Sparsity-Based Estimation of a Panel Quantile Count Data Model with Applications to Big Data*

Matthew Harding† and Carlos Lamarche‡

December 31, 2014

Abstract

In this paper we introduce a panel quantile estimator for count data with individual heterogeneity, by constructing continuous variables whose conditional quantiles have a one-to-one relationship with the conditional count response variable. The new method is needed as a result of the increased availability of Big Data, which allows us to track event counts at the individual level for a large number of activities from webclicks and retweets to store visits and purchases. At the same time, the presence of many different subpopulations in a large dataset requires us to pay close attention to individual heterogeneity. In this paper, we propose a penalized quantile regression estimator with fixed effects and investigate the conditions under which the slope parameter estimator is asymptotically Gaussian. We investigate solutions to the computational challenges resulting from the need to estimate tens of thousands of parameters in a Big Data setting and caution against penalizing in models with substantial zero inflation and endogenous covariates by using a series of Monte Carlo simulations. We present an empirical application to individual trip counts to the store based on a large panel of food purchase transactions.

JEL: C21, C23, C25, C55.
Keywords: Big Data; Quantile regression; Penalized Estimation; Count Data; Individual Effects.

*The authors would like to thank Shif Gurmu for comments on a previous draft as well as seminar participants at the University of Tennessee and conference participants at the 84th Annual Meeting of the Southern Economic Association.

†Sanford School of Public Policy, Duke University, Durham NC 27708; Phone: (919) 613 9300; Email: matthew.harding@duke.edu

‡Department of Economics, University of Kentucky, 335A Gatton College of Business and Economics, Lexington, KY 40506-0034; Phone: (859) 257 3371; Email: clamarche@uky.edu

1. Introduction

Event counts, or number of times an individual event occurs, are standard in empirical microeconomics. While research on count data has a long and rich tradition with numerous theoretical and applied studies across statistically focused disciplines (Cameron and Trivedi, 2013), the increased availability of Big Data opens up new challenges and possibilities. The key advantage of panel count data is that it offers the possibility of tracking a large number of counts at the individual level that are observed over a long period of time, leading to richer models that require more general forms of heterogeneity. The use of massive count data sets is naturally associated with the proposal of new computational frameworks (see, e.g., Taddy, 2013 and Taddy, 2014). This paper makes two contributions to the development of quantile regression methods for panel count data. First, we propose a semiparametric panel model for count data with individual heterogeneity and introduce penalized regression estimators. Second, an essential algorithm for Big Data settings and a simple inferential approach that takes advantage of the sparsity of the design are proposed. The modeling approach is then applied to an empirical application to search and choice behavior using detailed individual transaction data from grocery stores.

Numerous papers have proposed parametric and moment-based approaches for longitudinal count data including Hausman, Hall and Griliches (1984), Wooldridge (1999), Blundell, Griffith and Windmeijer (2002), and more recently, flexible models for unobserved heterogeneity are proposed by Gurmu, Rilstone and Stern (1999) and Burda, Harding and Hausman (2012). Quantile regression, originally introduced by Koenker and Bassett (1978), and quantile for counts, developed by Machado and Santos Silva (2005), are becoming increasingly popular in applied microeconomic research and offer a semiparametric alternative approach to standard methods. The extension of quantile methods to panel data is investigated in a series of papers by Koenker (2004), Lamarche (2010), Galvao (2011) and Kato, Galvao and Montes-Rojas (2012). Galvao, Lamarche, and Lima (2013) study the estimation of a censored panel data model. Moreover, alternative approaches and models are studied in Chernozhukov, Fernández-Val, Hahn, and Newey (2013) and Harding and Lamarche (2014), among others. These papers deal exclusively with the case of continuous response variables. This highlights the need for quantile panel methods suitable for the analysis of count data.

This paper introduces an innovative quantile regression approach to model heterogeneity in panel count data, while providing flexibility relative to standard parametric models. Standard approaches were developed with a continuous outcome variable in mind and suffer from a number of limitations when applied to count data. For example, Koenker and Bassett (1978) develop a quantile regression estimator for the τ-th quantile function conditional on a vector of independent variables $\boldsymbol{x} \in \mathcal{X} \subset R^p$, $Q_Y(\tau|\boldsymbol{x})$, that is consistent and asymptotically normally distributed. Inference procedures are also develop under sufficient conditions including that the conditional probability density function is continuous at the conditional quantile and have continuous derivatives. These conditions are violated if the response variable Y is a count variable. In a different context, valid asymptotic inference is possible by adopting a specific form of jittering proposed by Stevens (1950) and later extended by Machado and Santos Silva (2005). This paper extends the approach of Machado and Santos Silva (2005) to panel data by constructing continuous variables whose conditional quantiles have a one-to-one relationship with the count response variable conditional on observables and unobservables.

