Indian Economic Review (2021) 56:25–48
https://doi.org/10.1007/s41775-020-00101-6

ARTICLE

Heterogeneity in food expenditure among US families: evidence from longitudinal quantile regression

Arjun Gupta[1] · Soudeh Mirghasemi[2] · Mohammad Arshad Rahman[1]

Published online: 11 January 2021

Abstract

Empirical studies on food expenditure are largely based on cross-section data and for a few studies based on longitudinal (or panel) data the focus has been on the conditional mean. While the former, by construction, cannot model the dependencies between observations across time, the latter cannot look at the relationship between food expenditure and covariates (such as income, education, etc.) at lower (or upper) quantiles, which are of interest to policymakers. This paper analyzes expenditures on total food (TF), food at home (FAH), and food away from home (FAFH) using mean regression and quantile regression models for longitudinal data to examine the impact of economic recession and various demographic, socioeconomic, and geographic factors. The data are taken from the Panel Study of Income Dynamics (PSID) and comprise of 2174 families in the United States (US) observed between 2001 and 2015. Results indicate that age and education of the head, family income, female-headed family, marital status, and economic recession are important determinants for all three types of food expenditure. Spouse education, family size, and some regional indicators are important for expenditures on TF and FAH, but not for FAFH. Quantile analysis reveals considerable heterogeneity in the covariate effects for all types of food expenditure, which cannot be captured by models focused on conditional mean. The study ends by showing that modeling conditional dependence between observations across time for the same family unit is crucial to reducing/avoiding heterogeneity bias and better model fitting.

Keywords Bayesian quantile regression · Great Recession · Heterogeneity bias · Longitudinal data · Mixed-effects · Mortgage

JEL Classification C11 · C31 · C33 · D10 · D12 · R20

✉ Mohammad Arshad Rahman
marshad@iitk.ac.in; arshadrahman25@gmail.com

Extended author information available on the last page of the article

1 Introduction

Food expenditure forms an integral part of the total family (or household) expenditure and is often categorized into food at home (FAH), food away from home (FAFH), and food delivered at home (FDAH). This categorization is relevant from a health perspective and other reasons. First, the division permits us to analyze the nutrition quality of food among families. This is important, because there are health implications of consuming more FAFH, as it is considered to be less nutritious than FAH (Mancino et al. 2009) and more energy dense (Binkley 2008). Some authors have also linked more FAFH to overweight and obesity (Cai et al. 2008). Second, the division allows us to answer interesting policy-oriented questions. For example, what is the effect of a female-headed family on FAH expenditure or does having a home mortgage reduce FAH and/or FAFH expenditures? Third, food assistance programs are often designed to minimize the health risks arising from deficient nutrition particularly among unemployed and lower income groups. This categorization can help assess the efficacy of food assistance program on FAH expenditure of the vulnerable groups, particularly during times of economic crisis.

As a result, the study of expenditure on FAH and FAFH has attracted considerable attention in the literature. Few previous studies using cross-section data include Lee and Brown (1986), Nayga (1996), Aguiar and Hurst (2005), and Liu et al. (2013). Lee and Brown (1986) employ a switching regression model on the 1977–78 Nationwide Food Consumption Survey data to examine expenditures on FAH and FAFH among the US households. Nayga (1996) utilizes the 1992 US consumer expenditure survey (CES) data to estimate the effect of wife's education and employment on three subcategories of food expenditure—for prepared food, food prepared at home, and food away from home. The modeling scheme utilized is a generalized version of Heckman's sample selection model (Heckman 1979). Aguiar and Hurst (2005) employs an instrumental variable linear regression to investigate, among other things, the effect of anticipated (i.e., retirement) and unanticipated (i.e., unemployment) shock to income on TF, FAH, and FAFH expenditures. The data are taken from the Continuing Survey of Food Intake of Individuals (CSFII, collected by the US Department of Agriculture) and corresponds to interviews conducted between 1989–1991 and 1994–1996, but the households are different in the two interviews. Liu et al. (2015) use a trivariate sample selection procedure to study patterns in FAFH expenditure among the Chinese households. In studies such as Liu et al. (2013), the use of the sample selection framework is motivated to account for the occurrence of zero expenditures, particularly on FAFH or in its subdivision (e.g., full-service restaurants, fast-food restaurants, and others). To get a more complete picture, readers may look into Table 1 of Davis (2014) for a brief summary of 17 articles (out of 20) on studies related to food expenditure using cross-section data.

The relationship of food expenditure (at home and away from home) to other covariates has been the focus of analysis in several cross-section studies. They include the relation of food expenditure (of various types) to consumer preferences (Stewart et al. 2005), family composition (Liu et al. 2013), race (Lanfranco et al. 2002), homeownership and mortgages (Nayga 1996; Mian et al. 2013), wife's labor force participation (Redman 1980; Kinsey 1983; Darian and Klein 1989; Yen 1993; Nayga 1996), children's welfare (Handa 1996), and obesity (Drichoutis et al. 2012). Some authors have also examined the effects of tax on food expenditure. For example, Zheng et al. (2019) examines the impact of tax on expenditure in grocery food (i.e., FAH) and restaurant food (i.e., FAFH) using a weekly data observed between April 2012 and January 2013, collected by United States Department of Agriculture (USDA). They find that tax on grocery (restaurant food) reduces expenditure on grocery (restaurant food) and increases expenditure on restaurant food (grocery).

The above paragraphs clearly indicate that there are ample cross-section studies on food expenditure, but panel or longitudinal studies are rather lacking with few exceptions. Cai et al. (2008) presents a state-level analysis of different types of food expenditure on overweight rates, obesity rates, and combined rates (the sum of overweight and obesity rates) using data from the Behavioral Risk Factor Surveillance System. The primary finding is that FAH (FAFH) expenditure is negatively (positively) associated to obesity and combined rates, and both FAH and FAFH expenditures do not significantly affect overweight rates. The only panel study mentioned in Davis (2014) is the article by Gelber and Mitchell (2012), where they use PSID and time diary data between 1975 and 2004, and find that for a decrease in income tax (i.e., incentive to join the labor force increases), single women are much more likely to increase FAFH expenditure to substitute for housework compared to single men. At the same time, the effect on FAH expenditure is statistically insignificant. Kohara and Kamiya (2016) use a panel data on Japanese households for the period 2004–2006 and find that mothers' labor supply decision has a negative effect on food produced at home. Moreover, the negative effect is common for all economic classes and more pronounced for the low economic class. Besides, there are abundant studies that examine the impact on food expenditure from participating in Supplemental Nutrition Assistance Program (SNAP), formerly known as Food Stamp Program (FSP)[1]. Few articles from this literature[2] include Hoynes and Schanzenbach (2009), Wilde et al. (2009), Beatty and Tuttle (2014), and Burney (2018). However, these studies focus on the conditional mean of the response variable and thus cannot explain the relationship at the quantiles.

The current study takes a broader perspective and looks at expenditures on total food (TF), food at home (FAH), and food away from home (FAFH), and explains its variation based on various demographic, socioeconomic, and geographic factors including mortgage and recession. The data are taken from Panel Study of Income Dynamics (PSID) and are composed of 2174 family units observed over the period 2001–2015. Since ours is a panel data, we exploit a longitudinal or panel regression

[1] The American Recovery and Reinvestment Act (ARRA) of 2009 renamed the FSP to SNAP and increased benefits by an average of \$80 per household. However, a common variable to capture SNAP participation pre- and post-ARRA is not available in PSID.

[2] Within the SNAP literature, the central debate is whether households respond similarly to an increase in cash income and in-kind transfer (food coupons). While some researchers, such as Hoynes and Schanzenbach (2009), have found that the response is similar; others such as Beatty and Tuttle (2014) have found that households increase in food expenditure is more when given an in-kind transfer (food stamps) as compared to cash income.

framework that can accommodate both common (fixed-effects) and individual-specific (random-effects) parameters (hence also known as mixed-effects model in Statistics)[3]. However, mean longitudinal regression is not capable of capturing the heterogeneity in covariate effects across the conditional distribution of the response variable. To overcome this limitation, we study the heterogeneous effect of the covariates on food expenditure (TF, FAH, and FAFH) using a quantile model for longitudinal data that accommodates both common effects and individual-specific effects, also known as quantile mixed models.

This paper contributes to the literature in at least three different ways. First, quantile longitudinal regression provides a comprehensive understanding of food expenditure pattern of family units to variation in covariates by providing estimates at different quantiles. The method is robust compared to standard longitudinal models where the focus is on the mean, because among other things, it is unaffected by the presence of outliers in the data. Second, this study adds to the understanding of the differences in food expenditure pre, during, and after the Great Recession. This enables us to capture patterns linking recession and food expenditure by categories which we explore in this study. To our knowledge, this is the first attempt to examine the effects of the Great Recession on food expenditure at home and away from home within a quantile panel data framework. Third, longitudinal data allow us to model the behavior of family units over time, which provides an advantage to control for unobserved heterogeneity leading to more robust estimates. As shown in this paper, it is important to control for this repeated behavior, because models which treat unobserved heterogeneity as a part of error term often result in inconsistent estimates and may lead to incorrect policy inference.

The remaining paper is organized as follows. Section 2 lays out the basic framework of the mean regression and quantile regression models for longitudinal data that we employ in our analysis. Section 3 presents a descriptive summary of the data and discusses the trends in variables over the time period of our study. Section 4 presents the results from the aforementioned regression models and shows the consequences of not modeling individual-specific heterogeneity. Finally, Sect. 5 presents concluding remarks.

