Available online at www.sciencedirect.com
SCIENCE DIRECT®

NORTH-HOLLAND

China Economic Review 15 (2004) 457–476

China Economic Review

Investigation of patterns in food-away-from-home expenditure for China

Insik MIN[a], Cheng FANG[b], Qi LI[c,d,*,1]

[a]*Department of Economics, Kyung Hee University, Hoegidong 1, DongdaemoonGu, Seoul, 130-701, Korea*
[b]*Economics and Social Development, FAO—Food and Agriculture Organization of the United Nations, Via delle Terme di Caracalla-00100 Rome, Italy*
[c]*Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA*
[d]*Department of Economics, Tsinghua University, Beijing 100084, People's Republic of China*

Abstract

Using China's urban household-level survey data from 1992 and 1998, we study household food-away-from-home (FAFH) expenditure across the two time periods and across regions. We use a popular parametric linear specification and a newly developed nonparametric estimation method (with mixed categorical and continuous variables) to estimate the FAFH expenditure function. The nonparametric model shows that the income elasticities have increased from 1992 to 1998, while the parametric model suggests the contrary. The goodness-of-fit analysis, a model specification test, and in-sample analysis all suggest that the nonparametric method gives better description of the data than the parametric approach. The nonparametric estimation results also reveal other interesting FAFH consumption patterns which are not detected by the parametric method.

JEL classification: D12; C14
Keywords: Food away from home; China; Consumption

* Corresponding author. Department of Economics, Texas A&M University, College Station, TX 77843-4228, USA. Tel.: +1 979 845 954; fax: +1 979 847 8757.
E-mail address: qi@econmail.tamu.edu (Q. Li).
[1] Li's research is partially supported by the Private Enterprises Research Center, Texas A&M University.

1043-951X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.chieco.2004.06.008

1. Introduction

Until the 1980s, there was relatively little value added in China's food sector and food-away-from-home (FAFH) consumption. Consumers prepared most meals at home with grain, raw vegetables, and meat produced at home, or purchased from state-run food stores or directly from farmers. Along with China's rapid income growth of the last 20 years, Chinese consumers have begun eating more meals in restaurants, cafeterias, and dining halls. The FAFH share of total food expenditures has steadily increased from 5.03% in 1992 to 14.70% at 2000. In 2000 per capita annual FAFH expenditures reached 288 yuan with the total FAFH expenditure 132 billion yuan or US$15.9 billion in urban China. The FAFH expenditure growth in China is expected to continue because of the rising middle class and rapid urbanization, and also due to its relatively low FAFH share compared with the USA (40.3% in 2001, Economic Research Service, USDA), Canada (35.6% in 2001, Statistic Canada), and other developed countries (see Jensen & Yen, 1996).

China's service sector accounts for one-third of its GDP with food processing and food services a major component. With one-fifth of the world's consumers, many observers believe that China will finally emerge as a major force in world food service markets. The increase in FAFH expenditures implies the creation of demand for high value and specialty food products and restaurant services. A better understanding of the factors associated with FAFH expenditures has become increasingly important to understanding changes in the food commodity market, forecasting food demand, anticipating the implications of changes in eating patterns on diet and food imports, and the design of effective marketing programs for both domestic and international restaurant business. The U.S. Department of Agriculture has identified the FAFH issue as one of the key research topics related to China's food and agriculture sectors (Gale, 2002).

Previous studies have investigated Chinese household food expenditures at home (e.g., Chern, 2000; Fan, Crammer, & Wailes, 1994, 1995; Fang & Beghin, 2002; Wang & Chern, 1992; Wu, 1999). But no research has been conducted on China's FAFH. Our research attempts to fill this gap and to provide timely information on the important determinants of FAFH expenditures in China.

Several researchers have investigated FAFH issues with U.S. household survey data. McCraken and Brandt (1987) examine the FAFH expenditure behaviors by type of food facility. They employ a Tobit model in which they decompose the marginal effects on the probability of FAFH consumption and on the actual expenditure level. Yen (1993) studies the patterns of working wives FAFH expenditures with a Box-Cox double hurdle model. Byrne, Capp, and Saha (1996) use the generalized Heckman two-step estimation procedure, which is more flexible than a Tobit model, to study the U.S. FAFH expenditure pattern.

In this paper, we use a popular linear regression model and a newly developed nonparametric approach to examine some important determinants of FAFH expenditures. Nonparametric/semiparametric methods have been successfully applied to estimate various econometric models such as regulation impact analysis on telecommunication industry (Ai & Sappington, 2002), estimation of hedonic price function (Anglin & Gencay, 1993), estimation of consumer demand (Blundell, Duncan, & Pendakur, 1998), and cross-country growth (Liu & Stengos, 1999), to mention only a few. However, when dealing with a mixture of categorical and continuous variables,[2] the conventional nonparametric method is unsatisfactory because the conventional approach is to split the sample into many discrete cells and use the data in each discrete cell to estimate a nonparametric regression function using the remaining continuous variables. This sample splitting method often results in huge finite sample efficiency losses and it can even become infeasible when the number of discrete cells is large relative to the sample sizes. The recently developed cross-validation-based techniques on smoothing both the discrete and the continuous variables (Hall, Racine, & Li, in press; Li & Racine, 2003; Racine & Li, 2004) do not suffer the above-mentioned problem. Moreover, when there exist some irrelevant explanatory variables (variables that are in fact independent of the dependent variables), Hall et al. (in press) have shown that the cross-validation method has the amazing ability of *automatically* removing irrelevant (explanatory) covariates, a property not shared by any of the existing estimation methods. Indeed, we have found that, for empirical applications with economic data, "irrelevant" covariates are surprisingly common. For example, Li and Racine (in press) show that, via many types of empirical data (e.g., U.S. patent data, crop yield data, female labor market participation data, marketing data, medical treatment data, etc.), smoothing the discrete variables often lead to much smaller out-of-sample prediction mean square errors than the conventional sample-splitting nonparametric method and the commonly used parametric methods. The superior performance of the nonparametric cross-validation methods in the above applications is due to two facts: (i) it automatically removes 'irrelevant' covariates and (ii) it detects nonlinearities for relevant covariates. As we will show in this paper with China's Urban Household Survey data, this newly proposed nonparametric estimation method removes an irrelevant regressor, detects nonlinearities in other variables, and gives a much better estimation result than a parametric linear model. It reveals how FAFH consumption changes for different income levels, for different demographic regions, and over time, which may not be easily detected by commonly used parametric specifications.

This paper is organized as follows. In Section 2, we first briefly present a framework for modeling household FAFH, and discuss the data sources and data descriptions. Then we discuss the parametric and the nonparametric methods in modeling FAFH expenditure. Section 3 reports the estimation results based on both a parametric linear model and the nonparametric model discussed in Section 2. Conclusions are given in Section 4.

2. Data description and econometric models

In this section, we first review an empirical modeling approach for deriving a FAFH expenditure function by Yen (1993). We then provide an overview of the household data used in the present analysis. Finally, we present the parametric specification and the nonparametric estimation method proposed by Hall et al. (in press), Li and Racine (2003),

[2] In a regression model, the discrete or continuous variables refer to the explanatory variables. For regression models related to FAFH, the explanatory discrete variables usually include: gender dummy, location dummy, family size, job title of the household head, among others.

and Racine and Li (2004) for estimating FAFH equation using China's Urban Household Survey data.