The main challenge in the estimation of quantile counts for panel data is the combination of non-smoothness of the objective function and individual specific heterogeneity potentially correlated with the independent variables. The necessary smoothness is achieved by adding a uniform random variable to the count variable but omitting unobserved individual heterogeneity can lead to lack of identification of the conditional quantile function of the count variable. Our investigation shows that existing cross-sectional approaches that also employ jittered samples can be severely biased in small samples. We propose an ℓ_1-penalized quantile regression estimators for panel count data explicitly allowing for individual heterogeneity and we investigate the conditions under which the slope parameter estimator is asymptotically Gaussian. The proposed estimator shows desirable small sample properties in a class of count models with random and fixed individual effects.

We present an application to modeling consumer trip counts. In both real and virtual retail environment consumer choice begins with a visit to the store. Modeling the number of store visits is an important measure of consumer search and has important implications in a number of areas from marketing to urban planning and even public health. At a basic level, trip counts are determined by observable socio-demographics and unobserved attributes such as the opportunity cost of time. More recently, and in light of the Great Recession, economists have also asked what role local economic environments play in driving consumption behavior. In our example, we extend the analysis of trip counts by also studying the effect of house prices and unemployment on shopping trip behavior. It has been argued that both housing prices and unemployment may affect consumption (see, e.g., Campbell and Cocco, 2007, Attanasio et al., 2009), but the analysis is typically based on synthetic panels, which are constructed by the methodology introduced in Browning et al. (1985) and Deaton (1985) to obtain a panel of time series of cross-sectional data from a specific population.

In contrast with existing studies, we can track individual households purchases per month over a period of five years, avoiding potential biases arising from the use of pseudo-panels. Using data from the National Consumer Panel, we estimate a panel quantile count model for the number of shopping trips and number of shopping days. We find that the proposed method gives different shopping profiles relative to other quantile methods and the classical conditional Poisson estimator with fixed effects. While we find evidence of both demographic gradients and a wealth channel impacting search behavior at the upper quantiles, the results are also surprising in that they reveal a much larger role for unobserved preference heterogeneity in explaining the observed variation than we might have a priori expected. This suggests that marketing strategies based on other strategies than demographic profiling are likely to be more successful in developing predictive analytics of consumer behavior.

The next section introduces the model for count data and the corresponding estimator. In Section 3, we investigate the small sample behavior of the proposed approach in relation to other methods. Section 4 demonstrates how the estimator can be used in an empirical application using a large number of individual transactions. Section 5 offers a few concluding remarks.

2. Panel quantile for count data

Let $\{(y_{it}, \boldsymbol{x}'_{it}, \alpha_i)\}$ be a sequence of identically and independently distributed (i.i.d.) random variables for subject i at time t with $i = 1, \ldots, N$ and $t = 1, \ldots, T_i$. The variable y_{it} denotes a discrete count variable with support on the set of non-negative integers, $\mathbb{N}_0$, and has conditional probability equal to $P(y_{it} = k|\boldsymbol{x}_{it}, \alpha_i)$ for $k = \{0, 1, ..., y\}$. The vector $\boldsymbol{x}_{it} = (\boldsymbol{x}'_{1,it}, \boldsymbol{d}'_{1,it}, \boldsymbol{x}'_{2,i}, \boldsymbol{d}'_{2,i})' \in \mathcal{R}^p$ denotes a p-dimensional vector of independent variables, where $\boldsymbol{x}_{1,it}$ is a p_1-dimensional vector of time-varying continuous variables, $\boldsymbol{d}_{1,it}$ is a k_1-dimensional vector of time-varying discrete variables, $\boldsymbol{x}_{2,i}$ is a p_2-dimensional vector of time invariant continuous variables, and $\boldsymbol{d}_{2,i}$ is a k_2-dimensional vector of time invariant discrete variables. The vector $\boldsymbol{x}_{it}$ also includes an intercept. The variable α_i measures latent heterogeneity potentially dependent on the columns of the vector of independent variables, $\boldsymbol{x}_{it}$. Moreover, we denote by $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)$ the τ-th quantile of the conditional distribution of y_{it} given $\boldsymbol{x}_{it}$ and α_i, where τ is a given quantile in the interval $(0, 1)$. The quantile function is defined as $Q_Y(\tau|\boldsymbol{x}, \alpha) = \inf\{y : \Pr(Y \leq y|\boldsymbol{x}, \alpha) \geq \tau\}$.

