\dots,\mathrm{K}) \qquad (2.6)$$

In this form, vector μ_k has a singular distribution since the shares add up to 1. We can drop one equation and write the cross-section model (2.6) in the form:

$$y_1 = X\beta_1 + \varepsilon_1 \qquad y_i = (y_{i1}, \dots y_{ik}, \dots, y_{iK})'$$

$$\dots$$

$$y_{N-1} = X\beta_{N-1} + \varepsilon_{N-1} \qquad (2.7)$$

where y_i (i=1,2,…,N-1) is the vector of observations on individual expenditure shares of the ith commodity for all households, X is the matrix of observations on independent variables, and ε_i (i=1,2,…,N-1) is a vector of unobservable random disturbances.

Stacking equation (2.7), we obtain:

$$\mathrm{y} = [\mathrm{I} \otimes \mathrm{X}]\beta + \varepsilon \qquad (2.8)$$

where $\otimes$ is the Kronecker product and:

$$\mathrm{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix}, \ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{N-1} \end{bmatrix}, \ \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{N-1} \end{bmatrix} \qquad (2.9)$$

The matrix X is of full rank and the random vector ε is distributed normally with mean 0 and covariance matrix $\Sigma_\varepsilon \otimes \mathrm{I}$, where Σ_ε is obtained from the covariance matrix Ω_ε by striking the row and column corresponding to the omitted equation.

The OLS estimator of the vector of parameters β from the cross-section model is:

$\hat{\beta} = [\mathrm{I}\otimes(\mathrm{X}'\mathrm{X})^{-1}X']y$ (2.10)

The second step is to consider the time series aggregate expenditure patterns based on model (2.5). Again, we drop one share equation and express the N prices as relative prices, p_i / p_N, giving:

$$\tilde{w}_t = \frac{1}{D(p)}[\tilde{\alpha} + \tilde{B}\ln\tilde{p}_t - B_M\xi_t^M + B_A\xi_t^L] + v_t$$

where the tilde above the variables represent the N-1 dimensioned counterpart of those in (2.5), and $\xi_t^M = \sum_k M_{kt}\ln M_{kt} / M_t$, $\xi_t^L = \sum_k M_{kt}A_k / M_t$. Stacking this nonlinear function and relabeling the variables and parameters we get:

$\varphi = f(\beta, B) + \nu = f(\delta) + \nu$ (2.11)

where

$$\varphi = \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{N-1} \end{bmatrix}, f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{N-1} \end{bmatrix}, \delta = \begin{bmatrix} \beta \\ B \end{bmatrix}, \nu = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_{N-1} \end{bmatrix} \quad (2.12)$$

β represents α_p, B_M, B_A. Under our assumptions, the random vector ν is distributed normally with mean 0 and covariance matrix $\Sigma_\nu\otimes\mathrm{I}$, where Σ_ν is obtained from the covariance matrix Ω in (2.5) by striking the row and column corresponding to the omitted equation.

The first two steps above give us an initial estimate of the parameters. The cross-section data were used to identify the constants α_p, the coefficients of total expenditure B_M, and the demographic coefficients B_A. The price coefficients B are identified from aggregate time series data.

In the last step, we estimate a pooled model using non-linear 3-stage least squares (NL3SLS). We minimize the following with respect to the vector of unknown parameters δ:

$$\mathrm{SSR}(\delta) = [y - Y\delta]'\left[\hat{\Sigma}_{\varepsilon}^{-1} \otimes I\right][y - Y\delta]$$

$$+[v - f(\delta)]'[\hat{\Sigma}^{-1} \otimes Z(Z'Z)^{-1}Z'][v - f(\delta)] \qquad (2.13)$$

where,

$$Y = \begin{bmatrix} I \otimes X & 0 \\ 0 & 0 \end{bmatrix}$$

is a matrix of observations on the independent variables in (2.8) and Z is a matrix of instrumental variables for the aggregate model (2.12)[4].

We choose estimates that minimize the objective function (2.13), subject to constraints implied by integrability and concavity discussed above. The integrability constraints are described in detail by Jorgenson, Lau and Stoker (1982), while the concavity constraints are discussed in detail by Holt and Goodwin (2009) and by Moschini (1999). The NL3SLS estimator obtained by minimizing the function (2.13), subject to the constraints, is a consistent estimator of the vector of unknown parameters δ.

The second method of estimating the demand system assumes that the disturbances in the demand system (2.3) are additive so that the system of estimating equations is:

$$\omega_{kt} = \frac{1}{D(p_{kt})} \cdot (\alpha + B \cdot \ln p_{rt} - B_M \cdot \ln M_{kt} + B_A \cdot A_{kt}) + \varepsilon_{kt} \qquad (2.14)$$

where the error vector ε_{kt} is assumed to have mean zero with variance-covariance matrix Σ. We drop one equation since the shares add to one, and express three prices relative to the fourth. We construct regional prices, p_{rt}, for each year 1992-2009, and estimate (2.14) as repeated cross-sections following Jorgenson and Slesnick (2008). The regional prices are explained in Section 3.

Holt and Goodwin (2009) also discuss the elasticities of translog demand systems. In our work, we follow their formulas in estimating the elasticities. Uncompensated price elasticities in the translog demand systems are given by:

$$\eta_{ij} = -\delta_{ij} + \frac{\beta_{ij}/\omega_i - \beta_{Mi}}{-1 + \sum_k \beta_{Mk} \cdot \ln(p_k/M)} \qquad (2.15)$$

[4] This is equation 3.20 in Jorgenson and Slesnick (1987).

Expenditure elasticities are given by:

$$\eta_{iy} = 1 - \frac{\sum_j \beta_{ij}/\omega_i}{-1+\sum_k \beta_{Mk} \cdot \ln(p_k/M)} \tag{2.16}$$

where δ_{ij} is the Kronecker indicator. The compensated (Hicksian) price elasticities are:

$$\eta_{ij}^{C} = \eta_{ij} + \omega_j \eta_{iy} \tag{2.17}$$

3. Data Sources and price construction

3.1 UHIES data

In China, the only comprehensive source of information on household income, consumption expenditures on disaggregated items, demographics and housing is the Urban Household Income and Expenditure Survey (UHIES) conducted by the National Bureau of Statistics (NBS). The UHIES is conducted every year, using a stratified design and probabilistic sampling. One third of the sample households are replaced each year. The national sample size since 2001 is more than 40,000 observations. The commodity categories we use are given in Appendix A.1. We group the expenditures into four bundles:

1. Food (FD) – expenditures on purchased and in-kind food (including dining out)
2. Consumer goods (CG) – expenditures on clothing, household equipment, medical goods, educational goods, transportation equipment, communications equipment, recreational goods, and other goods.
3. Services (SV) – expenditure on medical care, educational services, transportation services, communication services, recreation services, and other services.
4. Housing (HS) – expenditures on rental equivalents of housing, household utilities include water, electricity, and household fuels.

The data used in our study is a subsample of the UHIES, covering 9 provinces from

1992 through 2009[5]. The 9 provinces were selected to represent all regions of China: Beijing, Liaoning, Zhejiang, Anhui, Hubei, Guangdong, Sichuan, Shaanxi and Gansu. Both less developed provinces (Anhui and Gansu) and richer ones (Guangdong and Zhejiang) are included. After 2001, our sample covers about 90% of the cities in the 9 provinces, while only 60% are covered before that. The sample size is between 5,000 and 6,000 households per year before 2001, and 15,000-17,000 after that. We compared the average expenditure shares in our sample to the national ones given in Statistical Yearbook (2009), and find that they are within 0.8 percentage points of each other (Appendix A.2).[6]

Within each province, the large cities are significantly different from the smaller ones in many respects, in particular the level of consumer prices. We use the differences in regional prices to help identify price elasticities and find a distinct difference between the large cities and the small ones.[7] With this division between large versus small cities, our 9 provinces result in 17 distinct regions (Beijing is just one large city and thus not divided).

In the UHIES data, there are some obvious errors with extreme values. We replace the extreme values by the group average, for example, if we detect for some household the vegetable expenditure is as high as around 90,000 yuan per year, we than use the average expenditures on vegetables by that demographic group. We identified 27 such errors in our 9-province sample[8].

The household weights are based on NBS sampling weight for each city, and we rescale it so that can be comparable for the whole sample, not only within cities. The procedure is described in Appendix A.3. In-kind income[9] data are only available after

[5] This subsample comes from the China Data Center, Tsinghua University (CDC).

[6] Other surveys such as the China Household Income Panel survey (CHIP) and the China Family Panel Survey (CFPS) focus on income and have little expenditure data. While the panel structure is useful, their lack of consumption data renders them unsuitable for our purposes.

[7] 36 cities in the whole country are defined as big by the NBS. The big cities in the 9-province sample are: Beijing, Shenyang, Dalian, Hangzhou, Ningbo, Hefei, Wuhan, Guangzhou, Shenzhen, Chengdu, Xian and Lanzhou.

[8] Cao, Ho and Liang (2016) eliminated outliers that exceeded 4 times the 99th percentile of expenditures in a given category in the same dataset. After applying our elimination process, none of the observations exceeds that limit.

[9] In China, many employees get noncash subsidies from the state owned enterprises. The in-kind income usually

2002. Expenses on owner-occupied housing include new purchases and renovation that are classified as investment, not consumption, in the National Accounts. We do not use these actual expenditures but instead impute rentals equivalents as housing service flow as described in the next section.

The annual service flow from durable goods should ideally be calculated from data on stocks of different types of consumer durables and their depreciation rates as in Jorgenson and Slesnick (1987). Unfortunately, neither the UHIES nor other sources of data allow us to estimate the household stocks well. The UNIES does indicate that most households own durables such as refrigerators and vehicles, and gives the expenditures on purchase of the durables if the households purchased them in the survey year. We approximate the service flow by noting that in the steady state households replace each type of durable when it has completely depreciated. We thus divide the purchases of durables by households purchased in the survey year to all households in the similar decile group. We allocate the households into deciles according to the expenditures on non-durable goods per capita within each region, in each year. We sum over all households i that purchased the consumer durable CD_i in region r, in each decile I and then divide the total by the sum of all household weights (fw) in group rI :

$$S_{rI,t} = \sum_{i \in rI} (fw_{i,t} \cdot CD_{i,t}) / \sum_{i \in rI} fw_{i,t} \qquad I=1,2,\ldots 10 \qquad (3.1)$$

We interpret $S_{rI,t}$ as the annual service flow from durables in each household decile I of region r at time t.