2.1. Empirical model for FAFH

In principle, FAFH consumption function can be derived by maximizing household utility function subject to some constraints. Let $U(z_1, \ldots, z_m)$ be a typical household utility function, where $(z_1, \ldots, z_m)$ are m basic activity choices that the household faces. Let

$$z_i = z_i(x_i, t_{i1}, \ldots, t_{iL})$$

be the production function of producing z_i, x_i denotes the consumer good used to produce z_i, and t_{ij} is the time spent by the household member j in producing z_i. Let h_j be the time input by household member j in market product. The time constraint is

$$T_j = h_j + \sum_{i=1}^{m} t_{ij}, \qquad (j = 1, \ldots, L)$$

Finally, let w_j be the wage rate of household member j, and p_i the price of x_i. The budget constraint is

$$\sum_{j=1}^{L} w_j h_j + v = \sum_{i=1}^{m} p_i x_i$$

where v is the nonlabor income. Maximizing the utility function, subject to the production, time and budget constraints, leads to the demand function

$$x_i = f_i(p_1, \ldots, p_m, w_1, \ldots, w_L, v).$$

Consider a family with one- or two-earner husband-and-wife families. Yen (1993) has derived the following FAFH expenditure function (subscripts 1 and 2 indicate husband and wife)

$$p_i x_i = g_i(h_2, w_2, v', D)$$

where $v' = w_1 h_1 + v$ is the household's income excluding wife's wage earning, D is a vector of socidemographic variables. For China's FAFH data set we have, wife's wage is not unavailable. Instead, we have the total family income (husband and wife's income combined). Therefore, we replace (w_2, v') by $w = w_1 + w_2$ in the FAFH expenditure function. Also, we do not have wife's working hour information. So the model we will estimate is based on

$$p_i x_i = g_i(w, D).$$

The socidemographic variable D includes age, education level, and job title of the household head, family size, city-size dummy variable, etc. We will give detailed description of the variables we choose to use to estimate the FAFH equation in the next subsection. The data we will use are similar to the U.S. data used by Yen (1993) with the main difference being that our income and consumption are annual data, while Yen used weekly data.

In order to estimate the FAFH expenditure function empirically, a common practice is to assume a particular functional form for $g_i(\cdot)$, say if $g_i(\cdot)$ is a linear model, then one can simply apply the least squares method to estimate the model. If the dependent variable contains a substantial portion of zeros, a Tobit-type or a more general (Box-Cox) double hurdle model can be adopted. However, a linear model is likely to be misspecified in practice. For example, a linear model implies a constant marginal income effect which can lead to misleading predictions (as we will show later in this paper). To avoid functional form misspecification, in addition to estimating a linear FAFH regression model, we also estimate a nonparametric FAFH regression model. We will then compare the similarities and differences of the estimation results obtained from the two approaches. We will show that the nonparametric estimation results reveal some interesting FAFH consumption pattern which are not detected by the parametric method.

2.2. Data description

The data are from urban household surveys conducted by the State Statistical Bureau of the People's Republic of China. The household surveys are carried out by local agencies. The families selected in the surveys are drawn from a very large population frame. Sampled households maintain a daily diary and transaction books that record all expenses and consumption in the households for a given year (Han, Wailes, & Cramer, 1995). This data set is unique because it encompasses an annual FAFH consumption for the surveyed households. We use this comprehensive survey data across all provinces in 1992 and 1998 to examine China's FAFH consumption pattern.

These data were used by Fang and Beghin (2002) to investigate the urban demand for edible oils and fats in China. The survey is administered directly by the National Statistics Bureau through its provincial and local survey network. The survey covers 30 provinces, 146 sample cities, and more than 80 counties. The sample is drawn based on several stratifications. The first step is to determine the sample size in each of the six large regions with the sample size being proportional to the region's population. Then, within each region, all the provincial capitals are chosen to represent large cities while mid-size cities and county towns are randomly selected. Next, within the selected cities and towns, the neighborhood committees and finally households are chosen by a further random selection. The data to which we have access include all sampled households in 28 selected cities, one representative city from each province (two provinces are dropped due to missing data). The dependent variable is the household per capita FAFH expenditure (in 1992 yuan). The explanatory variables include: (i) Household per capita income (in 1992 yuan); (ii) household size; (iii) education level of the household head with seven categories: (1) below elementary school, (2) elementary school, (3) middle school, (4) high school, (5) middle-level specialized training, (6) 2-year college, (7) bachelor's degree or above; (iv) age of the household head; (v) a 0–1 dummy variables indicating whether the household lives in a large city; (vi) a gender dummy of the household head; and finally (vii) a job title (of the household head) variable which takes eight different values giving job classification of the household head, these eight job titles are: (1) technical personnel, (2) chief in government agencies and social organizations, (3) clerks and staffs, (4) workers in wholesale and retail sectors, (5) social services, (6) workers in farming, forestry, animal husbandry and fishery, (7) workers in manufacturing, transport, etc., and (8) others who are hard to classify.

The total sample sizes we use in this study are 3459 for 1992 and 3359 for 1998. The ratio of average expenditure between the samples used in our study and all samples surveyed by National Statistics Bureau is close to one and therefore there is no serious sampling bias. Summary statistics of the above variables are presented in Table 1.

Following Fang and Beghin (2002), the FAFH expenditure and income are deflated by the provincial urban consumer price index (UCPI). Each UCPI is set equal to one in 1980, assuming the same cost of living standard for all regions in that year. This assumption is reasonable since prices were under strict control by the state government and the cost of living was basically uniform in 1980 for different regions (Kanbur & Zhang, 1999). After 1980, purchasing power parity holds approximately among different regions by using the UCPI deflator.

2.3. The parametric model

Previous studies with U.S. data usually involve estimation of some censored regression models such as a Tobit model (McCracken & Brandt, 1987), or a two-step estimation procedure (Byrne et al., 1996) because zero FAFH expenditure occurs from 15% to 45% for U.S. data. However, for our Chinese survey data, the zero FAFH expenditure is less than 3%, because the FAFH expenditure in China is an annual amount, while in the United States it is the amount spent for a 2-week period. The estimation results based on a Tobit model and from a simple linear model (with the least squares method) are almost identical. Therefore, there is no need to use a censored regression model with China's data due to the annual nature of the diary data, and we will only report the least squares estimation results.

Since the nonparametric estimation method discussed below treats the continuous and the discrete variables differently, we need to distinguish these two types of variables. We use x to denote the continuous variable and z to denote the discrete variable. The

Table 1
Summary statistics

	1992				1998			
	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
y (FAFH)	99.17	124.84	0	1815.3	158.69	203.20	0	2381.4
x_1 (real income)	2444.3	918.7	556.1	5999.2	3406.5	1603.1	802.4	10 907.9
x_2 (household size)	3.33	0.978	1	11	3.13	0.827	1	8
x_3 (education level)	4.01	1.48	1	7	4.19	1.42	1	7
x_4 (age)	45.80	11.78	19	75	47.18	11.29	22	75
z_1 (gender)	0.697	0.458	0	1	0.612	0.487	0	1
z_2 (city size)	0.299	0.457	0	1	0.309	0.462	0	1
z_3 (job title)	39.20	2.91	1	8	50.71	1.57	1	8

parametric specification we use in this paper is a standard linear regression model with some socioeconomic and demographic variables,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 z_{1i} + \beta_6 z_{2i} + \sum_{j=2}^{8} \alpha_j z_{3j,i} + u_i, \tag{2.1}$$

where x_1, x_2, x_3, and x_4 are per capita annual income, household size, education level of the household head, and age of the household head, respectively. z_1 is the gender dummy (1 for male). z_2 is a dummy variable that equals 1 if the household lives in a large city, and equals 0 otherwise. z_{3j} (j=2, ..., 8) covers 7 job title dummy variables, $z_{3j,i}$ equals 1 if individual i has the jth job title, and zero otherwise. The error u_i is assumed to have a zero mean ($E(u_i|x_i, z_i)$=0) and a finite variance.