2 Methodology

This section presents the mean regression for longitudinal data model and outlines the Bayesian approach for its estimation (Chib and Carlin 1999; Greenberg 2012). Thereafter, we present the Bayesian quantile regression for longitudinal data model and its estimation algorithm, which is inspired from Luo et al. (2012) and Rahman and Vossmeyer (2019).

[3] The terms fixed-effects and random-effects have been used to mean different things in the literature and there is no agreed-upon definition. In this paper, fixed-effects refers to regression coefficients that do not differ across i (or individuals) and random-effects mean regression coefficients that differ across i (see Greenberg 2012, Ch. 10). Andrew Gelman lists five different definitions of fixed-effects and random-effects at https://statmodeling.stat.columbia.edu/2005/01/25/why_i_dont_use/. But again, there are other popular definitions such as in Classical econometrics where fixed-effects means that the unobserved individual-specific heterogeneity is correlated with the regressors, while random-effects imply zero correlation (or more strongly statistical independence) between individual-specific heterogeneity and the regressors (see Cameron and Trivedi 2005; Wooldridge 2010; Hsiao 2014; Greene 2017).

2.1 Mean regression for longitudinal data

The longitudinal data model can be expressed in terms of the following equation:

$$y_{it} = x'_{it}\beta + s'_{it}\alpha_i + \epsilon_{it}, \quad \forall i = 1, \dots, n, \quad t = 1, \dots, T, \tag{1}$$

where y_{it} denotes the value of the response y for the ith individual at the tth time period, x'_{it} is a $1 \times k$ vector of explanatory variables, β is $k \times 1$ vector of common (fixed-effects) parameters, s'_{it} is a $1 \times l$ vector of covariates (often a subset of x_{it}) with individual-specific effects, α_i is an $l \times 1$ vector of individual-specific (random-effects) parameters included to capture the marginal dependence between observations on the same individual, and ϵ_{it} is the error term assumed to be independently and identically distributed (*iid*) as a normal distribution, i.e., $\epsilon_{it} \overset{iid}{\sim} N(0, h^{-1})$ for all values of $i = 1, \dots, n; t = 1, \cdots, T$, where h^{-1} is the variance. The distributional assumption on the error implies that y_{it} conditional on α_i are independently distributed as a normal distribution, i.e., $y_{it}|\alpha_i \sim N(x'_{it}\beta + s'_{it}\alpha_i, h^{-1})$ for all $i = 1, \dots, n; t = 1, \dots, T$.

In this paper, the response variable y will either be TF, FAH, or FAFH expenditures. The vector x_{it} will consist of a common intercept and a host of covariates related to demographic, socioeconomic and geographic factors. Finally, the vector of covariates with individual-specific effects s'_{it} will consist of an intercept and inverse-hyperbolic sine transformation of income.

To proceed with the Bayesian estimation of the longitudinal model, we first stack the model for each individual i. This is convenient for multiple reasons including reducing the computational burden. We define $y_i = (y_{i1}, \dots, y_{iT})'$, $X_i = (x'_{i1}, x'_{i2}, \dots, x'_{iT})'$, $S_i = (s'_{i1}, s'_{i2}, \dots, s'_{iT})'$, $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iT})'$. The resulting stacked model can be written as:

$$\begin{aligned} y_i &= X_i\beta + S_i\alpha_i + \epsilon_i, \quad \text{for } i = 1, \dots, n, \\ &\alpha_i|\Sigma \sim N_l(0, \Sigma), \\ &\beta \sim N_k(\beta_0, B_0), \qquad \Sigma^{-1} \sim Wish(\nu_0, D_0), \qquad h \sim Ga(c_0/2, d_0/2), \end{aligned} \tag{2}$$

where we assume that $\alpha_i|\Sigma$ are mutually independent and identically distributed as $N_l(0, \Sigma)$, and the last line represents the prior distributions, with N, *Wish*, and *Ga* denoting the normal, Wishart, and gamma distributions, respectively. The model given by Eq. (2) implies that the conditional density $y_i|\alpha_i \sim N(X_i\beta + S_i\alpha_i, h^{-1}I_T)$ for $i = 1, \dots, n$. The complete data density is then given by:

$$f(y, \alpha|\beta, h, \Sigma) = \prod_{i=1}^{n} f(y_i, \alpha_i|\beta, h, \Sigma) = \prod_{i=1}^{n} f(y_i|\beta, \alpha_i, h)\pi(\alpha_i|\Sigma),$$

which is equivalent to the complete data likelihood when viewed as a function of the parameters.

Algorithm 1

1. Sample (β, α) in one block as follows:
   (a) Let $\Psi_i = S_i \Sigma S_i' + h^{-1} I_T$. Sample β marginally of α from $\beta|y, h, \Sigma \sim N(\widetilde{\beta}, \widetilde{B})$, where,
   $\widetilde{B}^{-1} = \left(\sum_{i=1}^{n} X_i' \Psi_i^{-1} X_i + B_0^{-1} \right)$, and $\widetilde{\beta} = \widetilde{B} \left(\sum_{i=1}^{n} X_i' \Psi_i^{-1} y_i + B_0^{-1} \beta_0 \right)$.
   (b) Sample $\alpha_i | y, \beta, h, \Sigma \sim N(\widetilde{a}, \widetilde{A})$ for $i = 1, \cdots, n$, where,
   $\widetilde{A}^{-1} = (h S_i' S_i + \Sigma^{-1})$, and $\tilde{a} = \widetilde{A} \left(h S_i' (y_i - X_i \beta) \right)$.
2. Sample $\Sigma^{-1} | \alpha \sim Wish(v_1, D_1)$, where $v_1 = (v_0 + n)$, and $D_1^{-1} = \left(D_0^{-1} + \sum_{i=1}^{n} \alpha_i \alpha_i' \right)$.
3. Sample $h | y, \beta, \alpha \sim Ga(c_1/2, d_1/2)$ where:
   $c_1 = (c_0 + nT)$, and $d_1 = d_0 + \sum_{i=1}^{n} (y_i - X_i \beta - S_i \alpha_i)' (y_i - X_i \beta - S_i \alpha_i)$.

By Bayes' theorem, the complete data posterior density can be written as product of the complete data likelihood times the prior distributions as follows:

$$
\begin{aligned}
\pi(\beta, \alpha, \Sigma^{-1}, h|y) &\propto \left\{ \prod_{i=1}^{n} f(y_i|\beta, \alpha_i, h) \pi(\alpha_i|\Sigma) \right\} \pi(\beta) \pi(\Sigma^{-1}) \pi(h) \\
&\propto h^{nT/2} \exp \left[-\frac{h}{2} \sum_{i=1}^{n} (y_i - X_i\beta - S_i\alpha_i)'(y_i - X_i\beta - S_i\alpha_i) \right] \\
&\times |\Sigma|^{-\frac{n}{2}} \exp \left[-\frac{1}{2} \sum_{i=1}^{n} \alpha_i' \Sigma^{-1} \alpha_i \right] \exp \left[-\frac{1}{2} (\beta - \beta_0)' B_0^{-1} (\beta - \beta_0) \right] \\
&\times |\Sigma^{-1}|^{\frac{(v_0 - l - 1)}{2}} \exp \left[-\frac{1}{2} tr(D_0^{-1} \Sigma^{-1}) \right] \times h^{\frac{c_0}{2} - 1} \exp \left[-\frac{d_0 h}{2} \right].
\end{aligned}
\tag{3}
$$

The conditional posterior distributions are derived from the complete data posterior (Eq. 3) and the model is estimated using Gibbs sampling, a well-known Markov chain Monte Carlo method (Geman and Geman 1984; Casella and George 1992). The MCMC algorithm for estimating the model is presented in Algorithm 1. The parameters (β, α) are sampled jointly to avoid correlation between the parameters, because the covariates in s_{it} are often a subset of x_{it} (Greenberg 2012, Chap.10). Specifically, we first sample β (marginally of α, but conditional on other model parameters) from an updated normal distribution and then sampled α (conditional on β and other model parameters) from its updated normal distribution. The precision matrix Σ^{-1} is sampled from an updated Wishart distribution and, finally, the precision parameter h is sampled from an updated gamma distribution.

2.2 Quantile regression for longitudinal data

The quantile regression for longitudinal data can be expressed in terms of the following equation:

$$y_{it} = x_{it}'\beta + s_{it}'\alpha_i + \epsilon_{it}, \quad \forall i = 1, \dots, n, \quad t = 1, \dots, T, \tag{4}$$

where all the notations are same as in Sect. 2.1, except that the errors are assumed to be *i.i.d.* as an asymmetric Laplace (AL) distribution, i.e., $\epsilon_{it} \overset{iid}{\sim} AL(0, h^{-1}, p)$, where h^{-1} is the inverse of the scale parameter and p denotes a quantile. This implies that y_{it}, conditional on α_i, are independently distributed as an AL distribution, i.e., $y_{it}|\alpha_i \sim AL(x_{it}'\beta + s_{it}'\alpha_i, h^{-1}, p)$ for $i = 1, \cdots, n, t = 1, \dots, T$. Note that the error distribution is assumed to be AL to form a working likelihood, because the quantile loss function appears in the exponent of an AL distribution (see Yu and Moyeed 2001; Rahman 2016). The resulting conditional quantile function for response y_{it} is:

$$Q_{y_{it}}(p|x_{it}, \alpha_i) = x_{it}'\beta + s_{it}'\alpha_i,$$

where $Q_{y_{it}} \equiv F_{y_{it}}^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the outcome variable conditional on the individual-specific parameters and the covariates.