3.2 Measuring Owner Occupied Housing

In China today, most urban households own their homes, our sample shows that only less than 10% live in rented units. This makes it a special challenge to construct a proper measure of annual consumption expenditures.

There are three ways to estimate rental equivalents of owner-occupied housing

---

consists of food, daily necessities and sometimes bus tickets and tourism. The shares of in-kind expenditures are now much lower, the highest one is about 4% for food in 2009 compared to 8% in 2002.

(OOH). The first is to use the households' estimates of the rental equivalents in response to the question asking how much they would charge to lease their homes out. The second way is to find a similar house in the rental market and use its rent. However, neither method can be used here. The UHIES does ask respondents to give a rental equivalent, however, these are implausibly low in practice. Rented housing mostly involves public housing or employer-owned housing where rents are not set in a regular market. The rents paid do not give a correct sense of the few units that are proper market units. The characteristics of the apartments are not given in detail and make it very difficult to find a similar house in the same location. The third way is to use capital stocks of housing to estimate the owner occupied rent as in Li, Luo and Sicular (2011).

The UHIES survey does ask for the current price of the homes beginning in 2002 (i.e. the stock price) [10]. We combine this with studies of rental-price ratios from the Hung Lung Center for Real Estate at Tsinghua University, to obtain a rental estimate. We begin with the values of the OOH homes (*Hvalue*) in the 2009 UHIES. Our goal is to estimate the annual rental equivalent:

$$Rent_t^{r_i} = \rho_t^r \cdot Hvalue_t^{r_i} \tag{3.2}$$

where r_i is the i^{th} household in region *r*, and ρ is the rent-price ratio for region *r* in year *t*.

The Hung Lung Center for Real Estate at Tsinghua University estimates the rental-price ratio in about 110 cities in China from 2009 to 2013, using households' estimates of the rental equivalents and the current price of their homes. In 2009, 29 cities out of 87 in our 9-province sample is covered by this Tsinghua study. All 17 of our regions are represented by these 29 cities, including both large and small cities. For those cities not included in Hung Lung dataset, we use the regional average rental-price ratio.

---

[10] Housing prices per square meter vary hugely over the 17 regions in our sample, or even within the same city. In our sample, about 7% of households have zero or missing values for house price, which we replaced with the regional average. There are no extreme values found, all were less than twice the 99^{th} percentile value within each region. For the regional average, we use the middle 90% of the sample, excluding 5% in each tail. For those households with no apartment size data (less than 1%), we use the regional average size.

Our next step is to compute regional rental equivalents (per square meter) for 1992 to 2008 in the 17 regions. We first assume that the rental equivalents (per square meter) of the owner occupied units change at the same rate as the rent index in the CPI:

$$RP^r_{t+1}/RP^r_t = \widehat{CPI}^{rent}_{r,t+1} \quad t = 1992, \ldots 2008 \tag{3.3}$$

where RP^r_t is the rental equivalent (per square meter) of region *r* at time *t*.

The survey gives us house values after 2001, and so we can calibrate the inflation of rent-price ratios ρ^r_t for this period using the inflation rate of rental equivalents:

$$\frac{\sum_{k=1}^{rK}(\rho^r_{t+1} \cdot Hvalue^{rk}_{t+1})/\sum_{k=1}^{rK}(Area^{rk}_{t+1})}{\sum_{k=1}^{rK}(\rho^r_t \cdot Hvalue^{rk}_t)/\sum_{k=1}^{rK}(Area^{rk}_t)} = \frac{RP^r_{t+1}}{RP^r_t} \qquad t=2002,..,2008 \tag{3.4}$$

From these time series of regional ρ^r_t's, we get the annual rental equivalents for each household using eq. (3.2).

Prior to 2002, we have to apply the regional rental equivalent, RP^r_t, to estimate the rents for all households. As we noted, many of the rental units are not at market rents, and so we replace the reported rents with the rental equivalents per m^2, multiplied by the size of the rental unit.

Total expenditures

Our final estimate of total expenditures is thus the sum of spending on nondurables and services plus the service flows from consumer durables and owner occupied housing.

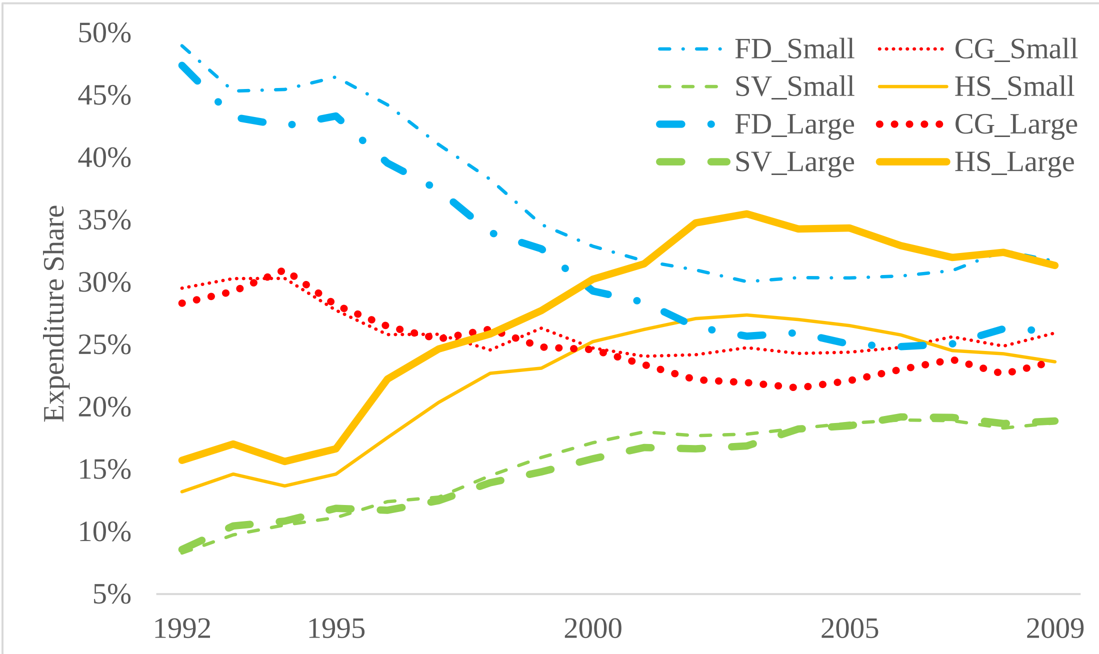


Figure.1 Expenditure Shares (Small Cities Vs. Large Cities).

FD: Food, CG: Consumer Goods, SV: Services, HS: Housing

In Figure 1, we plot the four expenditure shares separately for large and small cities. During the 1992-2009 period, per-capita GDP rose at 9.5% per year with the Services share rising from 9% to 19% in the mid-2000s, while the Food share fell from 49% to 30% in the small cities. The housing shares rose rapidly in the 1990s, peaking in 2002 along with the privatization of public housing. After that, the estimated housing share fell slowly, driven by the relatively slower rise in imputed rentals. The shares of consumer goods also fell until the mid-2000s, and then rose slightly. The trends are similar overall between the large and small cities; however, there are some differences, such as a smaller food share and larger housing share in the large cities. As we show later, this is due to the higher incomes and higher housing prices in the large cities.

3.3 Measuring Price Levels in China

The UHIES records the RMB expenditures on hundreds of items, but quantities are only given for food, clothing, durable goods and energy. No information on prices paid are provided, so we need to link with price data from alternative sources. Unfortunately, there is little comprehensive data on price levels in China. From 1990 to 1993, the NBS published price levels on specific products (defined uniformly across the country) for big cities in all provinces. At that time, China was still only emerging from the planned economy, and many industrial consumer goods were still subject to a central distribution system, making this endeavor feasible. This data is the only official source available of price level data in China.

Brandt and Holz (2006) estimate the provincial price levels based on this 1990 NBS data, and these are used by other researchers. They are, however, not suitable for our purposes. First, their regional price index is based on a fixed basket of goods for each province (separately for urban and rural areas) which does not allow for variation within provinces. They define baskets for urban and rural households in each province, and then use the total expenditures for the given baskets to calculate the aggregate price over time; ideally, one should use chain weighted indexes to allow for changes in the basket. Second, they only use prices in the capital city of each province; as discussed, we find a big gap between large and small cities. Third, the price of OOH is estimated simply from the price of building materials. Fourth, their baskets are constructed using 1990 data. As shown in Figure 1, the consumption patterns in 2009 are very different.

We therefore estimate our own regional price levels using five different sources of data for 2009. The first is the National Development and Reform Commission (NDRC) surveys, which provide price data for many items, and we chose the ones for homogenous services such as local bus fares, taxi fares, telephone fees, and cable TV fees. There are such price data for almost all the cities in our sample for 2009.

The second source of prices is the unit values derived from the UHIES data on expenditures and quantities for fairly homogenous items such as food, clothing, water, electricity, and fuels. The prices of these items are also recorded in the NDRC surveys. To avoid the potential endogeneity of the imputed prices from the UHIES data, we compare the unit values with the NDRC prices. There are no significant differences between them. As there are more disaggregated items in the UHIES data, we use the unit values where available.

The third type of price data comes from provincial Development and Reform Commissions (DRC). The prices of some services are regulated by the provincial DRCs, and these are reported on their official websites[11]. These include medical services and school fees, and we collect the prices for all of our 17 regions for 2015. The fourth source of prices is the websites of service suppliers such as tutoring companies. We picked the prices of about 4 services in each city for 2015.

The remaining consumption items are more complex ones such as household goods, transportation equipment, communication equipment, medical goods, educational goods and recreational equipment. The prices of some of these items are recorded in the NDRC surveys, however, the products are not matched exactly, giving implausible prices. Given the lack of data on exactly matched items in different locations, we believe it is most straightforward to assume that these items have the same price in all locations in 2015. By 2015, purchasing online is very common in urban China, and almost all goods in the above list are available online. We believe that the small differences in delivery charges make our assumption a reasonable one, i.e. making 2015 the year to begin extrapolating prices backwards from a common starting level.