2.4. The nonparametric model

The nonparametric regression model we consider is

$$y_i = g(x_i, z_i) + u_i \tag{2.2}$$

where x_i=(x_{1i}, x_{2i}, x_{3i}, x_{4i}), and z_i=(z_{1i}, z_{2i}, z_{3i}) are the continuous and the discrete variables, respectively. $g(x, z)$=$E(y_i|x_i, z_i)$ is the conditional mean function, and the functional form of $g(\cdot)$ is not specified so that we allow a flexible (nonlinear) functional form. $E(u_i|x_i, z_i)$=0 and var$(u_i|x_i)$=$\sigma^2(x_i)$. The functional form $\sigma^2(\cdot)$ is not specified. We use the kernel method to estimate the unknown function $g(\cdot)$.[3] Strictly speaking, age, education level, and the household size are (ordered) discrete variables, by treating them as ordered discrete variables or as continuous variables lead to very similar estimation results. For a more detailed treatment of nonparametric regression function estimation with continuous variables, see Pagan and Ullah (1999), and for the case of mixed discrete and continuous variables, see Li and Racine (in press) and Racine and Li (2004).

Let $k(\cdot)$ denote the univariate kernel function for a continuous variable, and $K((x_i-x)/h)=\Pi_{s=1}^{4} k((x_{si}-x_s)/h_s)$ is the product kernel for the continuous variables, where x_{si} and x_s are the sth components of x_i and x, respectively. h_s is the smoothing parameter associated with x_s(s=1, 2, 3, 4). In the empirical application, we choose a standard normal kernel function: $k(v)=\mathrm{e}^{-v^2/2}/\sqrt{2\pi}$ (see Pagan & Ullah, 1999 for a more detailed discussions on kernel estimation of a nonparametric regression function). For the discrete variable z, we use the following kernel function as suggested by Racine and Li (2004).

$$l(z_{si}, z_s, \lambda_s) = \begin{cases} 1, & \text{if } z_{si} = z_s, \\ \lambda_s, & \text{otherwise,} \end{cases} \tag{2.3}$$

where z_{si} and z_s are the sth components of z_i and z, respectively (s=1, 2, 3).

The range of the smoothing parameter λ_s is [0, 1]. If λ_s=0, then $l(z_{si}, z_s, 0)$ becomes an indicator function, and if λ_s takes the upper bound value of 1, then $l(z_{si}, z_s, 1)\equiv 1$ is a constant for all values of (z_{si}, z_s).

[3] One can also use other nonparametric methods such as the series method to estimate $g(\cdot)$, see Ai and Chen (2003) for a general treatment on efficient estimation with nonparametric series method.

The product kernel for the discrete variable is

$$L(z_i, z, \lambda) = \prod_{s=1}^{3} l(z_{si}, z_s, \lambda_s). \tag{2.4}$$

We use a Nadaraya–Watson type kernel to estimate $g(x, z)$

$$\hat{g}(x,z) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x_i - x}{h}\right) L(z_i, z, \lambda)}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) L(z_i, z, \lambda)}. \tag{2.5}$$

Racine and Li (2004) have shown that $\hat{g}(x, z)$ consistently estimate $g(x, z)$. They also establish the asymptotic normal distribution of $\hat{g}(x, z)$.

Note that if λ_s=0 for all s=1, 2, 3, then $L(z_i, z, 0)$ becomes an indicator function, which takes value one if z_i=z, and zero otherwise. In this case, $\hat{g}(x, z)$ defined by Eq. (2.5) becomes the conventional frequency estimators of $g(x, z)$, where one splits the sample into different discrete cells in estimating $g(x, z)$.

It is well known that the selection of smoothing parameters is of crucial importance in nonparametric kernel estimations. Hall et al. (in press), Li and Racine (2003), and Racine and Li (2004) show that the least squares cross-validation method performs very well in simulations and with a number of real data sets. Therefore, we will use the cross-validation method to select the smoothing parameters. That is, we choose (h, λ)=$(h_1, h_2, h_3, h_4, \lambda_1, \lambda_2, \lambda_3)$ by minimizing the following cross-validation function:

$$CV(h, \lambda) = \sum_{i=1}^{n} [y_i - \hat{g}_{-i}(x_i, z_i)]^2, \tag{2.6}$$

where $\hat{g}_{-i}(x_i, z_i)$=$\sum_{j\neq i}^{n} y_j K_{ij} L_{ij} / \sum_{j\neq i}^{n} K_{ij} L_{ij}$ is the leave-one-out kernel estimator of $g(x_i, z_i)$, K_{ij}=$K((x_i - x_j)/h)$ and L_{ij}=$L(z_i, z_j, \lambda)$. From the results given in Racine and Li (2004), we know that, asymptotically, CV(h, λ)=$E\{[\hat{g}(X, Z) - g(X, Z)]^2\}$+terms unrelated to (h, λ). Therefore, the cross-validation selected smoothing parameters are asymptotically optimal in the sense that they minimize the asymptotic estimation mean square errors.

Note that as we mentioned earlier, if for some $t \in \{1, 2, 3\}$, that z_{si} is in fact an irrelevant variable, say z_{si} is independent of y_i, then the cross-validation method will choose λ_s=1, which leads to $l(z_{si}, z_s, 1) \equiv 1$ for all values of (z_{si}, z_s), the corresponding variable z_s is completely smoothed out, as the estimated conditional mean function $\hat{g}(x, z)$ becomes independent of z_s. In this case, our nonparametric kernel estimator is much more efficient than the conventional nonparametric method which uses λ_s=0 (sample splitting).

Also note that if h_s=∞, then $k((x_{si} - x_s)/h_s)$=$k(0)$ becomes a constant. In this case, the continuous variable x_s is automatically removed from the regression model since the estimated conditional mean function $\hat{g}(x, z)$ becomes independent of x_s. Hall et al. (in press) have shown that the least square cross-validation method can (asymptotically) automatically remove irrelevant variables by assigning λ_s=1 and choosing a sufficiently large value for h_s so that the irrelevant variables z_s and x_s are smoothed out (removed) from the regression model.

The asymptotic distribution of $\hat{g}(x_i, z_i)$ is derived in Racine and Li (2004) as

$$\sqrt{nh_1h_2h_3h_4}[\hat{g}(x,z) - g(x,z)] \rightarrow N(0, V(x,z)) \text{ in distribution}, \tag{2.7}$$

where $V(x, z)=\sigma^2(x, z)[\int K(v)^2 dv]/f(x, z)$, $\sigma^2(x, z)=E(u_i^2|x_i=x, z_i=z)$. A consistent estimator of $V(x, z)$ is given by

$$\hat{V}(x,z) = \hat{\sigma}^2(x,z)\left[\int K(v)^2 dv\right]/\hat{f}(x,z), \tag{2.8}$$

where $\hat{\sigma}^2(x,z) = \sum_{j=1}^n \hat{u}_j^2 K\left(\frac{x_j-x}{h}\right)L\left(z_j,z,\lambda\right)/\sum_{j=1}^n K\left(\frac{x_j-x}{h}\right)L\left(z_j,z,\lambda\right)$. The 95% confidence bands for $\hat{g}(x, z)$ are given by $\hat{g}(x,z)\pm 1.96\sqrt{\hat{V}(x,z)/(nh_1h_2h_3h_4)}$.