We can directly work with the AL distribution; however, it is not convenient for Gibbs sampling. Therefore, as proposed in Kozumi and Kobayashi (2011), we make use of the normal–exponential mixture representation of the AL distribution:

$$\epsilon_{it} = h^{-1}\theta w_{it} + h^{-1}\tau\sqrt{w_{it}}\, u_{it}, \quad \forall\ i = 1, \dots, n;\ t = 1, \dots, T, \tag{5}$$

where $u_{it} \sim N(0, 1)$ is mutually independent of $w_{it} \sim \mathcal{E}(1)$, $\theta = \frac{1-2p}{p(1-p)}$, $\tau = \sqrt{\frac{2}{p(1-p)}}$, and the symbol $\mathcal{E}$ denotes an exponential distribution. The resulting quantile regression for longitudinal data model can be expressed as:

$$y_{it} = x_{it}'\beta + s_{it}'\alpha_i + \theta v_{it} + \tau\sqrt{h^{-1}v_{it}}\, u_{it}, \quad \forall i = 1, \dots, n, \quad t = 1, \dots, T, \tag{6}$$

where we have used the transformation $v_{it} = w_{it}/h$, since the presence of the scale parameter in the conditional mean is not conducive to Gibbs sampling (Kozumi and Kobayashi 2011; Rahman and Karnawat 2019). See also Bresson et al. (2020) and Ojha and Rahman (2020), where the scale is fixed at 1 to identify the parameters of quantile regression with binary outcomes.

To proceed with the Bayesian estimation, we again stack the model across i for reasons mentioned earlier. Define $y_i = (y_{i1}, \dots, y_{iT})'$, $X_i = (x_{i1}', x_{i2}', \dots, x_{iT}')'$, $S_i = (s_{i1}', s_{i2}', \dots, s_{iT}')'$, $D_{\tau\sqrt{\frac{v_i}{h}}} = diag(\tau\sqrt{\frac{v_{i1}}{h}}, \dots, \tau\sqrt{\frac{v_{iT}}{h}})$, $u_i = (u_{i1}, \dots, u_{iT})'$, and, finally, $v_i = (v_{i1}, \dots, v_{iT})'$. The resulting stacked quantile regression for longitudinal data can be written as:

$$\begin{aligned} y_i =& X_i\beta + S_i\alpha_i + \theta v_i + D_{\tau\sqrt{\frac{v_i}{h}}}\, u_i, \quad \text{for } i = 1, \dots, n, \\ & \alpha_i|\Sigma \sim N_l(0, \Sigma), \qquad v_{it} \sim \mathcal{E}(1/h), \qquad u_{it} \sim N(0, 1), \\ & \beta \sim N_k(\beta_0, B_0), \qquad \Sigma^{-1} \sim Wish(\nu_0, D_0), \qquad h \sim Ga(c_0/2, d_0/2), \end{aligned} \tag{7}$$

where we assume that $\alpha_i|\Sigma$ are mutually independent and identically distributed as $N_l(0, \Sigma)$, and the last line represents the prior distributions of the model parameters. The quantile model given by Eq. (7) implies that the conditional density

$y_i|\alpha_i \sim N(X_i\beta + S_i\alpha_i + \theta\nu_i, D^2_{\tau\sqrt{\frac{\nu_i}{h}}})$ for $i = 1, \ldots, n$. The complete data density is then given by $f(y, \alpha|\beta, \nu, h, \Sigma) = \prod_{i=1}^{n} f(y_i, \alpha_i|\beta, \nu_i, h, \Sigma) = \prod_{i=1}^{n} f(y_i|\beta, \alpha_i, \nu_i, h)\pi(\alpha_i|\Sigma)$

Algorithm 2

1. Sample (β, α) in one block as follows:

   (a) Let $\Omega_i = \left(S_i \Sigma S_i' + D^2_{\tau\sqrt{\frac{\nu_i}{h}}} \right)$. Sample β marginally of α from $\beta|y, \nu, \Sigma, h \sim N\big(\widetilde{\beta}, \widetilde{B}\big)$, where,

   $$\widetilde{B}^{-1} = \left(\sum_{i=1}^{n} X_i' \Omega_i^{-1} X_i + B_0^{-1} \right), \quad \text{and} \quad \widetilde{\beta} = \widetilde{B}\left(\sum_{i=1}^{n} X_i' \Omega_i^{-1} (y_i - \theta\nu_i) + B_0^{-1}\beta_0 \right).$$

   (b) Sample $\alpha_i|y, \beta, \nu, h, \Sigma \sim N\big(\widetilde{a}, \widetilde{A}\big)$ for $i = 1, \ldots, n$, where,

   $$\widetilde{A}^{-1} = \left(S_i' D^{-2}_{\tau\sqrt{\frac{\nu_i}{h}}} S_i + \Sigma^{-1} \right), \quad \text{and} \quad \widetilde{a} = \widetilde{A}\left(S_i' D^{-2}_{\tau\sqrt{\frac{\nu_i}{h}}} (y_i - X_i\beta - \theta\nu_i) \right).$$

2. Sample $\nu_{it}|y_{it}, \beta, \alpha_i, h \sim GIG\left(0.5, \widetilde{\lambda}_{it}, \widetilde{\eta}\right)$ for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where,

   $$\widetilde{\lambda}_{it} = h\left(\frac{y_{it} - x_{it}'\beta - s_{it}'\alpha_i}{\tau} \right)^2 \quad \text{and} \quad \widetilde{\eta} = h\left(\frac{\theta^2}{\tau^2} + 2 \right).$$

3. Sample $\Sigma^{-1}|\alpha \sim Wish\big(\nu_1, D_1\big)$, where $\nu_1 = (\nu_0 + n)$, and $D_1^{-1} = \left(D_0^{-1} + \sum_{i=1}^{n} \alpha_i\alpha_i' \right)$.

4. Sample $h|y, \beta, \alpha, \nu \sim Ga\big(c_1/2, d_1/2\big)$ where,

   $$c_1 = \left(c_0 + 3nT\right), \quad \text{and} \quad d_1 = d_0 + 2\sum_{i=1}^{n}\sum_{t=1}^{T} \nu_{it} + \sum_{i=1}^{n}\sum_{t=1}^{T} \frac{\left(y_{it} - x_{it}'\beta - s_{it}'\alpha_i - \theta\nu_{it}\right)^2}{\tau^2\nu_{it}}.$$

Once again, we employ Bayes' theorem to obtain the complete data posterior as the product of the complete data likelihood times the prior distributions as follows:

$$\begin{aligned}
\pi(\beta, \alpha, \nu, \Sigma^{-1}, h|y) &\propto \left\{ \prod_{i=1}^{n} f(y_i|\beta, \alpha_i, \nu_i, h)\pi(\alpha_i|\Sigma)\pi(\nu_i) \right\} \pi(\beta)\pi(\Sigma^{-1})\pi(h) \\
&\propto \prod_{i=1}^{n} \left\{ |D^2_{\tau\sqrt{\frac{\nu_i}{h}}}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y_i - X_i\beta - S_i\alpha_i - \theta\nu_i)' D^{-2}_{\tau\sqrt{\frac{\nu_i}{h}}} (y_i - X_i\beta - S_i\alpha_i - \theta\nu_i) \right] \right\} \\
&\times |\Sigma^{-1}|^{\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} \alpha_i'\Sigma^{-1}\alpha_i \right] \times h^{nT} \exp\left[-h\sum_{i=1}^{n}\sum_{t=1}^{T} \nu_{it} \right] \times h^{\frac{c_0}{2}-1} \exp\left(-\frac{d_0 h}{2} \right) \\
&\times \exp\left[-\frac{1}{2}(\beta - \beta_0)' B_0^{-1} (\beta - \beta_0) \right] \times |\Sigma^{-1}|^{\frac{(\nu_0 - l - 1)}{2}} \exp\left[-\frac{1}{2} tr(D_0^{-1}\Sigma^{-1}) \right].
\end{aligned} \tag{8}$$

The conditional posteriors can be derived from the joint posterior distribution (Eq. 8) and the model can be estimated using Gibbs sampling, as presented in Algorithm 2. Specifically, we sample β and α in a single block to elude the problem of poor mixing due to correlation between the parameters for reasons mentioned earlier (see also Rahman and Vossmeyer 2019; Bresson et al. 2020). The common effects parameters β, marginally of α, are sampled from an updated normal distribution and the individual-specific parameters α_i's are sampled from their respective updated normal distribution. The mixture variable ν is sampled component-wise from an updated generalized inverse Gaussian (GIG) distribution (Devroye 2014). The precision matrix Σ^{-1} is sampled from an updated Wishart distribution and the parameter h is sampled from an updated gamma distribution.

3 Data

The current study utilizes data from the Panel Study of Income Dynamics (PSID), which began in 1968, and is the longest running longitudinal household survey in the world. We constructed a balanced panel of 2174 family units with data for each alternate year, i.e., 2001, 2003, 2005, 2007, 2009, 2011, 2013, and 2015. This is because beginning 1997, the PSID collects data every alternate year. Our constructed data have information on different types of food expenditures, considered as dependent variables, and a host of socioeconomic, demographic, and geographic variables which are used as covariates or independent variables in our study. Table 1 presents a descriptive summary of the variables considered in our analysis.