Finally, the time series data is the CPI for provinces and large cities from the *China Price Statistics Yearbook*. We first divide the cities into the large and small types. We calculate the "large-city" urban CPI by averaging the CPI from 36 large cities in our 9-province sample, weighted by total city consumption expenditures in each province. The *Yearbook* also provides the provincial average urban CPI. We use this provincial average and our estimated "large-city" index to impute a residual for the "small-city" urban CPI for each province using data on provincial consumption expenditures.

[11] E.g. Beijing medical services is given at service2.bjpc.gov.cn/bjpc/mediprice/MedicalService1.jsp#top

We compute the price of consumption bundles using a Tornqvist index following Slesnick (1998, 2002). The price of bundle θ in region *j* relative to the reference region B (for Beijing), P_j^θ, is defined as:

$$\ln P_j^\theta = \sum_{l=1}^{\theta_L} \frac{1}{2}(\omega_{lB} + \omega_{lj}) \cdot \ln(P_j^l / P_B^l) \tag{3.5}$$

where θ_L is the number of items in bundle θ, ω_{lj} is the expenditure share of item *l* in bundle θ in region *j*, P_j^l is the price of item *l* in *j*. All prices in Beijing in 2009 are normalized at 1.

To illustrate the price trends we constructed price bundles separately for the large and small-city groups, averaged over the 9 provinces using regional expenditure weights:

$$\ln P_{9t}^\theta = \sum_j \varepsilon_{jt}^\theta \cdot \ln P_{jt}^\theta \tag{3.6}$$

where ε_{jt}^θ is the expenditure share of region *j* on bundle θ in our 9-province sample. These 9-province average prices of the four bundles are given in Figure 2 with dark lines for large cities and light lines for small-cities. The prices in large cities are higher relative to those in small cities for almost all years. The prices in the large and small cities have the same trends for all bundles except for housing. The housing prices increase much faster in the large cities, especially since the housing reforms begun in 1998. Food price rose rapidly before 1997, became stable until 2003 and then rose again. The services price rose quickly before 2000 and then decelerated, in line with the general inflation trend. The prices for consumer goods also rose with the general inflation prior to 1997, but then fell, partly caused by the Information Technology price trends. We compare our results with Brandt and Holz's in Appendix A.4.

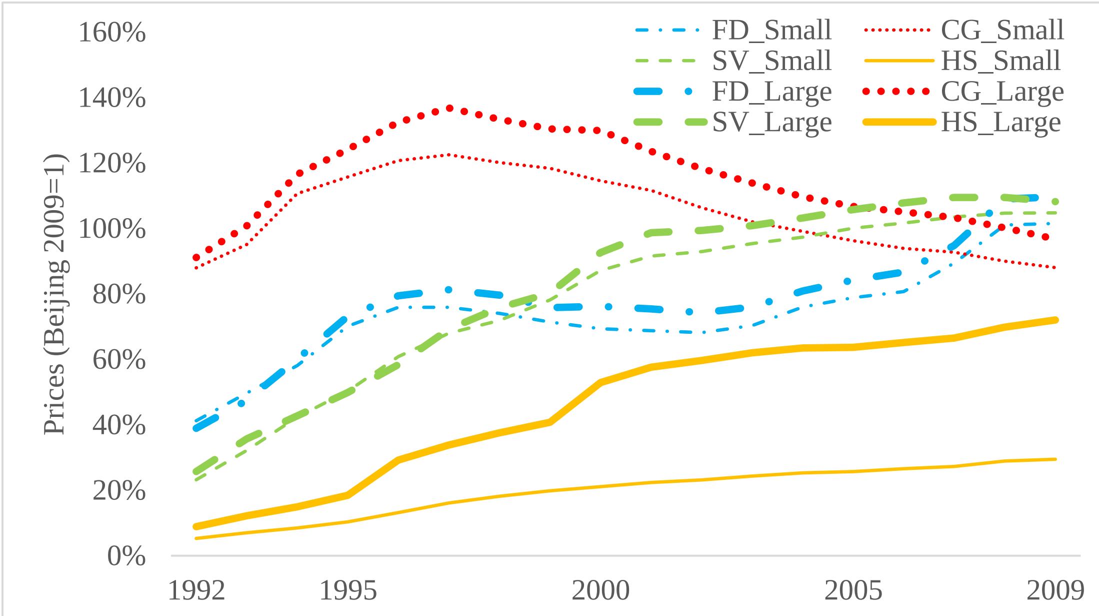


Figure 2. Consumption Prices (Small Cities Vs. Large Cities)

FD: Food, CG: Consumer Goods, SV: Services, HS: Housing

4. Aggregate Consumer Behavior

We estimate the parameters of the model using a demand system defined over four commodity groups described in the previous section. The demographic characteristics used to control for heterogeneity in household behavior include:

1. Age of household head: Under 35, 35-55, Above 55.
2. Gender of household head: Female, Male.
3. Occupation of household head: Private Sector, Public Sector.
4. Education of household head: Less than Secondary School, Secondary School, and College (or above).
5. Has Child: A 0-1 indicator showing if there is someone under age 16 in the household.
6. Has Aged: A 0-1 indicator showing if there someone aged 60 or older.
7. Number of members in the household: 1-2, 3, 3+.
8. Location: West, East and Middle.

In Table 1 we present summary statistics of the variables used. There are 184,000 observations over 1992-2009, with about 15,000 per year in the recent years. On average, food comprises 34% of the total expenditures, while services is lowest at 14%. The price of housing shows substantial variation in the sample as the stock price of houses rose dramatically. Male-headed households accounts for over 75% of the sample and 31% of the household heads have college degrees. 21% have an elderly member in the household.

Table 1. Sample summary statistics (Sample size: 183,564, 1992-2009)

Variable	Mean	Standard error	Minimum	Maximum
Food Share	0.342	0.119	0.005	0.875
Cons. Goods Share	0.242	0.095	0.006	0.820
Services Share	0.143	0.091	0.001	0.829
Housing Share	0.273	0.135	0.011	0.972
Log PFD	-0.330	0.264	-1.208	0.287
Log PCG	0.026	0.142	-0.403	0.463
Log PSV	-0.220	0.387	-1.861	0.211
Log PHS	-1.349	0.731	-3.973	0.000
Age 35-55	0.637	0.481	0	1
Age 55+	0.198	0.398	0	1
Male	0.756	0.429	0	1
Public employee	0.598	0.490	0	1
Sec. school	0.624	0.484	0	1
College	0.313	0.464	0	1
Has Child	0.438	0.496	0	1
Has Aged	0.215	0.411	0	1
Size 3	0.587	0.492	0	1
Size 3+	0.189	0.391	0	1
East	0.486	0.500	0	1
Middle	0.238	0.426	0	1

As explained in Section 2, we estimate the demand system using two methods – pooled cross-section and time series (2.13) and repeated cross-sections (2.14). We use nonlinear full information maximum likelihood with the housing equation omitted and subject to the constraints in (2.2, 2.4), symmetry, and the concavity constraint. In the single time series method, we first estimate the system without any curvature restrictions on the B matrix, and test it for concavity. In the repeated cross section method, the unrestricted estimates of B satisfy the concavity condition.

The parameters estimated are presented in Appendix A.5. The level of precision of the estimates from the repeated cross section method is high as would be expected given the large number of observations. However, some estimates in the pooled system are not significant; given that there is only one year of the cross-section.

Table 2. Price and Income Elasticities (standard errors in parenthesis)
(Reference Household: 35-55, Male, Private sector, Secondary School, Has Child, No Aged, Size 3, East)

	Good	Uncompensated Price Elasticity	Compensated Price Elasticity	Expenditure Elasticity
Pooled Method	Food	-0.152 (0.064)	-0.072 (0.065)	0.737 (0.022)
	Consumer Goods	-0.151 (0.043)	-0.111 (0.044)	1.237 (0.024)
	Services	-0.456 (0.166)	-0.228 (0.167)	1.389 (0.046)
	Housing	-0.320 (0.050)	-0.034 (0.052)	0.892 (0.021)
Repeated Cross Section Method	Food	-0.366 (0.018)	-0.151 (0.018)	0.709 (0.008)
	Consumer Goods	-0.707 (0.012)	-0.422 (0.013)	1.339 (0.010)
	Services	-0.524 (0.029)	-0.287 (0.029)	1.442 (0.013)
	Housing	-0.564 (0.011)	-0.300 (0.012)	0.824 (0.007)

Table 2 presents the price and income elasticities (2.15-2.17) calculated for the reference household in 2002: number of members=3, with child, no aged member, East, and head of household is male, aged 35-55, secondary school educated, and employed in the private sector.

The expenditure (income) elasticities are estimated with very small standard errors and are similar under the two methods; food is income inelastic (0.74), as is housing, which includes utilities (0.89). While the housing share rose rapidly prior to 2003, this was driven by prices that rose even faster than the high growth in incomes. Consumer goods (1.24) and services (1.39) are income elastic. This pattern of income elasticities is similar to estimates of other countries (e.g. Jorgenson and Slesnick 2008).

The own compensated price elasticities are negative for all goods in both methods. From the repeated cross-section method, food is price inelastic (-0.15), while consumer goods have much higher elasticitiy (-0.42). The elasticities for services and housing are in between these values. The estimates for the pooled method are driven by the time series data and show smaller price elasticities, furthermore, some elasticities are not statistically different from zero.

5. Projecting Aggregate Consumption

Detailed projections of the economy are a key component of policy analysis, that is, we need, not just GDP, but also consumption and other variables. In this section, we project how the demographic changes (aging population and education improvements) as well as the change of prices and incomes, affect the expenditure shares of the urban households.

To do this, we project a set of exogenous variables that provide the distribution terms in the aggregate consumption share equations, as done in Jorgenson, Goettle, Ho and Wilcoxen (2013) for the U.S. The vector of aggregate shares for Food, Consumer Goods, Services and Housing is given in (2.5) which we rewrite as:

$$\omega_t = \frac{1}{D(p_t)} \cdot (\alpha_p + B \cdot \ln p_t - B_M \cdot (\xi_t^d + \ln M_t) + B_A \cdot \xi_t^L) \qquad (5.1)$$

where

$$\xi_t^d = \sum_k \frac{m_{kt}}{M_t} \cdot \ln \frac{m_{kt}}{M_t} \tag{5.2}$$

$$\xi_t^L = \sum_k m_{kt} \cdot A_k / M_t \tag{5.3}$$

m_k is the expenditure of household k, and $M_t = \sum_k m_k$. The distribution terms represent the effects of the changing composition of families on aggregate consumption demands; ξ_t^d represents the income distribution term and ξ_t^L represents the effect due to different household types having different baskets when faced with the same prices.