3. Empirical results

3.1. Parametric estimation results

Table 2 reports the parametric model estimation results. First, we observe that income is the most significant variable in both 1992 and 1998. Also, their numerical values are fairly close to each other for the two sample periods. The household's income elasticity is computed by $\beta_1(x_{1i}/y_i)$. The mean income elasticities on FAFH are 1.155 and 1.195 for 1992 and 1998, respectively. The estimated elasticities are significantly larger than the findings using U.S. data. Byrne et al. (1996), McCracken and Brandt (1987), and Yen (1993) find that the elasticities are between 0.20 to 0.36 using U.S. FAFH consumption data. FAFH consumption of Chinese households is more income elastic compared with U.S. households. About 40% of the surveyed households in China have elasticities greater than one which implies that FAFH is a luxury good for many households in China (see Table 6). While the elasticities from U.S. studies are much smaller than one.

The coefficient of household size is positive but is not significant. This result implies that the per capita FAFH expenditure does not depend on the size of the family in both 1992 and 1998. However, as we will see later, the nonparametric method reveals a

Table 2
Parametric model estimation results

	1992			1998		
	Coefficient	S.D.	t-ratio	Coefficient	S.D.	t-ratio
β_0 (constant)	29.0	22.3	1.30	88.3	31.8	2.78
β_1 (real income)	0.041	0.0036	11.5	0.044	0.0032	13.8
β_2 (household size)	3.02	2.72	1.11	0.789	3.97	0.20
β_3 (education level)	0.75	2.15	0.35	2.86	3.01	0.95
β_4 (age)	−1.05	0.25	−4.19	−1.97	0.394	−5.01
β_5 (gender)	−11.0	4.86	−2.26	−10.1	7.04	−1.44
β_6 (city)	31.8	4.66	6.82	53.2	8.97	5.92
α_2 (job title 2)	7.93	7.77	1.02	−2.41	15.2	−0.16
$\alpha_3$3 (job title 3)	5.81	6.63	0.88	−7.98	12.1	−0.66
α_4 (job title 4)	3.85	12.0	0.32	−23.5	17.8	−1.32
α_5 (job title 5)	−10.6	11.2	−0.95	2.31	21.0	0.11
α_6 (job title 6)	−5.91	10.7	−0.55	−19.9	15.2	−1.31
α_7 (job title 7)	0.99	7.32	0.13	−20.8	12.9	−1.62
α_8 (job title 8)	−31.3	12.6	−2.49	−53.1	31.8	−1.69

different pattern where it shows that the household size affects FAFH consumption nonlinearly.

The coefficient of the education level is positive but not significant. This result is similar to U.S. studies. Most U.S. studies indicate that the education level does not significantly affect FAFH expenditure. The (household head) age has a negative and significant effect on FAFH. The magnitude of the marginal age effects has conspicuously risen from 1.05 in 1992 to 1.97 in 1998. This reflects the *conjecture* that the older generations are more frugal and spend less on FAFH, a habit formed during the past years when the living standards were low in China, which could be difficult to change. In 1992, the Gender coefficient is negative and significant at the 5% level, but not at the 1% level. In 1998, the Gender coefficient is not significant even at the 10% level.

The city dummy variable has a significant positive effect on FAFH for both 1992 and 1998, indicating that households in large cities spend more on FAFH than those in small cities. One possible explanation for the positive (large city) effect is that large cities have more variety of food offerings and therefore FAFH consumption is more attractive to customers. The coefficient of the city dummy variable has increased from 31.81 in 1992 to 53.15 in 1998. This may suggest that the difference in FAFH consumption between large city and middle–small city is widened in 1998. However, as we will show using the robust nonparametric estimation method, the linear model is misspecified and in fact the opposite is true, i.e., the consumption patterns between large city and middle–small city become closer to each other in 1998 than in 1992. This may be due to the fact that middle–small start to have more variety of food offerings in 1998 than in 1992, and therefore, the FAFH consumption difference between large and middle–small cities becomes smaller in 1998.

Finally, the job title variables are mostly insignificant except for the job title 8 in 1992. An F test shows that the job title variables are jointly significant in 1992 at the 10% level, while they are not significantly different from zero at any conventional level for 1998.

3.2. Nonparametric estimation results

We present nonparametric estimation results in this subsection. As we discussed in Section 2, the data-driven cross-validation method may select λ_s=1 for some discrete variable z_s, and may choose a sufficiently large value for h_s for some x_s in which case it will imply that the corresponding z_s and x_s are irrelevant regressors, and these irrelevant variables are automatically removed from the regression model [since the nonparametric estimate $\hat{g}(x, z)$ becomes unrelated to these irrelevant variables]. For the continuous variable x_s, we write $h_s=c_s x_{s,\mathrm{sd}} n^{-1/8}$, where $x_{s,\mathrm{sd}}$ is the sample standard deviation of $\{x_{si}\}_{i=1}^{n}$. The cross-validation method selects c_s. If c_s is very large, the corresponding x_s variable will be automatically removed (smoothed out). Table 3 reports our cross-validation selected smoothing parameters λ_s (s=1,2,3), and c_s (s=1,2,3,4).

From Table 3 we see that the smoothing parameter for the job title variable takes the upper bound value 1 with both the 1992 and the 1998 data. This means that the job title is an irrelevant variable and it is automatically removed (completely smoothed out) from the nonparametric regression model. In contrast, the smoothing parameters for the city size dummy are quite small, its value is quite close to the lower bound 0 in 1992. This indicates that among the three discrete variables, city size variable has the most significant effect on

Table 3
Cross-validation values of λ and $h(h_s = c_s x_{s,\mathrm{sd}} n^{-1/7})$

Year	λ_1	λ_2	λ_3	c_1	c_2	c_3	c_4
1992	0.186	0.0004	1.000	1.098	1.544	3.21×10^6	0.850
1998	0.173	0.089	1.000	0.989	1.716	1.737	3.347

FAFH consumption. The fact that λ_2 (the smoothing parameter related to the city size variable) is close to 0 (in 1992) suggests that FAFH consumption patterns are quite different for large and middle–small city households in 1992, while a relatively larger value of λ_2 in 1998 indicates that this difference is reduced in 1998 (λ_2=0.0004 in 1992 and 0.089 in 1998). This is possibly due to the fact that in 1998 the variety of food offerings had increased (relative to 1992) faster for many middle–small cities compared with large cities.

For the continuous variables, education is smoothed out in 1992, i.e., household head education level does not affect the nonparametric estimation of the FAFH consumption in 1992.

The main advantage of the nonparametric approach is that it allows for a flexible regression functional form. However, it is difficult to present nonparametric estimation results with high dimensional data such as in our case. To overcome this problem, we choose to present the estimated FAFH expenditure as a function of a continuous variable and the city dummy variable, taking other continuous variables at their sample mean values, and the gender dummy variable at its mode value (male). Note that the job title variable and the education variable in 1992 are automatically smoothed out and therefore they are effectively removed from the explanatory variable set. We present results for large and middle–small city separately because the estimation results suggest that FAFH consumption behavior for households in large cities are different from those in middle–small cities (since λ_2 is close to zero).

We first consider the impact of income on FAFH consumption. We graph $\hat{g}_i(x_1, z_2) \equiv \hat{g}(x_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, z_1=1, z_2)$ as a function of x_1 (income) and z_2 (city size), this gives two curves, one for z_2=1 (large city), and one for z_2=0 (middle–small city), where $\bar{x}_s$ is the sample mean of $\{x_{si}\}_{i=1}^{n}$ for s=2, 3, 4, and z_1=1 (male) is the mode for the gender dummy z_1 (if it is 1992 data, x_3 is removed).