The primary variable of interest is the food expenditure of a family unit, which the PSID categorizes into three types: food at home (FAH), food away from home (FAFH), and food delivered at home (FDAH). The sum of these three expenditures yields total food (TF) expenditure of the family unit. The variable FAH represents the annualized expenditure of family unit at home and in our sample lies between \$0 and \$36400. There are only few observations with zero value for FAH. Similarly, the variable FAFH represents annualized food expenditure away from home and in the sample lies in the range \$0 to \$44,200. The zero values for FAFH are small at 5.7% of the total number of observations. All observations with zero TF expenditure were removed from the sample. Our study considers expenditure on TF, FAH, and FAFH as the dependent variable in different regressions. The expenditure on FDAH is dropped due to large number of zero values, which makes censoring important and a sample selection framework more appropriate.

An interesting characteristic about the distribution of food expenditures is that they are positively skewed. Figure 1 presents a box plot of the different types of food expenditure utilized in the study. Each box plot represents the distribution of food expenditure for a particular year. In each box plot, the solid line within the box shows the median value, while the bottom and top of the box represent the 25th and 75th percentiles, respectively. The vertical lines are whiskers and they show either the maximum/minimum values or 1.5 times the interquartile range of the data, whichever is smaller. Points more than 1.5 times the interquartile range below (above) the first (third) quartile are defined as outlier and plotted individually. As seen from Fig. 1, for each box plot (across different types of food expenditure), there are large number of outliers towards the higher values making the distribution positively skewed. Consequently, the mean food expenditure (which is pushed upward due to the presence of high values) and covariate effects at the conditional mean is inadequate for a complete picture. In the literature, studies have used logarithmic

Table 1 Data summary

Variables/Years	2001	2003	2005	2007	2009	2011	2013	2015
TF/1000	6.70 (3.62)	6.89 (3.73)	7.40 (4.18)	7.93 (4.64)	7.91 (4.52)	8.22 (4.75)	8.56 (5.18)	8.90 (5.50)
FAH/1000	4.60 (2.64)	4.70 (2.64)	5.00 (2.82)	5.42 (3.17)	5.57 (3.26)	5.79 (3.44)	6.02 (3.71)	6.17 (3.79)
FAFH/1000	1.99 (1.96)	2.04 (2.02)	2.28 (2.48)	2.40 (2.65)	2.24 (2.32)	2.33 (2.43)	2.44 (2.65)	2.63 (2.85)
Head Age	44.96 (12.36)	46.97 (12.37)	48.93 (12.37)	50.96 (12.37)	52.95 (12.36)	54.96 (12.37)	56.95 (12.36)	58.98 (12.33)
Head Edu	13.38 (2.68)	13.43 (2.64)	13.43 (2.64)	13.43 (2.64)	13.66 (2.62)	13.66 (2.62)	13.68 (2.63)	13.69 (2.63)
Spouse Edu	9.50 (6.47)	9.63 (6.44)	9.78 (6.40)	9.90 (6.36)	10.16 (6.51)	10.09 (6.53)	9.99 (6.61)	9.93 (6.66)
Family Size	2.94 (1.47)	2.91 (1.45)	2.87 (1.42)	2.82 (1.43)	2.76 (1.43)	2.66 (1.39)	2.58 (1.37)	2.48 (1.32)
Family Income/10000	7.72 (8.46)	7.97 (12.32)	8.78 (16.05)	9.22 (9.48)	9.59 (9.74)	9.34 (10.31)	9.83 (13.23)	9.94 (10.03)
Head Emp	0.85	0.85	0.84	0.81	0.73	0.69	0.66	0.62
Head Female	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Married	0.68	0.69	0.70	0.71	0.71	0.71	0.70	0.70
Single	0.15	0.13	0.11	0.11	0.10	0.10	0.10	0.09
Homeowner	0.75	0.78	0.80	0.81	0.81	0.81	0.81	0.81
Mortgage	0.59	0.60	0.61	0.61	0.61	0.57	0.55	0.53
White	0.68	0.68	0.71	0.71	0.71	0.71	0.71	0.71
Non-White	0.32	0.32	0.29	0.29	0.29	0.29	0.29	0.29
Recession	1.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00
Northeast	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.15
West	0.20	0.19	0.19	0.20	0.19	0.19	0.19	0.19
South	0.38	0.38	0.38	0.38	0.39	0.39	0.39	0.40

The table presents the mean and standard deviation (in parenthesis) of the continuous variables and proportion of the categorical variables for each considered year

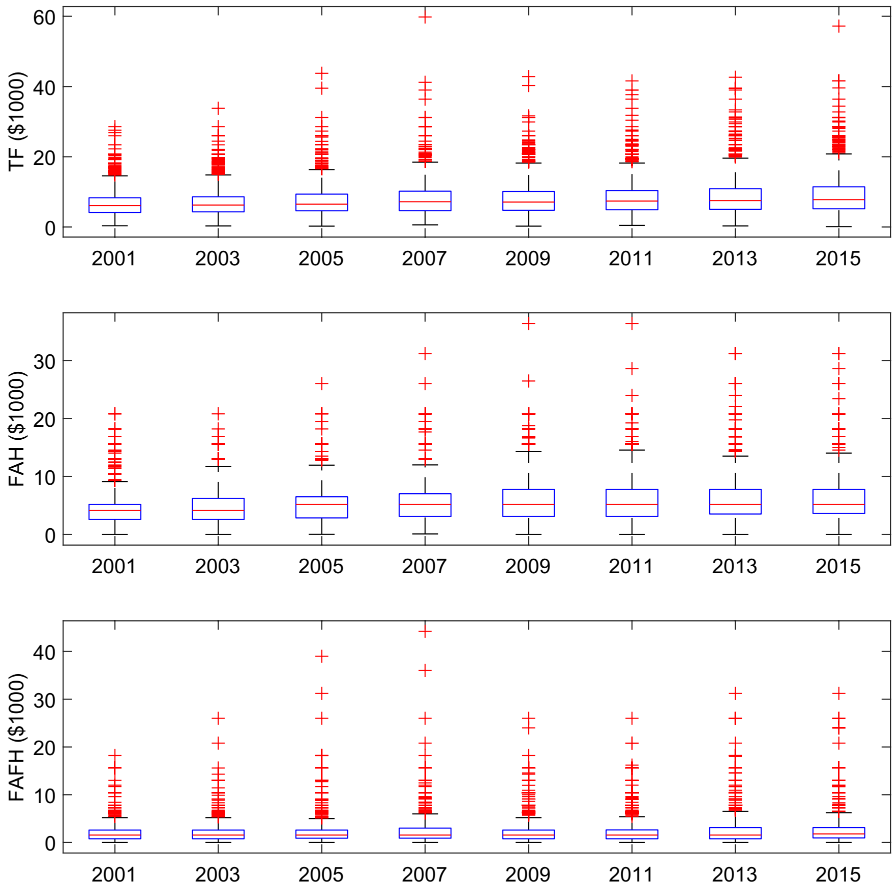

Fig. 1 Box plot for different types of food expenditure

transformation of food expenditure to alleviate this problem of heteroscedasticity (Liu et al. 2013). However, taking a logarithmic transformation cannot eliminate the non-normality or the heteroscedasticity problem. Besides, food and nutritional assistance programs (such as Supplemental Nutrition Assistance Program or SNAP) are typically interested in the lower tail (i.e., families/households with low food expenditure) to ensure food security.

The covariates or independent variables utilized in this study (see Table 1) include age of the head (*Head Age*), education of the head (*Head Edu*), and the spouse (*Spouse Edu*), measured as the number of years of schooling and takes value between 0 and 17 (17 represents post-graduate level work and above). *Family Size* represents number of members in a family unit. The variable *Family Income* indicates the actual value of income including transfer income in the previous year (negative values representing loss). We use the inverse-hyperbolic sine (IHS) transformation on income variable, because it adjusts for skewness and retains 0 and negative values (Friedline et al. 2015; Rahman and Vossmeyer 2019). The indicator for employment status of the head (*Head Emp*) equals 1 if the head is employed and 0 otherwise (omitted). The omitted category includes respondents that are temporarily laid off, looking for work, retired, permanently/temporarily disabled, keeping house, student, and others.

The indicator for gender of the head (*Head Female*) is coded as 1 if 'female' and 0 if 'male', while the marital status (of the head) is categorized into *Married*, *Single*, and *Separated* (omitted). The omitted category (*Separated*) consists of respondents who are widowed, divorced/annulled, or separated. Other variables included in our study are indicators for homeownership (*Homeowner*) and mortgages (*Mortgage*). The variable *Homeowner* takes the value 1 if the respondent is a homeowner and 0 otherwise. Similarly, we have the *Mortgage* variable which equals 1 if the respondent has a mortgage on property and 0 otherwise. Race is categorized into *White* and *Non-White* composed of Blacks, American Indian, Aleut, Eskimo, Asian, Pacific Islander, and Latino. Besides, we have indicators for recession years and the region in which the family resides. The recession dummy takes the value 1 for the years 2001, 2007, and 2009, because these were years with some recession period. Following the US Census Bureau, the region variable is classified into *Northeast*, *West*, *South*, and *Midwest* (omitted). Including regional indicators help us to look at differences, if any, in the expenditure behavior of the families across regions.