To project (5.1) we rewrite it in terms of group *K*, denoting the demographic categories used in the consumption model – the age of the head(a), sex of the head(s), job of the head(j), education of the head(e), has a child (c), has an aged(g), number of members(n), and location(l). For example, the number of households of type *K* is $n_K \equiv n_{asjecgnl}$.

For the sample period the distribution terms, ξ_t^d and ξ_t^L, are simply obtained from the UHIES sample. Beyond the sample period we assume that the relative average expenditures by type of households are fixed, that is, that the ratios m_{K_1}/m_{K_2} are fixed, but allow the number of households of type *K* to change according to population projections. We first define the mean expenditure shares of group *K* for the last year of the sample:

$$\bar{m}_K^0 = \frac{\bar{m}_{K,T=2009}}{M_{T=2009}} \tag{5.4}$$

We then fix the m_{kt}/M_t term in (5.2) and (5.3) at the base year share $\bar{m}_K^0$ and write the distribution terms for a future year t as:

$$\xi_t^d = \sum_k \frac{m_{kt}}{M_t} \cdot \ln \frac{m_{kt}}{M_t} = \sum_K n_{Kt} \cdot \frac{\bar{m}_K^0}{M_t^0} \cdot \ln \frac{\bar{m}_K^0}{M_t^0} \qquad \text{t=2010,2011,...} \tag{5.5}$$

$$\xi_t^L = \sum_k m_{kt} \cdot A_k / M_t = \sum_K n_{Kt} \cdot \frac{\bar{m}_K^0}{M_t^0} A_k \tag{5.6}$$

Details about the projection of ξ_t^d and ξ_t^L are in Appendix A.6.

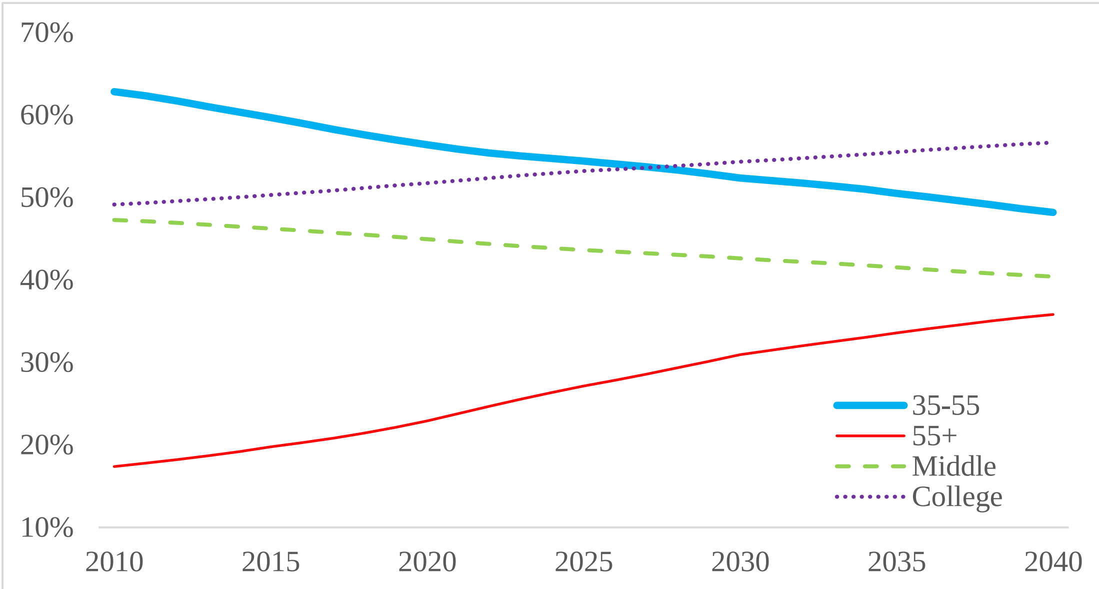


Figure 3. Projection of the shares of total expenditures by household type

In Figure 3, we plot the expenditure shares by age and education of the head of household. We project a rising share of older and better-educated households; the expenditure share of households with head over age 55 rises from 17.4% in 2010 to 35.8% in 2040, while the share of households with a college-educated head rises from 49.2% in 2010 to 56.66%.

The consumption function with the projections of the distribution terms are then combined with the projections of prices and incomes from a multi-sector growth model of China that has been used to analyze energy policies in Cao et al. (2017). This model endogenously calculates the prices of various commodities as well as household income based on a related consumption sub-model that covers both urban and rural households.

The projected price vector of the 4 consumption bundles, $\hat{p}_t$, and national disposable income is plugged into (5.1) giving the projected shares:

$$\omega_t = \frac{1}{D(\hat{p}_t)} \cdot (\alpha_p + B \cdot \ln \hat{p}_t - B_M \cdot (\xi_t^d + ln\widehat{M}_t) + B_A \cdot \xi_t^L) \tag{5.8}$$

The change in the expenditure share vector between the base year and t is:

$$\Delta\varepsilon_t = \frac{1}{D(\hat{p}_t)} \cdot (\alpha_p + B \cdot \ln \hat{p}_t - B_M \cdot (\xi_t^d + ln\widehat{M}_t) + B_A \cdot \xi_t^L)$$
$$-\frac{1}{D(\hat{p}_0)} \cdot \left(\alpha_p + B \cdot \ln \hat{p}_0 - B_M \cdot \left(\xi_0^d + ln\widehat{M}_0\right) + B_A \cdot \xi_0^L\right) \tag{5.9}$$

$\frac{1}{D(\hat{p}_t)}$ is very close to $\frac{1}{D(\hat{p}_0)}$, and when normalized to 1 in the base year, $D(\hat{p}_t) = -1$. With this simplification, the change in the expenditure shares is:

$$\Delta\varepsilon_t = -B \cdot \Delta \ln \hat{p}_t + B_M \cdot (\Delta\xi_t^d + \Delta \ln M) - B_A \cdot \Delta\xi_t^L \tag{5.10}$$

where $-B_{pp} \cdot \Delta \ln \hat{p}_t$ is the price effect, $B_M \cdot (\Delta\xi_t^d + \Delta \ln M)$ is the income effect, and $-B_{pA} \cdot \Delta\xi_t^L$ is the demographic effect.

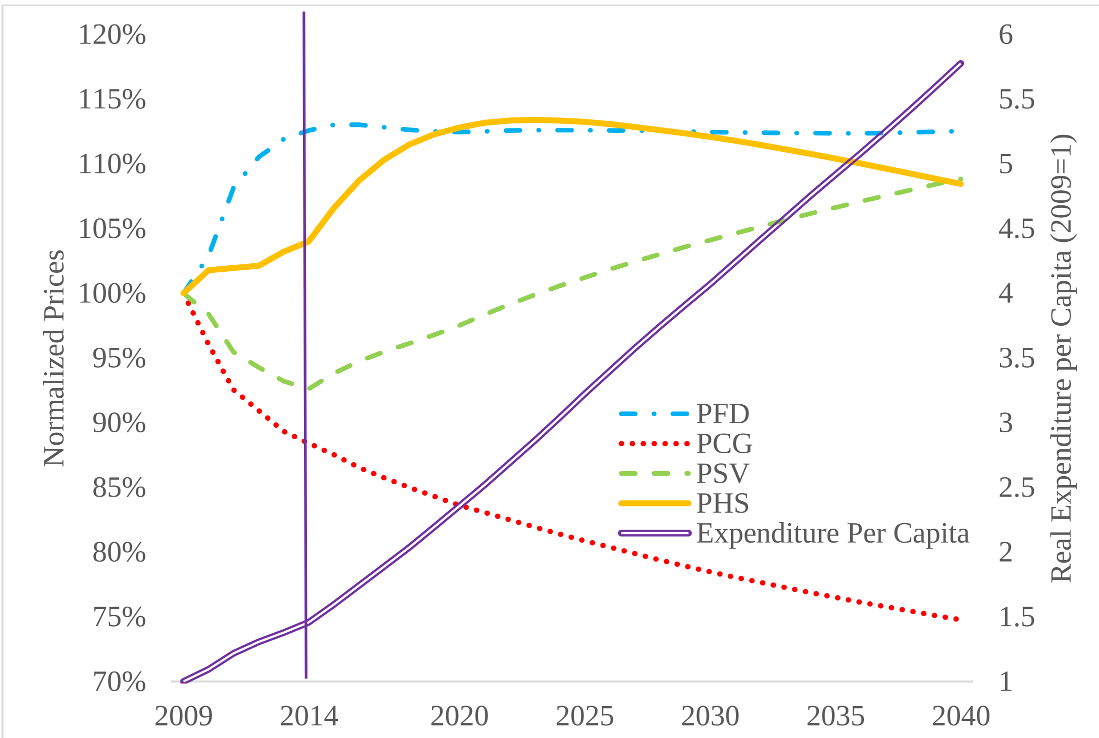
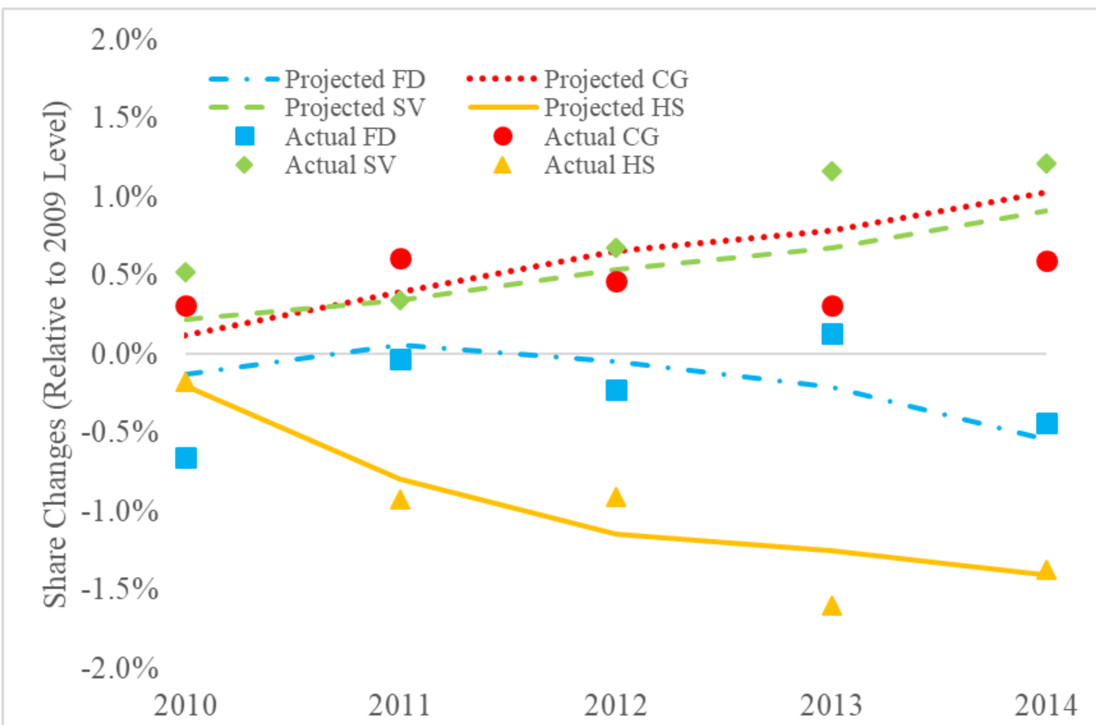