Figs. 1 and 2 give the estimated $\hat{g}_i(x_1, z_2)$ curves for 1992 and 1998, respectively, along with their 95% confidence bands. The estimated FAFH consumption curves show clearly some nonlinear patterns in income. For the 1992 curves, the large city curve peaks at income 3000 yuan (per capita) while the middle–small city curve peaks at income close to 4000 yuan. For income greater than 4000 yuan, the large city curve is relatively flat, while the middle–small city curve decreases. The decline in FAFH consumption for income higher than 4000 yuan may seem puzzling. One possible explanation is that household heads in high-income families are more likely to be business leaders or high rank government officials, and those people often have free meals (meals covered by business and government funds). Since FAFH only includes expenditures from one's own pocket and therefore free meals are not part of it, as a consequence, high-income households may spend less on FAFH consumption with their own funds. Another explanation is that high-

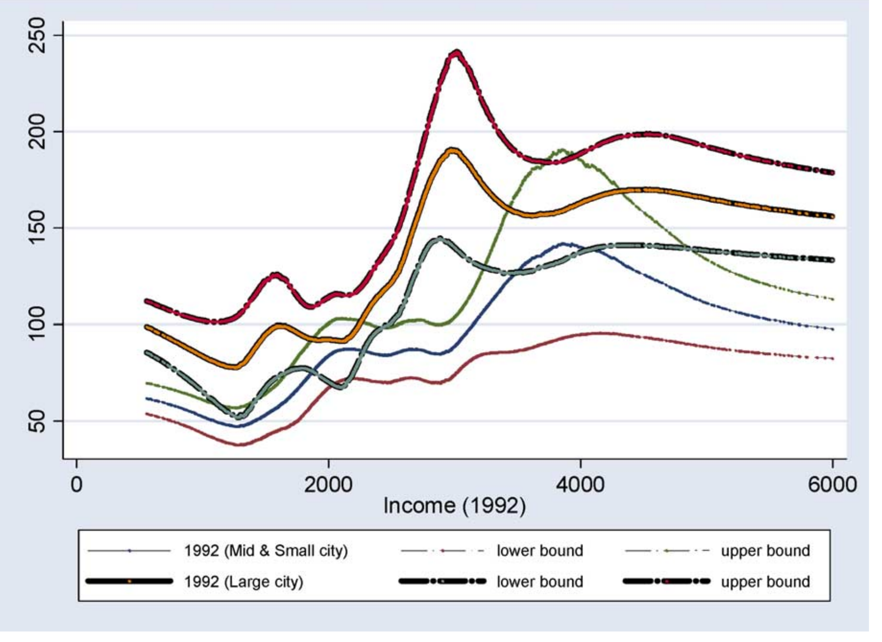

Fig. 1. Predicted FAFH expenditure versus income (1992).

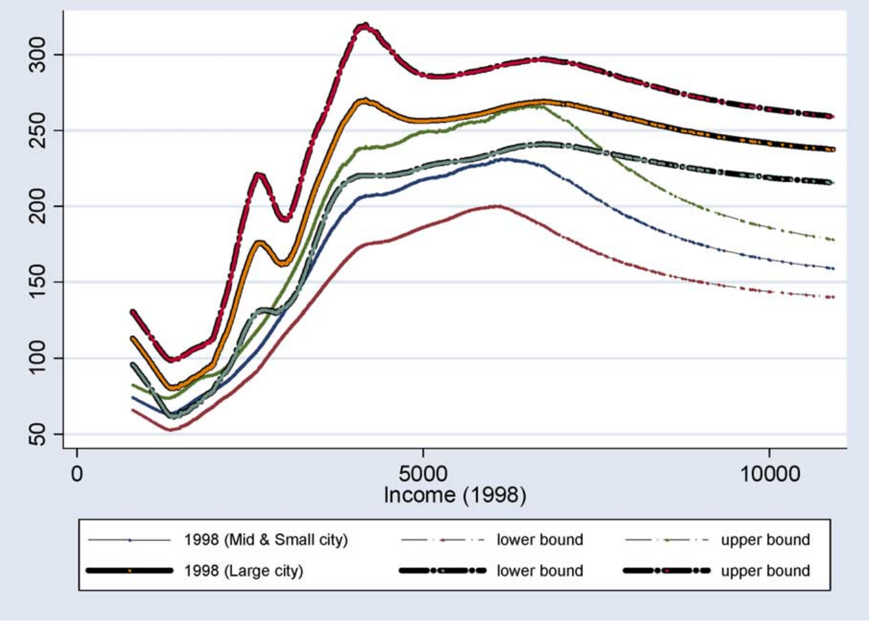

Fig. 2. Predicted FAFH expenditure versus income (1998).

income families are likely to hire part-time (flexible hour) housekeepers to help on household work including cooking dinners so that the need to eat outside is reduced and consequently, these high-income families may spend less on FAFH consumption. Fig. 2 shows that the two 1998 curves have similar shapes; this reinforces our earlier argument that different FAFH patterns due to city size are reduced in 1998. The FAFH consumption increases rapidly for income between 1500 and 4000 yuan. The large city curve keeps relatively flat after 4000 yuan, while the middle–small city curve decreases after 6500 yuan, a phenomena that is already observed with the 1992 data.

Both 1992 and 1998 estimated FAFH expenditure curves show that the FAFH consumption does not increase further for very high-income households. This is different from the parametric model estimation result, a linear model would predict higher FAFH consumption for higher-income households. To check which estimation results give better description of the data, we compute the goodness-of-fit of R^2 for both the parametric model and the nonparametric model. We also compute average FAFH consumption for different income households, and the average and the median income elasticities. The results are given in Tables 4–6.

From Table 4, we see that the nonparametric model has a larger R^2 than that of the parametric model, which means that the nonparametric method fits the data better than the parametric models. To decide whether the differences are statistically significant, we apply a recently developed model specification test by Hsiao, Li, and Racine (2003) to test the null hypothesis that the parametric linear model is in fact a correct specification. The alternative hypothesis is that the parametric model is misspecified, but the nonparametric model is the correct specification. Roughly speaking, the test is based on the difference between a parametric fit and a nonparametric fit using the cross-validation method to select smoothing parameters. Specifically, their test statistic is given by

$$\hat{J}_n = \frac{1}{n\sqrt{h_1 h_2 h_3 h_4}} \sum_{i=1}^{n} \sum_{j=1, j\neq 1}^{n} \hat{u}_i \hat{u}_j K\left(\frac{x_i - x_j}{h}\right) L(z_i, z_j, \lambda)/\hat{\sigma}_0, \tag{2.9}$$

where $\hat{u}_i$ is the least squares residual from the parametric model (2.1), and $\hat{\sigma}_0^2 = 2(n^2 h_1 h_2 h_3 h_4)^{-1} \sum_{j=1}^{n} \sum_{j=1, j\neq 1}^{n} \hat{u}_i^2 \hat{u}_j^2 K^2\left(\frac{x_j - x_i}{h}\right) L^2(z_i, z_j, \lambda)$.