We now look at the movement in average values of the variables for the sampled period. The average FAH expenditure for a typical family unit is around \$4590 in 2001, while the FAFH expenditure is around \$1990 for the same year. The average expenditure on TF is approximately \$6700 in 2001 and increases to \$8900 in 2015. Not surprisingly, the average expenditure on FAFH and TF was lower in the year 2009 compared to its respective values in 2007. This shows the adverse effect of the economic crisis on average food expenditure. The adverse effect seems to persist longer for FAFH expenditure, as its average value in 2011 is lower compared to 2007.

The average age of the head is around 45 years with a family size of approximately 3 members in 2001. In the sample, the family units are predominantly headed by males (about 82%) with an average of 13.38 years of schooling in 2001. The average years of schooling of the spouse are lower than that of the head and stands at 9.5 years in 2001, but increase to approximately 10 years in 2015. The sample clearly shows the effect of the Great Recession (December 2007–June 2009) on the variables *Family Income* and *Head Emp*. The mean annual family income is approximately \$77,000 in 2001 and increases to \$99,300 in 2015. However, there is a drop in average family income for 2011 compared to 2009. The effect of the economic crisis is much more pronounced on employment status of the head. In the sample, about 85% are employed in 2001, which started decreasing in 2005 and stood at 73% during 2009. However, the lowest percentage employed for the sample is 62% in 2015.

A large proportion of the sampled respondents are married (0.68 in 2001) and remain in the range 0.68–0.71 throughout the period of our study, while the proportion of single decreases from 0.15 to 0.09 between 2001 and 2015. Approximately 75% of the families own a house in 2001 and this proportion reaches 81% in 2007, and remains in that region for subsequent years. The proportion of respondents having a mortgage on property decrease from 0.59 to 0.53 between 2001 and 2015. Nonetheless, the mortgage percentage was higher than 0.59 between 2003 and 2009, which is another hallmark of the Great Recession. On the racial aspect, majority of the sampled families (about 68%) are White, while the remaining 32% consists of Blacks and other races, thus giving a diverse sample for the study. Our sample is also geographically heterogeneous. Most of the sampled respondents live in the South (38%), followed by Midwest (26%), West (20%), and Northeast (16%). This percentage is stable over the sample period, suggesting little geographic mobility across regions.

4 Results

This section discusses the results for the three types of food expenditure using the models presented in Sect. 2. In particular, the results from longitudinal mean regression are presented in Table 2, and the results from longitudinal quantile regression are exhibited in Table 3. The posterior estimates are based on 12,000 MCMC iterations after a burn-in of 3,000 iterations. Trace plots of the MCMC draws, not presented for the sake of brevity, mimic that of white noise and confirm that the chains have converged. Moderately diffused priors are utilized for the parameters in both the models: $\beta \sim N_k(0, 100 * I)$, $\alpha_i \sim N_l(0, I)$, $\Sigma^{-1} \sim Wish(5, 10 * I_l)$ and $h \sim IG(10/2, 9/2)$. Note that the definition of h are different in the mean and quantile regression models. Besides, there are two components of α_i, individual-specific intercept and individual-specific coefficient for inverse-hyperbolic sine transformation of income. With respect to individual-specific effects, results from Tables 2, 3 show that the standard deviations of α_i (i.e., $(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}})$ are different for the mean and quantile regression models. As such, a modeling approach with identical variances should be avoided. We now discuss the results for the common parameters in all the econometric models.

The results from the longitudinal mean regression, presented in Table 2, show that (logarithm of) *Head Age* positively affects expenditures on TF, FAH, and FAFH. Comparing the coefficients across categories, we observe that the coefficient for logarithm of *Head Age* in the FAH equation is much higher (more than four times) than its corresponding value in the FAFH equation. The result agrees with the intuition that people prefer eating at home as they get older, because FAH is considered to be much healthier. Another argument put forward by Liu et al. (2013) is that social activity reduces with age leading to lower rise in FAFH expenditure. Other studies that have found a positive coefficient for *Head Age* include Redman (1980), Nayga (1996), Stewart and Yen (2004), and Zheng et al. (2019). Moving to the results from quantile regression shown in Table 3, we observe that there is considerable variation in the coefficients for logarithm of *Head Age*. For example, in the FAH (FAFH) equation, the ratio of coefficients from *Head Age* between 80th and 20th quantiles is 1.86 (3.58). These differences show considerable heterogeneity in the effect of *Head Age* on different types of food expenditure.

Table 2 Posterior mean (MEAN) and standard deviation (STD) of the parameters from longitudinal mean regression

	TF		FAH		FAFH	
	MEAN	STD	MEAN	STD	MEAN	STD
Intercept	– 12.53	0.84	– 10.28	0.63	– 2.12	0.43
log (Head Age)	3.65	0.19	2.94	0.14	0.66	0.10
Head Edu	0.08	0.02	0.04	0.01	0.05	0.01
Spouse Edu	0.06	0.01	0.05	0.01	0.01	0.01
FamilySize	0.76	0.03	0.73	0.02	0.03	0.02
IHS Income	2.79	0.14	1.37	0.10	1.35	0.07
Head Emp(HE)	0.15	0.08	0.04	0.06	0.12	0.05
Head Female (HF)	– 1.51	0.20	– 0.79	0.15	– 0.71	0.11
HE × HF	0.62	0.17	0.37	0.13	0.26	0.09
Married	– 0.09	0.19	– 0.01	0.14	– 0.02	0.10
Single	0.96	0.17	0.44	0.12	0.43	0.09
Mort × Home	0.10	0.07	0.10	0.06	0.02	0.04
Non-White	– 0.59	0.12	– 0.35	0.09	– 0.26	0.06
Recession	– 0.26	0.05	– 0.19	0.04	– 0.08	0.03
Northeast	0.63	0.17	0.47	0.12	0.14	0.09
West	0.64	0.15	0.53	0.11	0.12	0.08
South	0.54	0.13	0.35	0.10	0.21	0.07
h	0.13	0.01	0.22	0.01	0.42	0.01
$\sigma_{11}^{\frac{1}{2}}$	2.31	0.10	1.57	0.08	0.99	0.06
$\sigma_{22}^{\frac{1}{2}}$	3.07	0.13	1.87	0.10	1.66	0.07
$\rho_{1,2}$	– 0.42	0.05	– 0.32	0.07	– 0.28	0.06

$h = \sigma^{-2}$ in mean regression

The two education variables *Head Edu* and *Spouse Edu* positively affect TF and FAH expenditures. Zheng et al. (2019) also finds a positive effect of head's education on FAH expenditure. For the FAFH expenditure, only *Head Edu* has a positive effect, but *Spouse Edu* has no effect (statistically speaking), because the credible interval for *Spouse Edu* contains zero. This implies that higher educated spouses (mostly females in our sample) are more knowledgable to understand the importance of healthy diet and consequently spend more on FAH, but not on FAFH. Our findings are similar to those reported by Redman (1980), and Kohara and Kamiya (2016). The results from quantile regression show considerable heterogeneity in the covariate effects, but a comparison of the coefficients for *Head Edu* and *Spouse Edu* seems more interesting. For FAH expenditure, across quantiles, the coefficient for *Spouse Edu* is always higher than that of *Head Edu* (at the median, the coefficient of *Head Edu* is 0.49 and that of *Spouse Edu* is 0.51). This implies that spouses (mostly female in our sample) have a larger positive impact on FAH expenditure across its distribution. In contrast, for FAFH expenditure, the coefficients for *Head Edu* are higher across quantiles compared to *Spouse Edu*). This implies that an increase in head's education leads to a higher increase in consumption of outside food. A

Table 3 Posterior mean (MEAN) and standard deviation (STD) of the parameters from longitudinal quantile regression