Figure 4. Projection of Prices and Expenditures
FD: Food, CG: Consumer Goods, SV: Services, HS: Housing

In Figure 4, we plot the prices and expenditure per capita from Cao et al. (2017). The model has a 2014 base year, and we add in the actual CPI for 2009-2014. The prices are normalized to 1 in 2009. The total expenditure is taken from the growth of real consumption per capita from the National Accounts. Expenditure per capita rises over the whole projection period, but decelerates with the slowing GDP growth. By 2040, the real expenditure per capita is almost 6 times the 2009 level. The model projects that the relative price of consumer goods will fall substantially while the price of services will rise. The projected behavior after 2014 is somewhat different from actual 2009-2014 price trends which were dominated by the aftershocks of the Financial Crisis. By 2040, the price of consumer goods (PCG) relative to food will be only two-thirds of the 2009 relative price. The housing price first rises, peaking around 2025, and then falling; this is due to the rate of capital accumulation, which eventually lowers the cost of capital (the model does not take into account land prices). We should also note that the projected population starts falling around 2025.

Plugging the prices and expenditures from the simulation into (5.3), we get the three effects for aggregate consumption. We compare our simulated results with the actual national urban consumption shares during 2010 to 2014. We first adjust the national data to match our definitions of the 4 consumption bundles.[12]



[12] We split the transportation, communication, medical and education categories in the China Statistical Yearbook (CSY) into two parts – goods versus services – using the relative shares in our detailed household data in 2009. We impute the rental equivalent of OOH but the (CSY) tables derived from the household survey do not. However, the National Accounts estimate of Urban Consumption does include a housing imputation, and so we use it to give the rental equivalents for 2010-14. The NBS also changed its methods for estimating in-kind consumption and housing after 2013 and we adjust for that.

Fig 5. Change in expenditure shares relative to 2009; Projected versus Actual.

The consumption shares from the actual 2010-14 surveys and those projected from our estimated parameters are plotted in Figure 5. On average, we overestimate the consumer goods share and underestimate the service share a little. However, the actual shares change in the same way as we project, and the errors are less than 1.5 percentage points, five years out.

The price, income and demographic effects on consumption shares are shown in Appendix A.7, Figures A.2-A-5. With rapid rise of food prices, the price effect is strong for all 4 bundles in the first few years. However, the income effect gradually dominates the price effect as income keeps growing but changes in relative price moderates. Interestingly, the net demographic effect is tiny over the whole period, though the demographic distribution changes significantly. Two opposite effects occur as the population becomes older and better educated. A household with an aged head prefers spending more on food and housing, while a household with a better-educated head prefers consumer goods and services. These two trends cancel each other and the net demographic effect is small over the whole period.

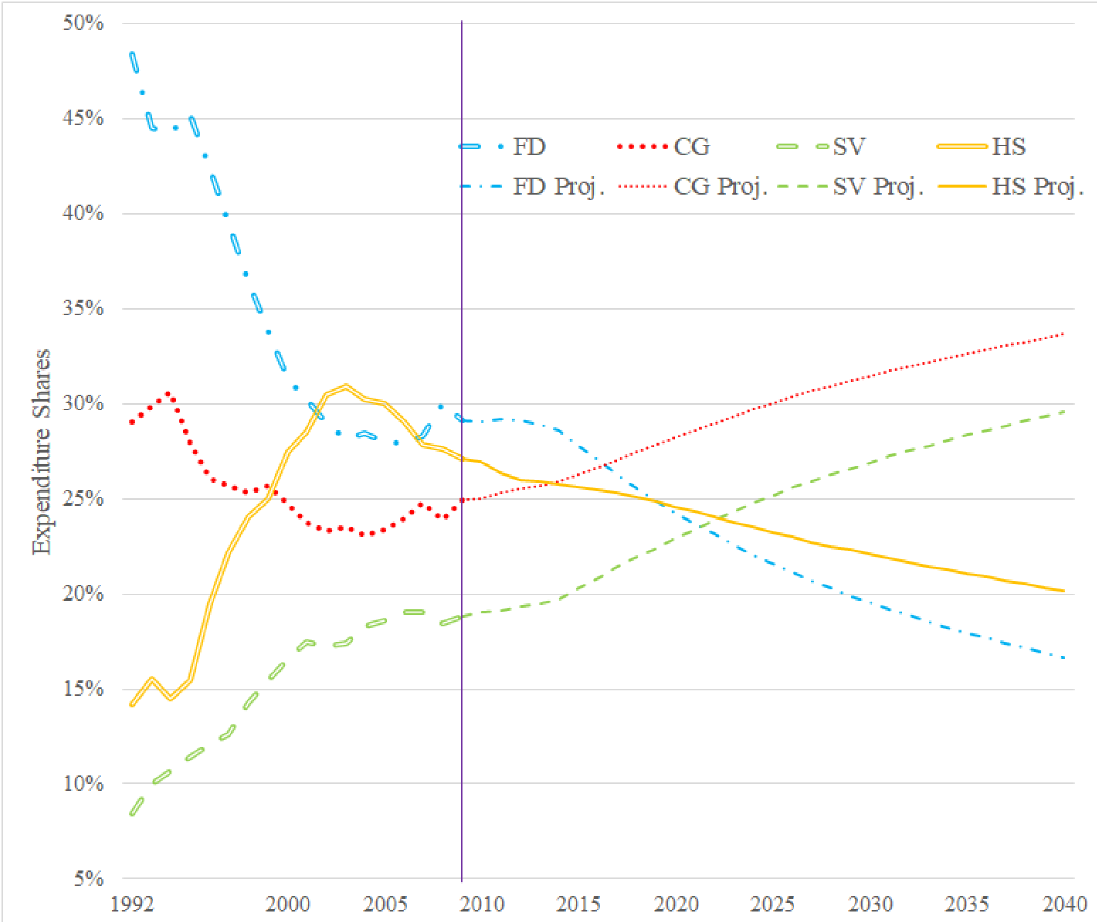


Figure 6. Expenditure Shares; sample period and projections.

In Figure 6, we plot the projected expenditure shares. The shares before 2010 are the actual shares from our sample, the shares for 2010-13 are from Figure 5, and the rest are simulated from the growth model. The services share continue their historical rise due to higher incomes, increasing from 19% in 2009 to 29% in 2040. Consumer goods trended like services in the projection period, unlike the historical fall in its share due to falling goods prices. Its rises from 25% in 2009 to 33.7% in 2040.

As necessities, food and housing shares fall with the rapidly rising incomes. The food share drops from 29% in 2014 to 16.6% in 2040. Note that this projected fall is in contrast to the somewhat stable share between 2002 and 2014 that is driven the high food prices and demographic effects. The housing share continues to fall, driven by the income elasticity of 0.8 (from 26% in 2014 to 21% in 2040). This is quite different from the historical behavior when it rose from 14.2% in 1992 to 30.5% in 2002 due to the rising house prices.

6. Conclusion

Consumption provides an alternative to income as a measure of material well-being. Disaggregated consumption data allows us to measure better the changes in prices and behavior brought about by the rapid economic growth in China. We have estimated a household demand function that can be consistently aggregated to national consumption. This is the first effort to this kind of research in China. Such a function allows a more precise measure of historical changes in welfare, and a more refined analysis of future economic policies.

We integrated 18 years of household survey data with information on price levels that vary across regions and time. We made a special effort to impute rents for the majority of households who own their homes, thus avoiding a common underestimate of consumption expenditures. The resulting data set cover over 180,000 households, allowing us to estimate the demand function that covers all of consumption in urban China. The large sample size and long time series enable us to create synthetic cohorts that facilitate the estimation. Using both synthetic cohorts and even longer time series of national aggregates allow us to contrast the price effects from cross-sectional variation with time-series variation.

Food is estimated to be less price elastic than the other three commodity bundles. Housing and services have very similar price elasticities. Consumer goods and services are income elastic, while food and housing are income inelastic.

We projected consumption patterns using the estimated parameters and a projection of the demographic changes coupled with a growth model that simulates commodity prices and household disposable incomes. The growth model gives us the income effects as well as the projected increase in service prices relative to goods prices. The net effect is a substantial projected fall in the food and housing shares and accompanied by rising consumer goods and services shares. We show how the demographic changes of aging and improving education cancel each other out.

While we have to employ various simplifying assumptions due to the lack of various data, we believe that the above estimates are informative. The results are statistically significant and accords with our expectations of price and income responses. From this empirical foundation, we aim to construct measures of social welfare and inequality as in Jorgenson and Slesnick (1984). Much remains to be done, in particular, extending the analysis to rural consumption that confronts even greater data challenges.