The $\hat{J}_n$ test defined in (2.9) has an asymptotic standard normal distribution under the null. However, the simulation results show that the asymptotic test is undersized, while the wild bootstrap procedure leads to good estimated sizes. Therefore, we use the bootstrap method to obtain the critical values for the test. The test statistic yield values of 1.68 and 3.79 for 1992 and 1998 data, respectively, while the 5% critical values obtained from 1000 bootstrap procedure (see Hsiao et al., 2003 for details) are 1.41 and 1.88, respectively. Hence, the test rejects the parametric linear specification at the 5% level using both 1992

Table 4
Goodness-of-fit R^2

	Parametric model		Nonparametric model	
Sample period	1992	1998	1992	1998
R^2	.128	.170	.382	.348

Table 5
Average FAFH expenditure by income category

	1992				1998		
Income	Data average	OLS average	NP average	Income	Data average	OLS average	NP average
Below 3K	82.78	82.12	83.26	below 4K	119.40	120.1	121.9
3K–5K	154.9	152.0	148.0	4K–9K	257.0	250.0	254.0
Above 5K	143.9	221.7	144.9	Above 9K	205.2	441.1	232.9

(since 1.68>1.40) and 1998 (since 3.79>1.88) data.[4] Thus, we conclude that the nonparametric method gives significantly better fit than the parametric model.

Table 5 gives the average FAFH expenditure for different income intervals, along with the predicted average values using the linear model and the nonparametric model. It shows that high-income households indeed spent less on FAFH than those households with slightly lower income. In 1992, households with income more than 5000 yuan spent about 11 yuan less on FAFH than those with income between 3000 and 5000 yuan. In 1998, households with income above 9000 yuan spent 52 yuan less on FAFH than those with income between 4000 and 9000 yuan. We observe that while for the low- and middle-income levels, both the parametric and the nonparametric models predict well for the average FAFH expenditure. However, for the high-income level, the parametric model gives misleading prediction results, while the nonparametric method performs much better in this case. The reason for this is that the true relationship between FAFH expenditure and income is nonlinear, the inflexibility of a linear model leads to its inferior prediction performance.

Table 6 reports the estimated mean and median income elasticities. For the nonparametric model, the elasticity η_i for house i is computed via $\eta_i = \frac{\partial \hat{g}(x_i, z_i)}{\partial x_{1i}} \frac{x_{1i}}{y_i}$, where x_{1i} is the income of the ith household.[5] For the linear model, $\eta_i = \beta_1 (x_{1i}/y_i)$. Table 6 reveals some interesting phenomena. Comparing the elasticities of 1992 and 1998, the nonparametric estimation results show that both the mean and median income elasticities have increased for both large city and middle–small city households. In contrast, the parametric model only shows that the mean elasticity for middle–small city has increased, while the mean elasticity for large city, and the median elasticity for both large and middle–small city are reduced from 1992 to 1998. The conflicting estimation results are due to the misspecification of the linear model. As we discussed earlier, the linear model overestimates FAFH expenditure for high-income households due to its imposing an incorrect linear income (trend) component on FAFH consumption (see Table 5 and the discussions there). The misspecified linear model also overestimates the elasticities for high-income families, leading to the erroneous prediction that the median elasticity has decreased from 1992 to 1998.

Also from Table 6, we observe that the nonparametric estimation result shows that the (mean and median) elasticity between mid–small city and large city becomes

[4] We also applied series-based nonparametric test proposed by Hong and White (1995) and obtain the same conclusion.

[5] Note that we have used a standard normal kernel $k((x_{1j}-x_{1i})/h_1)=\exp(-(x_{1j}-x_{1i})^2/(2h_1^2))/\sqrt{2\pi}$ so that $\hat{g}(x_i, z_i)$ is differentiable with respect to x_{1i}.

Table 6
Mean and median income elasticities

	Parametric				Nonparametric			
	Large city		Mid–small city		Large city		Mid–small city	
	1992	1998	1992	1998	1992	1998	1992	1998
Mean	0.878	0.848	1.274	1.351	0.626	0.826	0.900	0.947
Median	0.848	0.832	1.074	0.994	0.751	0.851	0.938	0.960
% of η_i>1	18.5	13.9	69.6	48.8	35.0	39.9	45.9	45.8

Note: η_i denotes income elasticity of household i.

closer to each other in 1998 than in 1992, reinforcing our earlier claim that the difference in FAFH consumption between mid–small city and large city households becomes smaller in 1998 than in 1992, while the parametric linear model fails to detect this trend.

As mentioned earlier, U.S. and other developed countries have income elasticity around 0.25 to 0.40. In contrast, China has a larger income elasticity of around 1, and the robust nonparametric estimation results show that the income elasticity is still on the rise. Table 6 also reports the percentage of households with income elasticity greater than 1. As can be seen from Table 6 for more than 40% of the surveyed households in China, FAFH is a luxury good and it is expected that as China continues to improve its living standard, people will spend significantly more on FAFH, and thus it is likely that the income elasticity will continue to rise in the near future before it decreases to a stabilized level.

To examine the effect of family size (x_2) on FAFH consumption, we graph the 1992 and 1998 functions $g_s(x_2, z_2)$=$g(\bar{x}_1, x_2, \bar{x}_3, z_1$=1, $z_2)$ in Figs. 3 and 4, respectively. From Fig. 3 we observe that in 1992, family size does not seem to affect FAFH much for large city households, while for middle–small city FAFH is the highest for a single-person family, then FAFH is basically flat for family size greater or equal to two. Fig. 4 shows that the 1998 functions have quite different patterns compared with the 1992 functions. For both middle–small and large cities, FAFH first increases with family size and reaches the maximum values for three-people family, then it exhibits a downward trend as family size increases further. There are 52% and 63% of families with size less or equal to three in 1992 and 1998, respectively. These changes are partly due to the encouragement of the one-child policy in China. Given the fact that the percentage of three-people family is on the rise in China, it is expected that the FAFH consumption should increase rapidly due to this family structure change, in addition to the impact of future real income increases.

As we discussed earlier, the education variable has no explanatory power to FAFH in 1992 (as it is automatically removed) and therefore we do not plot the FAFH–education graph for 1992. For 1998, Fig. 5 plots $g_E(x_3, z_2)$=$g(\bar{x}_1, \bar{x}_2, x_3, \bar{x}_4, z_1$=1, $z_2)$ using 1998 data. Fig. 5 shows that higher education leads to more FAFH consumption. The large city curve shows a more prominent upward trend than the middle–small city curve.

Finally, Figs. 6 and 7 graphs $g_A(x_4, z_2)$=$g(\bar{x}_1, \bar{x}_2, \bar{x}_3, x_4, z_1$=1, $z_2)$ using 1992 and 1998 data, respectively. Except for the 1992 curve with age less than 30, all curves show a

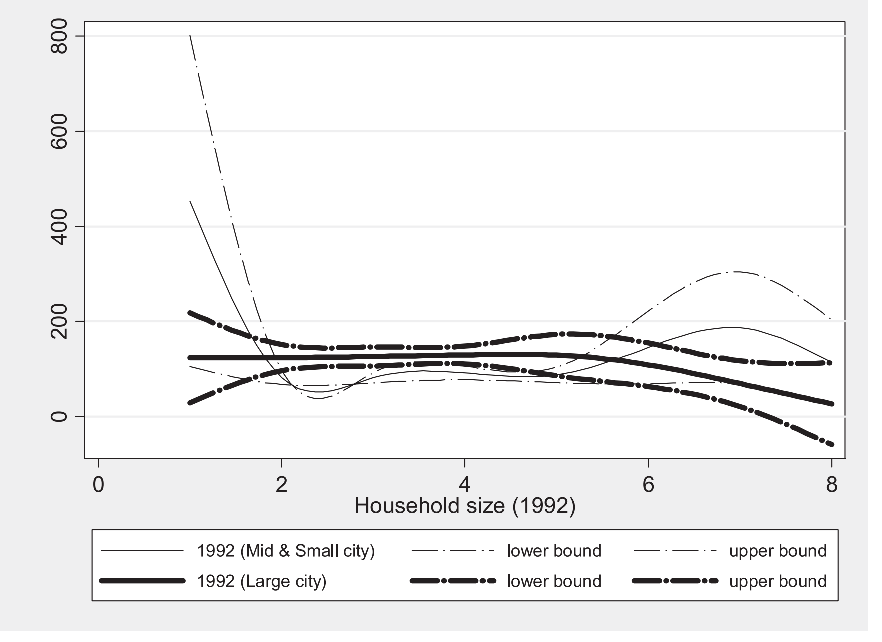

Fig. 3. Predicted FAFH expenditure versus household size (1992).