	TF						FAH						FAFH					
	20th		50th		80th		20th		50th		80th		20th		50th		80th	
	MEAN	STD	MEAN	MEAN	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
Intercept	−9.22	0.61	−11.72	0.70	−14.94	0.86	−7.83	0.44	−10.24	0.52	−13.15	0.66	−0.69	0.22	−1.12	0.30	−1.48	0.43
log(Head Age)	2.56	0.14	3.41	0.16	4.68	0.20	2.13	0.10	2.88	0.12	3.97	0.15	0.19	0.05	0.40	0.07	0.68	0.10
Head Edu	0.06	0.01	0.08	0.02	0.06	0.02	0.03	0.01	0.05	0.01	0.03	0.02	0.02	0.01	0.03	0.01	0.04	0.01
Spouse Edu	0.05	0.01	0.06	0.01	0.06	0.01	0.04	0.01	0.05	0.01	0.06	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Family Size	0.58	0.02	0.70	0.02	0.83	0.03	0.55	0.02	0.67	0.02	0.80	0.02	−0.01	0.01	0.01	0.01	0.04	0.01
IHS Income	2.33	0.10	2.71	0.12	3.22	0.16	1.06	0.07	1.24	0.08	1.50	0.11	1.00	0.04	1.33	0.06	1.71	0.08
Head Emp(HE)	0.21	0.06	0.16	0.07	0.08	0.08	0.05	0.04	0.06	0.05	0.03	0.06	0.06	0.02	0.08	0.03	0.08	0.04
Head Female (HF)	−0.99	0.15	−1.36	0.17	−1.80	0.23	−0.50	0.11	−0.74	0.13	−1.04	0.18	−0.29	0.05	−0.53	0.07	−0.83	0.12
HE×HF	0.35	0.11	0.45	0.13	0.47	0.15	0.24	0.08	0.26	0.09	0.22	0.12	0.15	0.04	0.18	0.06	0.14	0.08
Married	0.12	0.14	0.02	0.16	−0.11	0.20	0.11	0.10	0.01	0.12	−0.06	0.15	0.04	0.05	−0.02	0.07	0.01	0.10
Single	0.73	0.12	0.96	0.15	1.17	0.20	0.33	0.09	0.49	0.11	0.56	0.14	0.20	0.04	0.29	0.06	0.47	0.10
Mort×Home	0.02	0.05	0.06	0.06	0.09	0.07	0.06	0.04	0.07	0.04	0.06	0.05	0.01	0.02	0.01	0.03	0.01	0.03
Non-White	−0.61	0.09	−0.50	0.11	−0.45	0.14	−0.38	0.06	−0.30	0.08	−0.20	0.11	−0.19	0.03	−0.21	0.05	−0.24	0.07
Recession	−0.12	0.03	−0.15	0.03	−0.15	0.04	−0.06	0.02	−0.08	0.03	−0.07	0.03	−0.04	0.01	−0.04	0.02	−0.03	0.02
Northeast	0.61	0.12	0.69	0.15	0.66	0.20	0.41	0.09	0.44	0.11	0.52	0.15	0.04	0.05	0.12	0.07	0.25	0.10
West	0.38	0.12	0.56	0.14	0.66	0.18	0.29	0.08	0.43	0.10	0.62	0.13	0.03	0.04	0.09	0.06	0.15	0.09
South	0.35	0.10	0.48	0.12	0.68	0.15	0.20	0.07	0.26	0.09	0.41	0.11	0.11	0.04	0.17	0.05	0.26	0.08
h	1.76	0.02	1.10	0.01	1.49	0.01	2.31	0.02	1.44	0.01	1.93	0.02	3.64	0.03	2.18	0.02	2.84	0.02
$\sigma_{11}^{\frac{1}{2}}$	1.72	0.08	2.09	0.09	3.47	0.11	1.20	0.05	1.42	0.06	2.53	0.08	0.38	0.04	0.72	0.05	1.63	0.06
$\sigma_{22}^{\frac{1}{2}}$	2.45	0.11	2.81	0.12	4.52	0.15	1.48	0.08	1.55	0.10	2.87	0.12	0.95	0.04	1.40	0.06	2.44	0.08
$\rho_{1,2}$	−0.45	0.05	−0.36	0.05	−0.56	0.03	−0.31	0.06	−0.14	0.08	−0.52	0.04	0.14	0.14	−0.06	0.09	−0.43	0.04

$h = \sigma^{-1}$ in quantile regression possible explanation of such a result is higher involvement of males in sociable activities (Liu et al. 2013).

The variable *Family Size* positively affects TF and FAH expenditures, but not the FAFH expenditure. The positive effect on FAH is understandable as larger families tend to eat more at home and less outside, and is consistent with results reported by Zheng et al. (2019). However, the statistically zero effect on FAFH expenditure is in contrast to those reported in the literature. While some articles find a positive effect of family size (Stewart and Yen 2004; Liu et al. 2013; Zheng et al. 2019), others have reported a negative effect on FAFH expenditure (Redman 1980; Byrne et al. 1996). The quantile regression results once again show heterogeneity in covariate effects. For TF expenditure, the coefficient of *Family Size* is larger at higher quantiles, with the ratio of 80th-to-20th quantile coefficients at 1.43. For the FAH expenditure, the coefficient for *Family Size* is similar in size and sign to those from the TF expenditure equation. Interestingly, *Family Size* has no impact on FAFH expenditure for lower and middle quantiles.

Total family income is perhaps the most decisive variable that steers food expenditure. We use the IHS transformation of family income for reasons mentioned earlier (see also Friedline et al. 2015; Rahman and Vossmeyer 2019). As seen in Table 2, the transformed income variable positively affects expenditures on TF, FAH, and FAFH. The intuition is clear; increase in income translates to increase in food expenditures of all types. This result finds support in several other works such as Redman (1980), Lee and Brown (1986), Nayga (1996), Ziol-Guest et al. (2006), and Liu et al. (2013). Results from quantile regression show considerable heterogeneity in covariate effects with higher quantiles showing a larger impact of income on food expenditure. The ratios of 80th-to-20th quantile coefficients for *IHS Income* in the TF, FAH, and FAFH equations are 1.38, 1.41, and 1.71, respectively.

The next three variables in Table 2 are indicator variable for head's employment (*Head Emp*), indicator variable for female head (*Head Female*), and interaction of the two indicators. Head's employment has a positive effect on TF and FAFH expenditures, but statistically has no effect on FAH expenditure. These findings are similar to those in Aguiar and Hurst (2005), Huang et al. (2016), and Antelo et al. (2017). The indicator for *Head Female* is negative for all categories, which suggests that female-headed families tend to spend less on overall and each category of food. This can be attributed to two factors: females are better at managing family expenditure and an empowered woman better understands the importance of nutritious food and thus reduces FAFH expenditure. The interaction term (*Head Emp* × *Head Female*) in all three regressions is positive, which implies that an employed female head spends more on overall and each category of food purchase. Results from quantile regression, presented in Table 3, once again reveal heterogeneity in the covariate effect of the three indicator variables. Heads's employment positively affects TF expenditure at lower and middle quantiles, but not at upper quantiles. There are no effect on FAH expenditure and a positive effect on FAFH expenditure across quantiles. *Head Female* has a negative effect on overall and each category of food expenditure, and the negative effect increases at upper quantiles. The interaction term shows a positive effect on TF expenditure across quantiles, but a positive effect on FAH and FAFH expenditures only at lower and middle quantiles. Hence, at higher levels of FAH and FAFH expenditures, the employment of female head does not play an important role.

The impact of marital status on food expenditure is examined through the two indicator variables, *Married* and *Single*. The base or omitted category is *Separated*, explained in Sect. 3. As seen from Table 2, the coefficient for *Married* is not statistically different from zero. Therefore, being married has statistically no effect on food expenditure relative to the omitted category, *Separated*. However, being single has a positive effect on overall food expenditure and across categories. Our findings are consistent with results reported by Stewart and Yen (2004) and Liu et al. (2013), but contradictory to those by Byrne et al. (1996) and Zheng et al. (2019). The results from quantile regression reinforce the findings from the mean regression. Across quantiles, being married has no effect on food expenditure as compared to the omitted category. On the other hand, being single has a positive effect on food expenditure and are increasing with quantiles. The ratios of 80th-to-20th quantile coefficients for TF, FAH, and FAFH expenditures are 1.60, 1.70, and 2.35, respectively.

Homeowners having mortgages are resource constrained and have a lower cash flow for a given income. This may negatively affect food expenditure, particularly, FAFH expenditure. To explore this hypothesis, we include an indicator variable for homeowners having mortgages into our regression equations. Results from Table 2 and Table 3 show that families with mortgages have statistically no effect on food expenditure. Our results are opposite to those by Nayga (1996), where he finds that homeowners with mortgages do spend more on food prepared at home and FAFH, but not on prepared foods (e.g., frozen meals and prepared salads). Similarly, Liu et al. (2013) find that homeowners who are married (with and without children) have higher probability of different types of FAFH expenditures (e.g., full-service dining, fast-food and other facilities), but for single-person homeowners, this is true only for full-service dining.

Variations in food expenditure have often been linked to racial disparity. To investigate this conjecture, we include an indicator variable for *Non-White*, keeping *White* as the base or omitted category. Results from mean regression, presented in Table 2, exhibit that *Non-White* tends to have lower expenditure on overall food, as well as FAH and FAFH expenditures. Our findings are consistent with Nayga (1996) who finds that white households are likely to spend more on FAH and FAFH. Similarly, Lee and Brown (1986) report that Non-White are less likely to eat away from home. Our results are also in agreement with findings from other previous works such as Redman (1980), Stewart and Yen (2004), and Liu et al. (2013). Another reason for the negative coefficient, as noted by Byrne et al. (1996), is due to non-availability of ethnic foods at local restaurants. The results from quantile regression, shown in Table 3, largely agree with the finding from mean regression. *Non-White* have lower TF expenditure compared to *White*. Moreover, the impact is larger at lower quantiles and decreases as we move to upper quantiles. For, FAH expenditure, the *Non-White* have lower expenditures only at the lower and middle quantiles, but not at upper quantiles. In contrast, FAFH expenditure for *Non-White* is lower across quantiles and the negative impact increases with increasing quantiles.

Most expenditures, including consumption, typically decline during times of recession. To explore the negative effect on food expenditure, if any, we include an indicator for recession years (2001, 2007, and 2009) into our regression. Results from mean regression show that the coefficient for *Recession* is negative for all types of food expenditure, which implies that expected food expenditure (overall and category wise) declined during the recession years. As reported in Table 2, average TF, FAH, and FAFH expenditures declined by \$257, \$190, and \$75, respectively. Our findings are supported by Griffith et al. (2013), where they report decline in expenditure for food items for British households during and post the Great Recession. Similarly, Antelo et al. (2017) also find that food expenditure for Spanish households declined during the crisis period (i.e., 2008−2014) in Spain. Moving to quantile regression, we find that the quantile results reinforces the findings from mean regression. Both TF and FAH expenditures declined across quantiles during the recession years, and the effect is more or less uniform across the considered quantiles. For FAFH expenditure, we observe a decline only at lower and middle quantiles, but not at the upper quantile. Therefore, families whose expenditure on FAFH is high are not affected by recession years.