Appendix

Appendix A.1: Classification of Expenditures

Table A1.Classification of Expenditures in the UHIES

FD (Food)	**SV (Services)**
Food and tobacco	Transportation Service
Dining out	Transport Fuel
	Transportation fees
CG (Consumption Goods)	Vehicle Maintenance
Clothing	Communication Services
Household Facilities and Articles	Communication Fee
Transportation equipment	Postage Fee
Vehicles	Medical Services
Communication Goods	Health Care
Medical Goods	Education Services
Medicine	Tuition
Medical Devices	Child-Care Fee
Education Goods	Recreation Services
Recreation Goods	Tourism
Recreational Durables	Exercise
Recreational Articles	**HS (Housing)**
Magazines	Housing rents
Miscellaneous Goods	Water, Electricity and Fuels

Appendix A.2: 9-Province Sample Households vs. National Households

We compare the expenditure shares of our 9-Province sample with the national ones to show the degree of similarity. The national shares of urban expenditures are given in the *China Statistical Yearbook* (NBS 2010, Table 10-16).

Our 4 consumption bundles are built up from hundreds of detailed items as illustrated in Table A1 and do not have a simple link to the broad categories in the *Yearbook* tables. We rearranged our detailed data to conform with the published tables. In Table A2, we compare the expenditures per capita and expenditure shares from our sample with the national estimates for 2009. The absolute expenditures per capita in our sample are a about 10% higher for most categories. However, in terms of expenditure shares, they are much closer. For the most important group – food – they are identical at 38.0%, for transportation (goods and services) the shares are 9.4% versus 8.8%.

Next, we compare the time trends of the expenditure shares. For this purpose, we constructed 4 bundles out of the *Yearbook* categories in the following manner:

1. Food: Food
2. Goods: Cloth, Household Facility, Recreation Goods, Other Goods

3. Services: Transportation, Communication, Medical, Education, Recreation Service, Other Service
4. Water, electricity and fuel: Water & Electricity and fuel

Table A2. Comparison of expenditure per capita and shares in 2009

Category	Expenditure / Capita		Shares	
	Sample	Yearbook	Sample	Yearbook
Food	4988.31	4476.89	38.0%	38.0%
Cloth	1313.48	1276.43	10.0%	10.8%
Household Facility	717.7875	731.96	5.5%	6.2%
Transportation	1229.2753	1040.88	9.4%	8.8%
Communication	724.04716	641.7	5.5%	5.4%
Medical	901.6655	856.41	6.9%	7.3%
Education	743.26949	645.89	5.7%	5.5%
Recreation Goods	392.6002	381.32	3.0%	3.2%
Recreation Service	547.3996	445.56	4.2%	3.8%
Water, electricity and fuel	761.576	746.62	5.8%	6.3%
Miscellaneous Goods	445.4874	308.56	3.4%	2.6%
Miscellaneous Service	366.0941	230.05	2.8%	2.0%

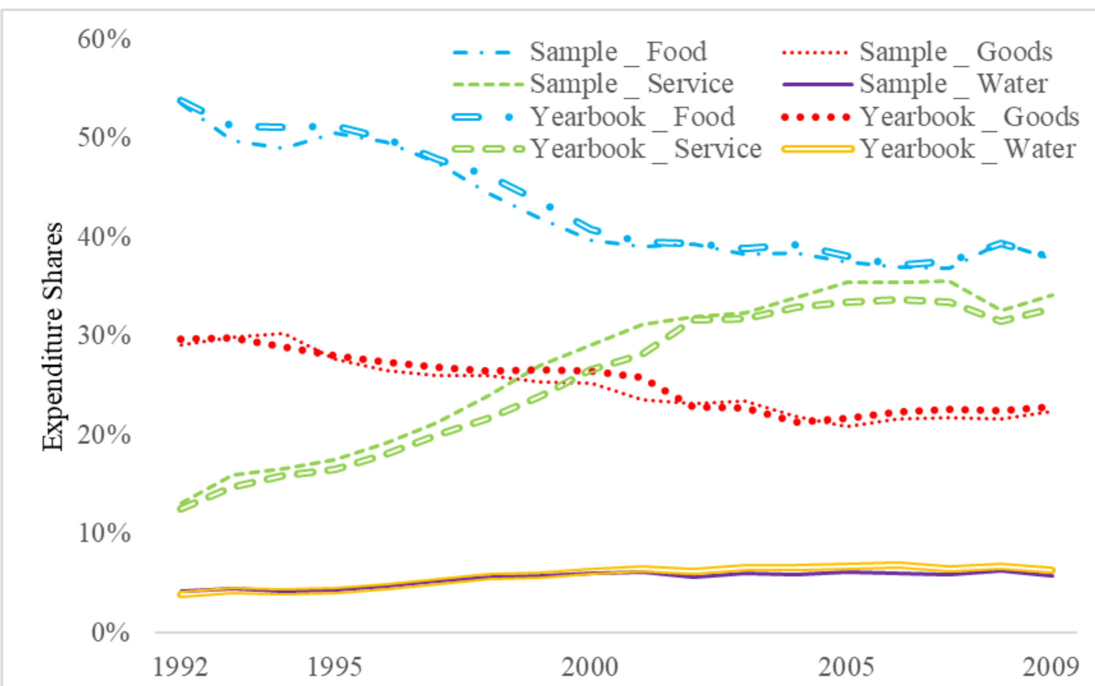


Fig A1. Comparison of expenditure share trends between our sample and national estimate in the Statistical Yearbook (1992-2009).

This allocation does not exactly match our bundles since the transportation, communication and education categories include both goods and services The fourth bundle does not include housing rental equivalents since the official tables do not include such an item.

The shares from our sample and the national shares are plotted in Figure A1. We can see that the differences are very small over the whole period.

Appendix A.3: Weights of the Sample Households

From 2002 to 2009, the UHIES provides a raw weight for each household, but from 1992 to 2001, no weights were given. The raw weights cannot be used directly, because they are the relative weights within a city.

We rescale the weights by requiring that the samples from a given group should be representative of the group. Thus, for region *r* in period *t* we have:

$$\sum_{i=1}^{I_{r,t}}(fw_{i,t} \cdot size_{i,t}) = UrbanPop_{r,t} \quad \text{(A.1)}$$

where $fw_{i,t}$ is the estimated weight of household i, $size_{i,t}$ is the size of the household and $UrbanPop_{r,t}$ is the urban population of region r.

To make use of the raw weights, we have:

$$fw_{i,t} = \rho_{r,t} \cdot raw_{i,t} \quad \text{(A.2)}$$

where $raw_{i,t}$ is the raw weight of household i, and $\rho_{r,t}$ is the calibrating parameter of region r, at time t.

The regions are defined in section 3.1. Each province (except Beijing) is split into two regions: large cities and other cities.

Appendix A.4: Regional Price Levels

Table A3 gives the relative price levels for large and small cities in 1992 and 2009. The prices in large cities are always higher than the prices in small cities within a province. For most provinces except Anhui, the price differences between large and small cities became bigger during the 18 years. In 1992, the prices in Guangdong were much higher than the other cities, while in 2009, the price in Beijing was the highest. In addition, the price differences among different provinces got bigger during the period.

Table A3.Price Levels for Large Cities and Small Cities in the 9 Provinces

Province	1992		2009	
	Large City	Small City	Large City	Small City
Beijing	100.0%	--	100.0%	--
Liaoning	101.3%	90.6%	84.5%	66.5%
Zhejiang	102.9%	95.9%	95.8%	77.4%
Anhui	98.4%	80.6%	72.9%	61.8%

Hubei	101.8%	80.8%	79.5%	57.9%
Guangdong	141.3%	114.8%	99.3%	72.9%
Sichuan	91.0%	76.4%	82.3%	61.1%
Shaanxi	91.2%	75.5%	72.4%	60.5%
Gansu	81.5%	69.7%	63.5%	54.5%

Brandt and Holz (2006) is an oft-cited source of regional price comparisons using data from 1992. We extrapolate their prices using provincial CPIs, and compare our results with theirs in Table A4. All prices are relative to Beijing levels. Our prices for the less developed provinces (Anhui, Sichuan, Shaanxi, Gansu) are lower than their estimates for both 1992 and 2009. The comparison for the rich Guangdong and Zhejiang provinces are mixed.

Table A4. Comparison with Brandt and Holz (2006)

	1992		2009	
	Ours	Brandt, Holz	Ours	Brandt, Holz
Beijing	100.0%	100.0%	100.0%	100.0%
Liaoning	95.4%	94.8%	75.7%	84.4%
Zhejiang	98.1%	93.3%	84.1%	88.0%
Anhui	82.9%	89.3%	63.5%	82.0%
Hubei	87.2%	94.1%	65.0%	92.2%
Guangdong	122.4%	123.3%	83.0%	102.9%
Sichuan	80.1%	87.7%	67.8%	90.9%
Shaanxi	83.1%	94.9%	65.4%	87.5%
Gansu	74.3%	91.9%	57.4%	85.8%

Appendix A.5: Estimation Results

In Tables A5 and A6, we show the estimation results from the pooled cross-section and time series method, and repeated cross section method respectively. Recall that the cross section consist of prices from 17 regions per year, while the prices in the pooled method span 1981 to 2011 with the assumption that all regions face the same price.

The price coefficient estimates from the pooled method have larger standard errors, and some are statistically insignificant. The repeated cross section coefficients are more precisely estimated given the larger degrees of freedom and most of the estimates are statistically significant.