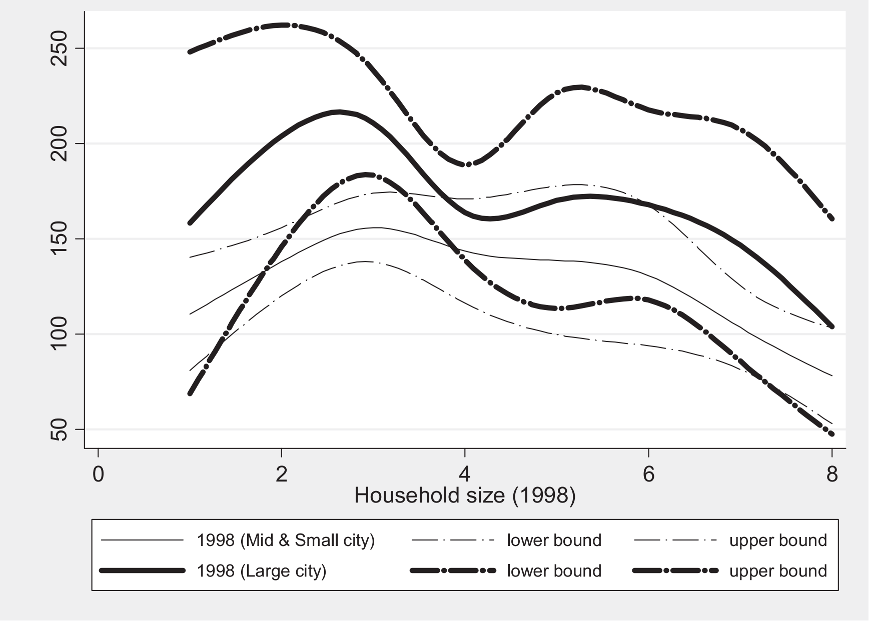

Fig. 4. Predicted FAFH expenditure versus household size (1998).

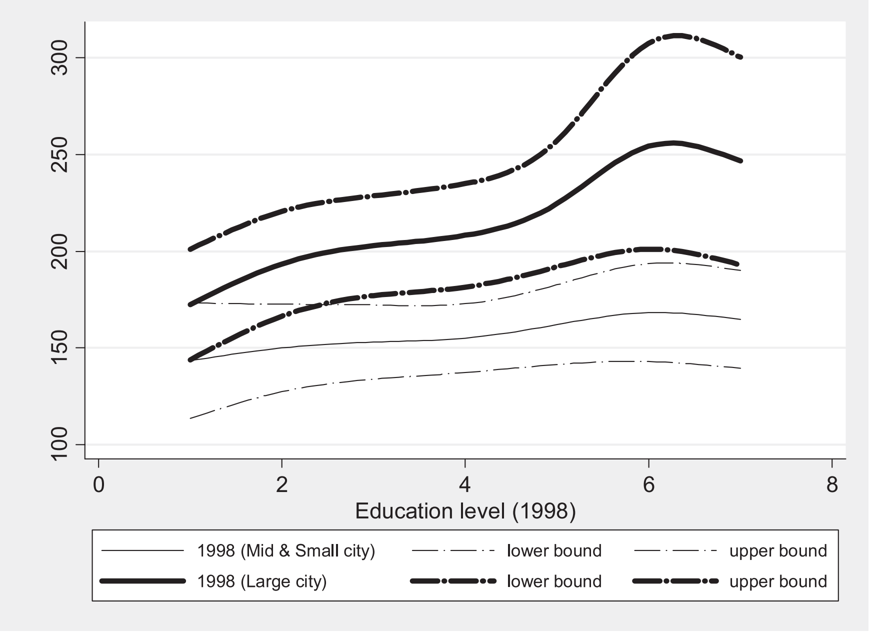

Fig. 5. Predicted FAFH expenditure versus education level (1998).

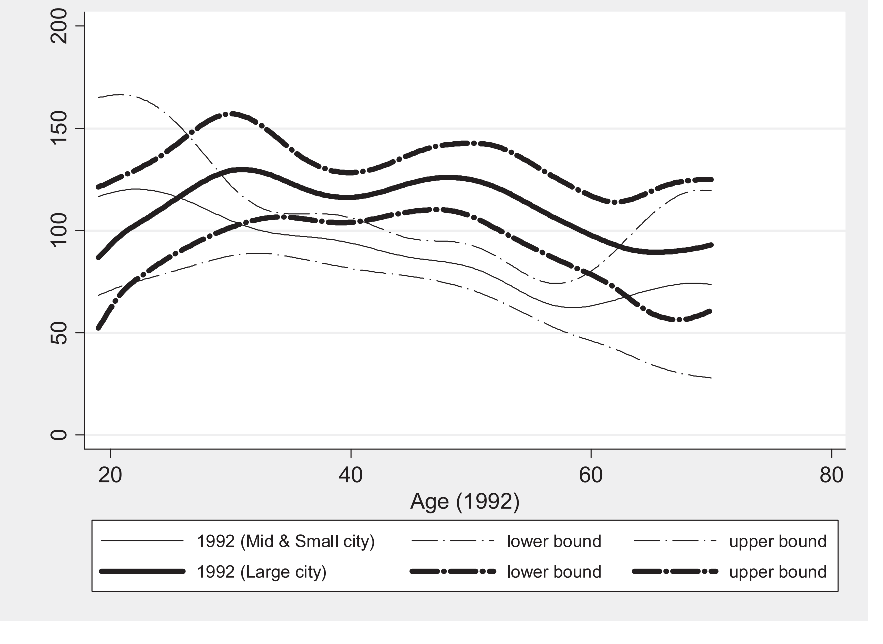

Fig. 6. Predicted FAFH expenditure versus age (1992).

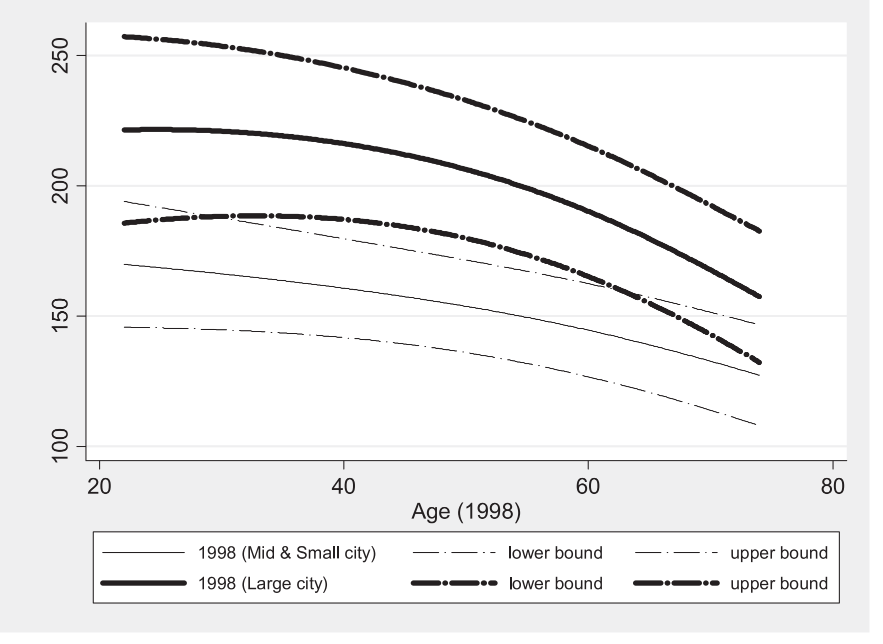

Fig. 7. Predicted FAFH expenditure versus age (1998).

downward trend, suggesting that the younger household head tend to spend more on FAFH.