Finally, we include regional indicators to examine geographical differences in food expenditure. These differences may be due to varying levels of urbanization, climatic conditions, and diverse food culture. We include indicators for *Northeast*, *West*, and *South* into our regression equations. *Midwest* is used as the base or omitted category. Our regional classification follows the definition of the US Census Bureau. Results from mean regressions (see Table 2) reveal that an average family living in *South* (relative to *Midwest*) has higher TF, FAH, and FAFH expenditures. However, for families living in the *Northeast* and *West*, the average expenditure is more on TF and FAH but not on FAFH. Other studies, such as Lee and Brown (1986), Nayga (1996), Byrne et al. (1996), and Liu et al. (2013), also find disparity in regional food expenditures. Moving to results from quantile regression (see Table 3), we see that for TF and FAH expenditures, all the quantile coefficients for *Northeast*, *West*, and *South* are positive and increase with quantiles. This suggests that families living in the three regions have higher quantile expenditures (compared to those living in *Midwest*) and the differential impact increases at higher quantiles. For FAFH expenditure, only *South* and *Northeast* (at the upper quantiles only) have a positive effect on FAFH expenditure.

In summary, the results from quantile regression reveal considerable heterogeneity in covariate effects which cannot be uncovered from mean regression. The additional information from quantile regression may be useful for policy making in the government or business, such as aiming sections of the population for welfare schemes or running campaigns to promote business.

4.1 Heterogeneity bias

Unobserved heterogeneity is a large component of food expenditure, and we control for this in our (mean and quantile) regression models with individual-specific parameters in the intercept and income. To demonstrate the heterogeneity bias and poorer model fit that can occur, we estimate the quantile models without including the individual-specific effects (i.e., without including the conditional dependence

Table 4 Posterior mean (MEAN) and standard deviation (STD) of the parameters from longitudinal quantile regression without random-effects

	TF						FAH						FAFH					
	20th		50th		80th		20th		50th		80th		20th		50th		80th	
	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD	MEAN	STD
Intercept	−0.80	0.39	−1.26	0.52	−1.70	0.66	−1.77	0.29	−2.75	0.36	−3.46	0.45	0.86	0.16	1.02	0.25	0.35	0.39
log(Head Age)	0.46	0.09	0.70	0.12	1.12	0.15	0.58	0.06	0.93	0.08	1.24	0.10	−0.23	0.04	−0.16	0.05	0.18	0.08
Head Edu	0.01	0.01	0.03	0.01	0.04	0.01	−0.01	0.01	0.02	0.01	0.04	0.01	0.01	0.00	0.02	0.01	0.02	0.01
Spouse Edu	0.05	0.01	0.01	0.01	−0.01	0.01	0.05	0.01	0.04	0.01	0.01	0.01	−0.01	0.01	−0.01	0.00	−0.02	0.01
Family Size	0.52	0.02	0.80	0.02	1.14	0.03	0.54	0.01	0.78	0.02	1.12	0.02	−0.02	0.01	−0.01	0.01	0.06	0.01
IHS Income	2.47	0.06	3.71	0.08	5.22	0.09	1.01	0.04	1.43	0.06	2.09	0.07	0.99	0.03	1.81	0.04	3.11	0.05
Head Emp(HE)	0.23	0.06	0.11	0.07	−0.12	0.09	0.07	0.04	0.15	0.05	0.05	0.07	0.09	0.02	0.06	0.03	−0.01	0.05
Head Female (HF)	−0.55	0.08	−0.81	0.11	−1.41	0.15	−0.18	0.06	−0.25	0.08	−0.43	0.11	−0.06	0.03	−0.40	0.05	−0.86	0.08
HE×HF	0.42	0.09	0.51	0.12	0.68	0.15	0.28	0.07	0.21	0.09	0.19	0.11	0.01	0.03	0.18	0.05	0.27	0.08
Married	−0.01	0.12	0.23	0.14	0.01	0.16	−0.04	0.08	0.09	0.10	0.45	0.11	0.03	0.04	−0.01	0.07	−0.18	0.09
Single	0.18	0.06	0.36	0.09	0.57	0.12	0.06	0.05	0.12	0.06	0.15	0.08	0.10	0.03	0.16	0.04	0.38	0.06
Mort×Home	0.11	0.04	0.02	0.05	0.03	0.07	0.14	0.03	0.03	0.04	0.01	0.05	0.01	0.02	−0.06	0.02	0.02	0.04
Non-White	−0.67	0.04	−0.80	0.06	−0.87	0.08	−0.48	0.03	−0.51	0.04	−0.50	0.05	−0.11	0.02	−0.22	0.03	−0.41	0.04
Recession	−0.23	0.04	−0.36	0.05	−0.51	0.06	−0.17	0.03	−0.20	0.03	−0.51	0.04	−0.05	0.02	−0.07	0.02	−0.08	0.03
Northeast	0.51	0.06	0.88	0.08	1.41	0.10	0.52	0.04	0.61	0.05	0.97	0.07	−0.01	0.03	0.12	0.04	0.32	0.05
West	0.36	0.05	0.46	0.07	0.79	0.09	0.40	0.04	0.49	0.05	0.70	0.06	−0.01	0.02	0.03	0.03	0.06	0.05
South	0.38	0.05	0.49	0.06	0.80	0.08	0.23	0.03	0.25	0.04	0.49	0.05	0.10	0.02	0.23	0.03	0.30	0.04
h	1.26	0.01	0.76	0.01	0.90	0.01	1.71	0.01	1.04	0.01	1.24	0.01	2.70	0.02	1.48	0.01	1.62	0.01

$h = \sigma^{-1}$ in quantile regression

Table 5 Model comparison between the longitudinal quantile regression with random-effects (with RE) and without random-effects (w/o RE). The log-likelihood (log-L), conditional Akaike Information Criterion (cAIC) and conditional Bayesian Information Criterion (cBIC) are evaluated at the posterior mean of the parameters

	20th quantile		50th quantile		80th quantile	
	With RE	W/o RE	With RE	W/o RE	With RE	W/o RE
TF Expenditure						
log-L	−38282	−45188	−38701	−46268	−40996	−51031
cAIC	76601	90411	77439	92573	82029	102099
cBIC	76802	90551	77640	92713	82230	102239
FAH Expenditure						
log-L	−33577	−39932	−34054	−40773	−36617	−45538
cAIC	67189	79901	68143	81582	73271	91111
cBIC	67390	80041	68344	81722	73472	91251
FAFH Expenditure						
log-L	−25753	−31996	−26797	−34630	−29883	−40896
cAIC	51542	64028	53630	69295	59802	81828
cBIC	51743	64168	53832	69435	60003	81968

between observations across time for the same family unit). This model can be estimated as a special case of Algorithm 2, by eliminating Step 1(a) and Step (3), and removing (α_i, Σ^{-1}) from the conditional posteriors of the remaining parameters.

The results from the longitudinal quantile models without the individual-specific effects are presented in Table 4 and they differ widely compared to those of Table 3, which presents the results from longitudinal quantile regression with individual-specific effects. For example, the coefficients for *Head Age*, *IHS Income*, *Head Female*, and *Recession* are noticeably different in the two models across quantiles and types of food expenditure. Again, there are variables whose coefficients either become statistically equivalent to or different from zero when the individual-specific parameters are excluded. In the former category, we have the coefficient for *Spouse Edu* at middle and upper quantiles for total food expenditure. In the latter category, we have the coefficient for homeowners with mortgages ($Mort \times Home$) at lower quantiles for expenditures on total food and food at home.

To highlight the importance of modeling the individual-specific effects (or random-effects), we compare model fitting at the considered quantiles using the conditional log-likelihood, conditional Akaike Information Criterion (cAIC), and conditional Bayesian Information Criterion (cBIC). The calculations of cAIC and cBIC are proposed and explained in Greven and Kneib (2010) and Delattre et al. (2014), respectively. These model comparison measures are presented in Table 5. The table clearly shows that across quantiles, the value of the conditional log-likelihood is higher and those of cAIC and cBIC are lower for each longitudinal quantile regression when individual-specific effects are included. Consequently, there is a strong evidence for modeling unobserved heterogeneity and ignoring it can lead to poor model fitting.

5 Conclusion

This article studies the relationship between different types of food expenditures (total food, food at home, and food away from home) and a host of economic, geographic, and demographic factors using data from the Panel Study of Income Dynamics for the period 2001−2015. Food expenditures are typically right skewed and, thus, covariate effects are likely to be heterogeneous across the conditional distribution of the response variable. Besides, unobserved heterogeneity is a large component of food expenditure. To explore these considerations, we study food expenditure within a longitudinal quantile framework that models dependence between the observations across time for the same family units. Results point to several important aspects including the presence of heterogeneity in the covariate effects. For example, we find that there are notable differences in the food expenditure behavior (of all types) between male and female headed households, expenditures on food away from home by employed female heads are heterogeneous across quantiles, and food expenditures (of all types) decrease during times of economic crisis and vary with quantiles. Besides, the paper provides strong empirical evidence that not considering unobserved heterogeneity can lead to heterogeneity bias and poor model fitting.