Table A5. Estimates of pooled cross-section and time series

Variable	Estimate	SE	Estimate	SE
	Food		Consumer Good	
CONST	-0.342	0.0226	-0.264	0.0241

Log PFD	-0.279	0.0311	0.086	0.0095
Log PCG	0.086	0.0095	-0.168	0.0127
Log PSV	0.048	0.0224	0.049	0.0095
Log PRD	0.065	0.0210	0.083	0.0106
Log EXPEN	-0.079	0.0026	0.050	0.0023
35-55	0.003	0.0071	0.013	0.0071
55+	-0.004	0.0132	0.006	0.0132
MALE	-0.014	0.0030	0.011	0.0030
PUBLIC	-0.004	0.0031	-0.018	0.0031
SECONDARY SCHOOL	0.007	0.0037	-0.003	0.0037
COLLEGE	0.029	0.0049	-0.010	0.0050
HAS CHILD	0.005	0.0036	-0.006	0.0036
HAS AGED	-0.010	0.0076	0.008	0.0076
SIZE 3	-0.026	0.0046	0.039	0.0046
SIZE 3+	-0.059	0.0082	0.071	0.0081
EAST	-0.009	0.0075	0.037	0.0059
MIDDLE	-0.024	0.0056	-0.019	0.0056
	Service		Housing	
CONST	-0.101	0.0220	-0.293	0.0229
Log PFD	0.048	0.0224	0.065	0.0210
Log PCG	0.049	0.0095	0.083	0.0106
Log PSV	-0.078	0.0238	0.045	0.0186
Log PRD	0.045	0.0186	-0.227	0.0167
Log EXPEN	0.063	0.0021	-0.034	0.0023
35-55	-0.008	0.0071	-0.008	0.0071
55+	0.015	0.0132	-0.018	0.0132
MALE	-0.003	0.0030	0.006	0.0030
PUBLIC	-0.005	0.0031	0.027	0.0031
SECONDARY SCHOOL	-0.004	0.0037	0.001	0.0037
COLLEGE	-0.007	0.0050	-0.012	0.0050
HAS CHILD	-0.010	0.0036	0.011	0.0036
HAS AGED	-0.003	0.0076	0.005	0.0076
SIZE 3	-0.016	0.0046	0.002	0.0046
SIZE 3+	-0.014	0.0082	0.002	0.0082
EAST	0.007	0.0074	-0.035	0.0069
MIDDLE	-0.002	0.0056	0.045	0.0056

Table A6. Estimates of Repeated Cross Section Method

Variable	Estimate	SE	Estimate	SE
	Food		Consumer Good	

CONST	-0.327	0.0067	-0.154	0.0065
Log PFD	-0.202	0.0055	0.030	0.0032
Log PCG	0.030	0.0032	-0.043	0.0027
Log PSV	0.071	0.0038	0.032	0.0025
Log PRD	0.020	0.0042	0.048	0.0028
Log EXPEN	-0.081	0.0025	0.066	0.0021
35-55	0.001	0.0028	0.011	0.0032
55+	-0.017	0.0041	0.016	0.0047
MALE	-0.005	0.0023	0.007	0.0033
PUBLIC	-0.003	0.0025	-0.013	0.0029
SECONDARY SCHOOL	0.014	0.0041	-0.012	0.0049
COLLEGE	0.032	0.0045	-0.018	0.0050
HAS CHILD	0.009	0.0022	-0.011	0.0031
HAS AGED	-0.010	0.0032	-0.001	0.0039
SIZE 3	-0.028	0.0031	0.032	0.0039
SIZE 3+	-0.052	0.0035	0.073	0.0044
EAST	-0.006	0.0026	0.003	0.0035
MIDDLE	-0.006	0.0031	0.012	0.0033
	Service		Housing	
CONST	-0.059	0.0067	-0.460	0.0067
Log PFD	0.071	0.0038	0.020	0.0042
Log PCG	0.032	0.0025	0.048	0.0028
Log PSV	-0.061	0.0046	0.026	0.0036
Log PRD	0.026	0.0036	-0.146	0.0035
Log EXPEN	0.067	0.0021	-0.052	0.0023
35-55	-0.011	0.0037	-0.001	0.0032
55+	0.007	0.0044	-0.006	0.0044
MALE	0.003	0.0029	-0.005	0.0028
PUBLIC	-0.001	0.0028	0.018	0.0027
SECONDARY SCHOOL	-0.007	0.0043	0.004	0.0044
COLLEGE	-0.013	0.0054	-0.001	0.0050
HAS CHILD	-0.015	0.0028	0.017	0.0027
HAS AGED	0.008	0.0032	0.002	0.0034
SIZE 3	-0.008	0.0035	0.005	0.0035
SIZE 3+	-0.003	0.0040	-0.018	0.0040
EAST	-0.005	0.0032	0.009	0.0031
MIDDLE	0.000	0.0036	-0.006	0.0033

Appendix A.6: Projection of demographic distribution terms

The projection of ξ_t^d and ξ_t^L requires a projection of the number of households of each type, $n_{asjecgnl}$. We construct a household bridge matrix (H) that links the distribution of household types to a population matrix of dimension 2 sexes, 16 age groups, and 3 education levels based on the 2010 population census and labor survey data. The population projection for these dimensions is from Cao, Ho, and Hu (2017). We assume that the distribution for location, household size, presence of child and aged, and employment type remain unchanged in the future, given the relative stability in the recent years. We focus on two characteristics that will likely change the most: age of the head; and educational attainment of the head.

The bridge matrix (H) links the population by age-sex-education to the number of households by age of head and education of head:

$$nf_{ae,t} = H_{ase}^{ae} \cdot \Lambda_{ase}^{ase} \cdot Pop_{ase,t} \quad \text{(A.3)}$$

$Pop_{ase,t}$ is the population matrix transformed into a vector by rearranging the indexes. Λ_{ase}^{ase} is a diagonal matrix that links the population distribution in our sample to the population distribution in the census data. The elements on the diagonal are the ratios for each type of population in our sample and the census data, with 0 off the diagonal. That is, $SamplePop_{ase,2009} = \Lambda_{ase}^{ase} \cdot Pop_{ase,2009}$.

We construct H_{ase}^{ae} from our sample in the following ways:

1. We define a matrix TX that allocates population into each type of household;

$$\mathrm{TX(I,J)} = \sum_k \sum_{i\in k} 1(k = I, k_i = J)/\sum_k \sum_{i\in k} 1(k_i = J) \quad \text{(A.4)}$$

where I is the type of household k, J is the type of population for member i in household k.

2. We define a matrix SX:

$$\mathrm{SX} = \mathrm{diag}\left(\frac{1}{Size_I}\right) \quad \text{(A.5)}$$

where $Size_I$ is the average size of household type I.

Then the bridge matrix H_{ase}^{ae} can be defined as:

$$H_{ase}^{ae} = \mathrm{SX} \cdot \mathrm{TX} \quad \text{(A.6)}$$

Appendix A.7: Three Effects for the Expenditure Shares in Projection

The change in expenditure shares over the projection period is driven by 3 effects: changes in demographic characteristics of households, the rise in disposable incomes and changes in relative prices. Figures A2 to A5 give the demographic effect, income effect and price effect for the four consumption bundles – food, consumer goods, services, and housing, respectively.

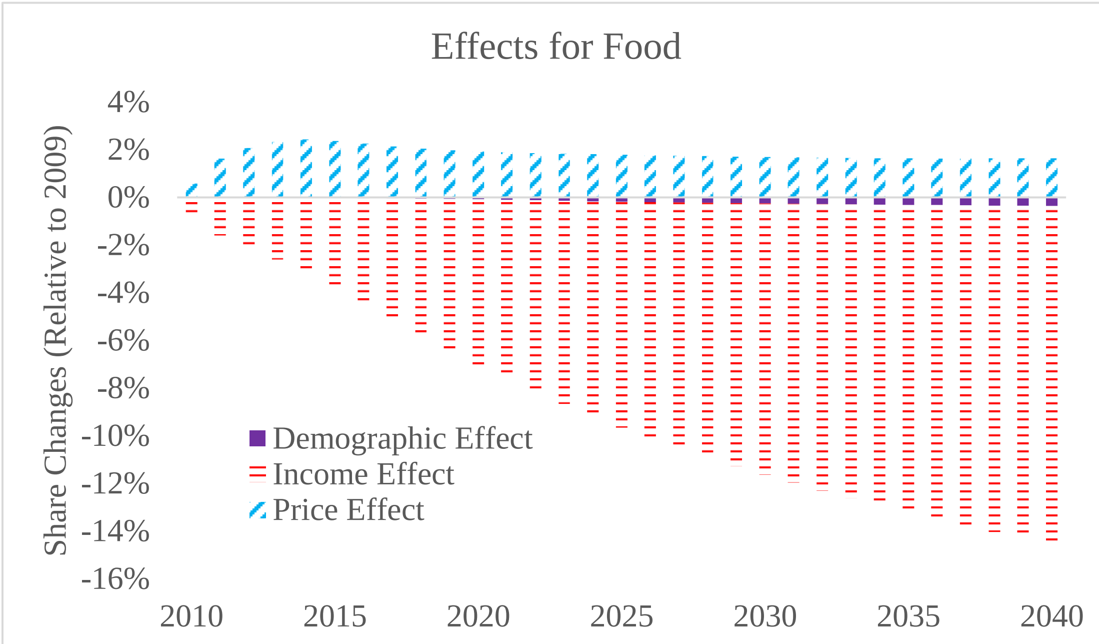


Figure A2. Effects for food expenditure share projections.

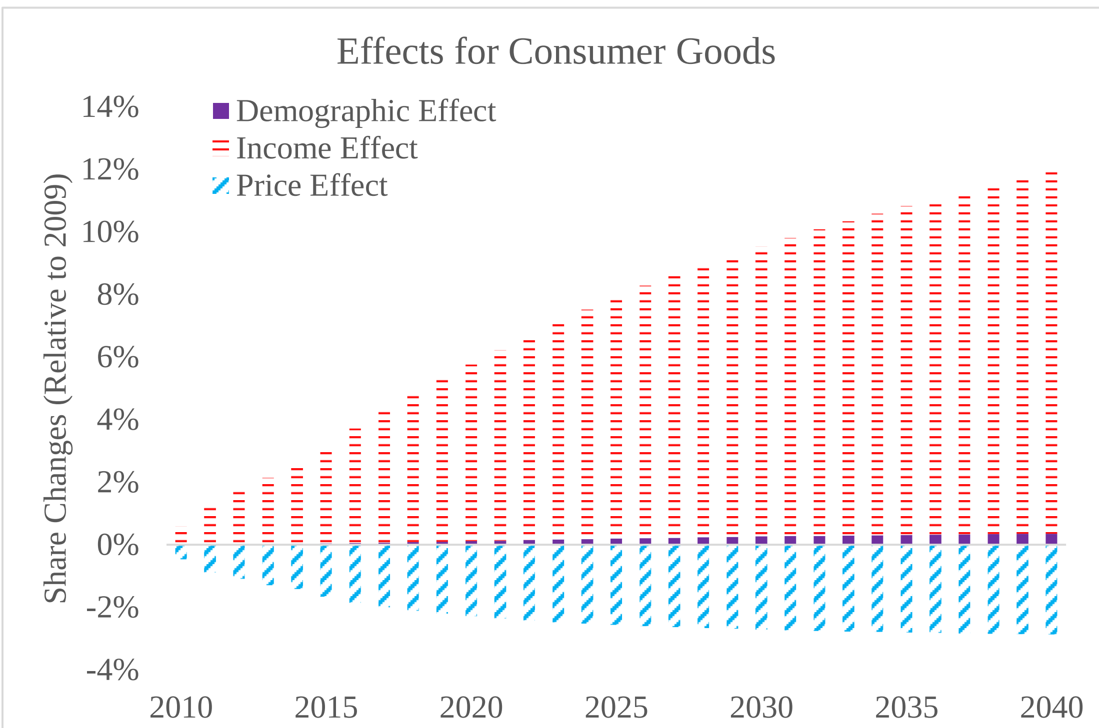


Figure A3. Effects for Consumer Goods expenditure share projections.