4. Conclusion

In the middle of fast economic liberalization and growth in China, changes in Chinese food consumption behavior have not been given the attention it deserves. This paper is the first to examine the pattern and trends of FAFH expenditure emerging during the 1990s in China. Moreover, we use a novel nonparametric approach that handles mixed discrete and continuous variables in a coherent way. Nationwide Urban Household Survey data from 1992 and 1998 makes possible a more detailed investigation of FAFH consumption patterns. The robust nonparametric method employed here captures the stylized facts and patterns of FAFH expenditures by China's urban residents. Our results show income is a significant determinant of FAFH consumption levels. Income elasticities in China are much higher than in the United States. China's income elasticity between 1992 and 1998 has increased (the parametric model fails to detect this tendency), which suggests that the FAFH consumption market in China has yet to reach its full potential. Household size affects FAFH expenditures in an interesting way. We find that the three-person family spends the most on FAFH. Given that families of this size are a growing percentage in China, this result suggests that family size is becoming a more important factor in determining FAFH consumption.

Our nonparametric estimation results show significant nonlinearities in the FAFH regression function and a specification test suggests that the parametric linear model is misspecified. The cross-validation selected smoothing parameters suggest that in 1992 the consumption behaviors are quite different for large city and middle–small city households, and this difference becomes smaller in 1998. In contrast, the parametric model is unable to detect this tendency. For the parametric model, the coefficient of the city size variable is 31.81 in 1992 and 53.15 in 1998. It yields no clue that the difference in consumption pattern between large and middle–small cities is reduced from 1992 to 1998.

This study shows that the robust nonparametric estimation method employed can provide comprehensive and detailed statistical interpretations of FAFH consumption behavior and patterns in China. This is of key importance in determining marketing methods and targeting potential customers for the fast food and family restaurant industry. In the last decade, China experienced a rapid growth in per capita real income and urban population with annual growth at about 7% and 2.3%, respectively. These growth trends are expected to continue or even accelerate with China's admission into the WTO. Based on the estimated income elasticities from this study, China's urban FAFH expenditure is projected to grow at more than 8.5% or US$1.35 billion annually, keeping other factors constant. The growth in FAFH consumption provides great opportunities to foreign companies in the restaurant and prepared food sectors. In recent years, U.S.-owned fast food and family style restaurants have become more popular with Chinese consumers. U.S. food products have gained a good reputation for high quality, unique taste, and reliable supply. More Chinese restaurants are adopting Western menu items on their traditional menus (Brabant, 1999; Moustakerski & Brabant, 2001). The techniques employed in this paper provide a useful methodology for the study of these trends as they unfold.

Acknowledgement

We are grateful to the insightful comments from two referees who greatly improved the paper.

References

Ai, C., & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, *71*, 1795–1843.

Ai, C., & Sappington, D. (2002). The impact of the state incentive regulations on the U.S. telecommunication industry. *Journal of Regulatory Economics*, *22*, 133–160.

Anglin, P. M., & Gencay, R. (1993). Semiparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, *11*, 633–648.

Blundell, R., Duncan, A., & Pendakur, K. (1998). Semiparametric estimation and consumer demand. *Journal of Applied Econometrics*, *13*, 435–461.

Brabant, L. E. (1999). People's Republic of China Hotel, Restaurant and Institutional Sector for East China. U.S. Department of Agriculture, Foreign Agricultural Service, GAIN Report. CH9836, December.

Byrne, P. J., Capps, O., & Saha, A. (1996). Analysis of food-away-from-home expenditure patterns for U.S. households, 1982–89. *American Journal of Agriculture Economics*, *78*, 614–627.

Chern, W. S. (2000). Assessment of demand factors affecting global food security. In W. S. Chern, C. A. Carter, & S. Y. Shei (Eds.), *Food security in Asia: Economics and policies*. Edward Elgar.

Fan, S., Crammer, G. L., & Wailes, E. J. (1994). Food demand in Rural China: Evidence from rural household survey. *Agricultural Economics*, *11*, 61–69.

Fan, S., Crammer, G. L., & Wailes, E. J. (1995). Household demand in rural China: A two-state LES-AIDS model. *American Journal of Agricultural Economics*, *77*, 54–62.

Fang, C., & Beghin, J. C. (2002). Urban demand for edible oils and fats in China: Evidence from household survey data. *Journal of Comparative Economics*, *30*, 732–753.

Gale, F. (2002). China's food and agriculture: Issues for the 21st century. U.S. Department of Agriculture, Economics Research Service, Agriculture Information Bulletin No. 775.

Hall, P., Racine, J., & Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *Journal of the American Statistical Association*, (in press).

Han, T., Wailes, E. J., & Cramer, G. L. (1995). Rural and urban data collection in the Peoples Republic of China. In the China Market Data and Information Systems. Proceedings of WCC-101 Symposium.

Hong, Y., & White, H. (1995). Consistent specification testing via nonparametric series regressions. *Econometrica*, *63*, 1133–1159.

Hsiao, C., Li, Q., & Racine, J. (2003). Consistent Model Specification Tests with Mixed Categorical and Continuous Data. Revised and re-submitted to *International Economic Review*.

Jensen, H. H., & Yen, S. T. (1996). U.S. food expenditures away from home by type of meal. *Canadian Journal of Agricultural Economics*, *44*, 67–80.

Kanbur, R., & Zhang, X. (1999). Which regional inequality? The evolution of rural–urban and inland–coastal inequality in China from 1983–1995. *Journal of Comparative Economics*, *27*, 686–701.

Li, Q., & Racine, J. (2003). Cross-validation of kernel density estimation with both categorical and continuous data. *Journal of Multivariate Analysis*, *86*, 266–292.

Li, Q., & Racine, J. (2004). Nonparametric econometrics: Theory and practice. *A graduate book manuscript*. Princeton University Press (in press).

Liu, Z., & Stengos, T. (1999). Non-linearities in cross-country growth regressions: A semiparametric approach. *Journal of Applied Econometrics*, *14*, 527–538.

McCracken, V., & Brandt, J. (1987). Household consumption of food away from home: Total expenditure and by type of food facility. *American Journal of Agricultural Economics*, *69*, 274–284.

Moustakerski, P. & Brabant, L. (2001). People's Republic of China Retail Food Sector Report. U.S. Department of Agriculture, Foreign Agricultural Service, GAIN Report. Ch1810, November.

Pagan, A., & Ullah, A. (1999). Nonparametric econometrics. Cambridge, UK: Cambridge University Press.

Racine, J., & Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, *119*, 99–130.

Wang, Z., & Chern, W. S. (1992). Effects of rationing on the consumer behavior of Chinese urban households during 1981–1987. *Journal of Comparative Economics*, *16*, 1–16.

Wu, Y. (1999). China's consumer revolution: The emerging patterns of wealth and expenditure. Northampton, MA: Edward Elgar.

Yen, S. (1993). Working wives and food away from home: The Box-Cox double hurdle model. *American Journal of Agricultural Economics*, *75*, 884–895.