While our paper emphasizes the modeling of heterogeneity in food expenditure, the findings reported also provide greater insights on total food expenditure and expenditures on food at home and food away from home, which may be of special interest to policy makers in the health and business sectors. For example, we find that spouse education has a positive effect across the distribution of food at home expenditure. Therefore, policy makers may provide higher incentives to female education to achieve better health outcomes in the country. Similarly, we find that being single or employed female heads have a positive effect on the distribution of food away from home expenditure. Consequently, restaurant and fast food chains may run campaigns targeting these specific groups to increase their sales. The above discussion and other findings reported in the paper may be utilized to better formulate policies and business decisions.

Funding Not Applicable.

Availability of data and material The data used in publicly available from the Panel Study of Income Dynamics, University of Michigan, Ann Arbor.

Code availability The codes are available on request.

Compliance with ethical standards

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Aguiar, M., & Hurst, E. (2005). Consumption versus expenditure. *Journal of Political Economy*, *113*(5), 919–948.

Antelo, M., Magdalena, P., & Reboredo, J. C. (2017). Economic crisis and the unemployment effect on household food expenditure: The case of Spain. *Food Policy*, *69*, 11–24.

Beatty, T. K. M., & Tuttle, C. J. (2014). Expenditure response to increases in in-kind transfers: Evidence from the supplemental nutrition assistance program. *American Journal of Agricultural Economics*, *97*(2), 390–404.

Binkley, J. K. (2008). Calorie and gram differences between meals at fast food and table service restaurants. *Applied Economic Perspectives and Policy*, *30*(4), 750–763.

Bresson, G., Lacroix, G., & Rahman, M. A. (2020). Bayesian panel quantile regression for binary outcomes with correlated random effects: An application on crime recidivism in Canada. *Empirical Economics (forthcoming)*,. https://doi.org/10.1007/s00181-020-01893-5.

Burney, S. (2018). In-kind benefits and household behavior: The impact of SNAP on food-away-from-home consumption. *Food Policy*, *75*, 134–146.

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 Agricultural Economics*, *78*(3), 614–627.

Cai, Y., Alviola, P. A., Nayga, R. M., & Wu, X. (2008). The effect of food-away-from-home and food-at-home expenditures on obesity rates: A state level analysis. *Journal of Agricultural and Applied Economics*, *40*(2), 507–521.

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge: Cambridge University Press.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*(3), 167–174.

Chib, S., & Carlin, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, *9*(1), 17–26.

Darian, J. C., & Klein, S. W. (1989). Food expenditure patterns of working-wife families: Meals prepared away from home versus convenience foods. *Journal of Consumer Policy*, *12*(2), 139–164.

Davis, G. C. (2014). Food at home production and consumption: Implications for nutrition quality and policy. *Review of Economics of the Household*, *12*, 565–588.

Delattre, M., Lavielle, M., & Poursat, M. A. (2014). A note on BIC in mixed-effects models. *Electronic Journal of Statistics*, *8*(1), 456–475.

Devroye, L. (2014). Random variate generation for the generalized inverse Gaussian distribution. *Statistics and Computing*, *24*(2), 239–246.

Drichoutis, A. C., Nayga, R. M., & Lazaridis, P. (2012). Food away from home expenditures and obesity among older Europeans: Are there gender differences? *Empirical Economics*, *42*(3), 1051–1078.

Friedline, T., Masa, R. D., & Chowa, G. A. (2015). Transforming wealth: Using the inverse hyperbolic sine (ihs) and splines to predict youth's math achievement. *Social Science Research*, *49*, 264–287.

Gelber, A. M., & Mitchell, J. W. (2012). Taxes and time allocation: Evidence from single women and men. *The Review of Economic Studies*, *79*(3), 863–897.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741.

Greenberg, E. (2012). *Introduction to Bayesian econometrics*. New York: Cambridge University Press.

Greene, W. H. (2017). *Econometric analysis* (8th ed.). New York: Prentice Hall.

Greven, S., & Kneib, T. (2010). On the behaviour of marginal and conditional AIC in linear mixed models. *Biometrika*, *97*(4), 773–789.

Griffith, R., Connell, M. O., & Smith, K. (2013). *Food expenditure and nutritional quality over the Great Recession*. London: Tech. Rep., Institute for Fiscal Studies.

Handa, S. (1996). Expenditure behavior and children's welfare: An analysis of female headed households in Jamaica. *Journal of Development Economics*, *50*(1), 165–187.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Hoynes, H. W., & Schanzenbach, D. W. (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program. *American Economic Journal: Applied Economics*, *1*(4), 109–139.

Hsiao, C. (2014). *Analysis of panel data (econometric society monograph)*. New York: Cambridge University Press.

Huang, J., Kim, Y., & Birkenmaier, J. (2016). Unemployment and household food hardship in the economic recession. *Public Health Nutrition*, *19*(3), 511–519.
Kinsey, J. (1983). Working wives and the marginal propensity to consume food away from home. *American Journal of Agricultural Economics*, *65*(1), 10–19.
Kohara, M., & Kamiya, Y. (2016). Maternal employment and food produced at home: Evidence from Japanese data. *Review of Economics of the Household*, *14*, 417–442.
Kozumi, H., & Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation*, *81*(11), 1565–1578.
Lanfranco, B. A., Ames, G. C., & Huang, C. L. (2002). Food expenditure patterns of the Hispanic population in the United States. *Agribusiness*, *18*(2), 197–211.
Lee, J. Y., & Brown, M. G. (1986). Food expenditures at home and away from home in the United States—a switching regression analysis. *Review of Economics and Statistics*, *68*(1), 142–147.
Liu, M., Kasteridis, P., & Yen, S. T. (2013). Who are consuming food away from home and where? Results from the consumer expenditure surveys. *European Review of Agricultural Economics*, *40*(1), 191–213.
Liu, H., Wahl, T. I., Seale, J. L., & Bai, J. (2015). Household composition, income, and food-away-from-home expenditure in urban China. *Food Policy*, *51*, 97–103.
Luo, Y., Lian, H., & Tian, M. (2012). Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, *82*(11), 1635–1649.
Mancino, L., Todd, J., & Lin, B. H. (2009). Separating what we eat from where: Measuring the effect of food away from home on diet quality. *Food Policy*, *34*(6), 557–562.
Mian, A., Rao, K., & Sufi, A. (2013). Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics*, *128*(4), 1687–1726.
Nayga, R. M. (1996). Wife's labor force participation and family expenditures for prepared food, food prepared at home, and food away from home. *Agricultural and Resource Economics Review*, *25*(2), 179–186.
Ojha, M., Rahman, M.A. (2020). Do online courses provide an equal educational value compared to in-person classroom teaching? Evidence from US survey data using quantile regression, https://arxiv.org/pdf/2007.06994.pdf
Rahman, M. A. (2016). Bayesian quantile regression for ordinal models. *Bayesian Analysis*, *11*(1), 1–24. https://doi.org/10.1214/15-BA939.
Rahman, M. A., & Karnawat, S. (2019). Flexible Bayesian quantile regression in ordinal models. *Advances in Econometrics*, *40B*, 211–251.
Rahman, M. A., & Vossmeyer, A. (2019). Estimation and applications of quantile regression for binary longitudinal data. *Advances in Econometrics*, *40B*, 157–191.
Redman, B. J. (1980). The impact of women's time allocation on expenditure for meals away from home and prepared foods. *American Journal of Agricultural Economics*, *62*(2), 234–237.
Stewart, H., & Yen, S. T. (2004). Changing households characteristics and the away-from-home food market: A censored equation system approach. *Food Policy*, *29*, 643–658.
Stewart, H., Blisard, N., Jolliffe, D., & Bhuyan, S. (2005). The demand for food away from home: Do other preferences compete with our desire to eat healthfully? *Journal of Agricultural and Resource Economics*, *30*(3), 520–536.
Wilde, P. E., Troy, L. M., & Rogers, B. L. (2009). Food stamps and food spending: An Engel function approach. *American Journal of Agricultural Economics*, *91*(2), 416–430.
Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data* (2nd ed.). Cambridge: MIT Press.
Yen, S. T. (1993). Working wives and food away from home: The Box-Cox double hurdle model. *American Journal of Agricultural Economics*, *75*(4), 884–895.
Yu, K., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics and Probability Letters*, *54*(4), 437–447.
Zheng, Y., Dong, D., Burney, S., Kaiser, H.M. (2019). Eat at home or away from home? The role of grocery and restaurant food sales taxes. Journal of Agricultural and Resource Economics pp 98–116
Ziol-Guest, K. M., DeLeire, T., & Kalil, A. (2006). The allocation of food expenditure in married-and single-parent families. *The Journal of Consumer Affairs*, *40*(2), 347–371.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Arjun Gupta[1] · **Soudeh Mirghasemi**[2] · **Mohammad Arshad Rahman**[1]

Arjun Gupta
arjung@iitk.ac.in; arjungupta3276@gmail.com

Soudeh Mirghasemi
Soudeh.Mirghasemi@hofstra.edu

1 Department of Economic Sciences, Indian Institute of Technology, Kanpur, Kanpur, India

2 Department of Economics, Hofstra University, Hempstead, USA