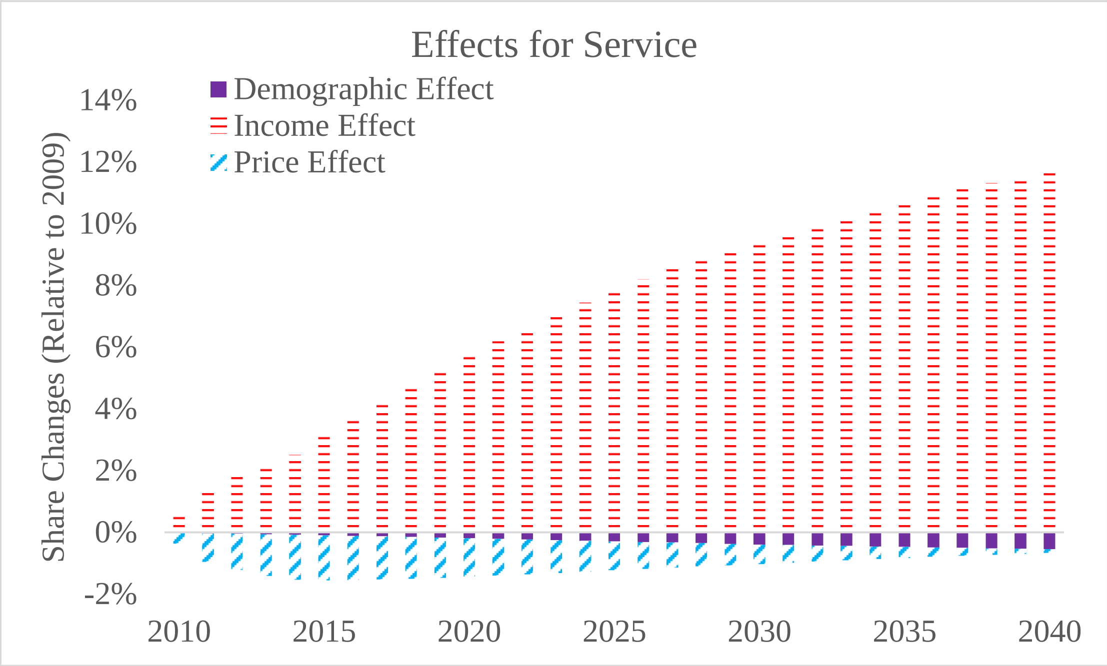


Figure A4. Effects for Services expenditure share projections.

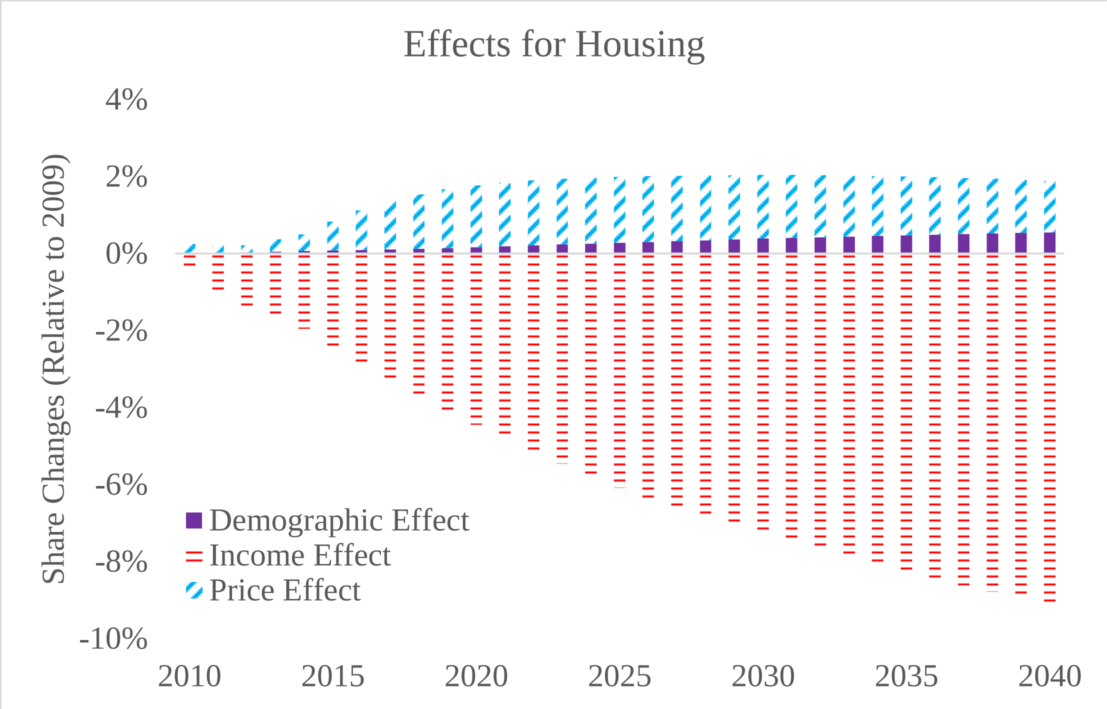


Figure A5. Effects for Housing expenditure share projections.

References

Attanasio, O., Pistaferri, L. (2016), "Consumption Inequality ", *Journal of Economic Perspectives*, 30(2), Spring, 1-27.

Brandt, L., Holz, C. (2006), "Spatial Price Differences in China: Estimates and Implications", *Economic Development and Cultural Change*, 55(1), 43-86.

Cao, J., Ho, M., Liang, H. (2016), "Household energy demand in Urban China: Accounting for regional prices and rapid income change", *Energy Journal*, 37, 87-110.

Cao, J., Ho, M., Hu W. (2017), "The Supply Side Outlook for China is not so Gloomy," Harvard China Project on Energy, Economy and Environment Working Paper.

Cao, J., Ho, M,, Jorgenson, D. (2017), "China Emissions Trading System and Carbon Tax Hybrid", Harvard China Project on Energy, Economy and Environment Working Paper.

Chamon, M., Prasad, E. (2010), "Why Are Saving Rates of Urban Households in China Rising?", *American Economic Journal: Macroeconomics*, 2(1), 93-130.

Chen, Y., Xing, W. (2011), "Quantity, quality, and regional price variation of cigarettes: Demand analysis based on a household survey in China", *China Economic Review*, 22(2), 221-232.

Fan, S., Wailes, E., Cramer, G. (1995), "Household Demand in Rural China: A Two-Stage LES-AIDS Model", *American Journal of Agricultural Economics*, 77(1), 54-62.

Holt, M., Goodwin, B. (2009), "The Almost Ideal and Translog Demand Systems", Chap 2, *Quantifying consumer preferences,* Emerald Press.

Hovhannisyan, V., Gould, B. (2014), "Structural Change in Urban Chinese Food Preferences", *Agricultural Economics*, 45(2), 159-166.

Jorgenson, D. (1998), "Did We Lose the War on Poverty?", *Journal of Economic Perspectives*, 12(1), Winter, 79-96.

Jorgenson, D., Lau, L., and Stoker, T., (1982), "The Transcendental Logarithmic Model of Aggregate Consumer Behavior," in Advances in Econometrics, ed. by Basmann, R. and Rhodes, R. Jr., Vol. 1. Greenwich: JAI Press.

Jorgenson, D., Slesnick, D. (1984), "Aggregate Consumer Behavior and the Measurement of Inequality", *Review of Economic Studies*, 51(3), 369-392.

Jorgenson, D., Slesnick, D. (1987), "Aggregate Consumer Behavior and Household Equivalence Scales", *Journal of Business & Economic Statistics*, 5(2), 219-232.

Jorgenson, D., Slesnick, D. (2008), "Consumption and Labor Supply", *Journal of Econometrics*, 147(2), 326-335.

Jorgenson, D., Goettle, R., Ho, M., and Wilcoxen, P. (2013), Double Dividend, Cambridge: MIT Press.

Lau, L. (1982), "A note on the fundamental theorem of exact aggregation", *Economics Letters*, 9(1), 119-126.

Li, Shi, Chuliang Luo, Terry Sicular. "Overview: Income Inequality and Poverty in China, 2002-2007." CIBC Centre for Human Capital and Productivity. CIBC Working Papers, 2011-10. London, ON: Department of Economics, University of Western Ontario (2011).

Meng, X. (2003), "Unemployment, consumption smoothing, and precautionary saving in urban China ", *Journal of Comparative Economics*, 31, 465-485.

Meng, X., Gregory, R., Wang, Y. (2005), “Poverty, inequality, and growth in urban China, 1986-2000", *Journal of Comparative Economics*, 33(4), 710-729.

Meyer, B., Sullivan, J. (2013), “Consumption and Income Inequality and the Great Recession", *The American Economic Review, Papers and Proceedings of the One Hundred Twenty-Fifith Annual Meeting of the American Economic Association*, 103(3), 178-183.

Moschini, G. (1999), "Imposing Local Curvature Conditions in Flexible Demand Systems", *Journal of Business & Economic Statistics*, 17(4), 487-490.

National Bureau of Statistics of China (NBS) (2010) *China Statistical Yearbook 2010*, China Statistics Press, Beijing

Ortega, D., Wang, H., Eales, J. (2009), “Meat demand in China" *China Agricultural Economic Review*, 1(4), 410-419.

Slesnick, D. (1998), "Empirical Approaches to the Measurement of Welfare” *Journal of Economic Literature*, 36, 2108-2165.

-------- (2002),"Prices and Regional Variation in Welfare", *Journal of Urban Economics*, 51, 446–468

Wei, S., Zhang, X. (2011), "The Competitive Saving Motive: Evidence from Rising Sex Ratios and Savings Rates in China", *Journal of Political Economy*, 119(3), 511-564.