2.1. Model and assumptions

This section introduces the model and assumptions associated with our panel quantile regression model for count data. Although the number of time series observations can vary with i as in the empirical application of the method described in Section 4, for simplicity in exposition, the remaining part of this section focuses in the case of $T_i = T$ for all i. Consider the following specification:

$$h(z_{it}, \tau) = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + v_{it}, \tag{2.1}$$

$$z_{it} = y_{it} + u_{it}, \tag{2.2}$$

where $h(\cdot)$ is a known monotone transformation, τ is a quantile of the conditional distribution of the count variable, v_{it} is an error term, z_{it} is a continuous variable, and u_{it} is a uniform random variable in $[0, 1)$. Equation (2.1) states that a known monotone function achieves linearity in terms of the parameter of interest $\boldsymbol{\beta}$ and unobserved heterogeneity α_i. If $h(\cdot)$ is equal to the logarithmic transformation, this model contains the case where latent individual heterogeneity $\exp(\alpha_i)$ enters multiplicatively in the equation. In equation (2.2), the continuous variable z_{it} is constructed by adding to the count variable, y_{it}, a random variable u_{it} which is uniformly distributed in the interval $[0, 1)$. Work by Stevens (1950) and Pearson (1950) lead to a nowadays conventional idea expressed in equation (2.2) that has a long tradition in Statistics since the discussion by Anscombe to the Royal Statistical Society (1948). We consider the following assumption:

ASSUMPTION 1. *The variable u_{it} is distributed as uniform in $[0, 1)$ and is independent of the count variable y_{it} and $\{(\boldsymbol{x}'_{it}, \alpha_i, v_{it})\}$ $\forall (i, t)$. The variables $\{v_{it}\}$ are i.i.d. for each i and all $t \geq 1$.*

Assumption 1 is standard in the literature on quantile for counts (Machado and Santos Silva (2005) and Hong and He (2010)). The use of a uniform distribution is for convenience, because it allows algebraic and computational simplifications. This assumption can be relaxed by considering any continuous distribution with support on $[0, 1)$ and density function bounded away from 0 and ∞. Note that a different choice of the jittering distribution leads to a different transformation of z without changing the distribution of $h(z)$. The second part of Assumption 1 makes easier to find a consistent and asymptotically normal estimator for $\boldsymbol{\beta}$ (see, e.g., Galvao, Lamarche and Lima (2013)). Note that the distribution of the v_{it}'s is left unspecified.

As discussed before, the assumption on the distribution of the random variable u_{it} can be relaxed but this formulation leads to simplifications in the derivation of conditional quantile functions. For instance, the conditional distribution of z_{it} conditional on $\boldsymbol{x}_{it}$ and α_i can be written as:

$$
\begin{aligned}
P(z_{it} < z|\boldsymbol{x}_{it}, \alpha_i) &= P(y_{it} < y|\boldsymbol{x}_{it}, \alpha_i) + P(u_{it} < u|\boldsymbol{x}_{it}, \alpha_i)P(y_{it} = y|\boldsymbol{x}_{it}, \alpha_i) && (2.3)\\
&= P(y_{it} < y|\boldsymbol{x}_{it}, \alpha_i) + uP(y_{it} = y|\boldsymbol{x}_{it}, \alpha_i) && (2.4)\\
&= \sum_{k=0}^{y-1} P(y_{it} = k|\boldsymbol{x}_{it}, \alpha_i) + uP(y_{it} = y|\boldsymbol{x}_{it}, \alpha_i). && (2.5)
\end{aligned}
$$

Denoting $y \equiv Q_Y(\tau|\boldsymbol{x}, \alpha)$ and $u \equiv Q_Z(\tau|\boldsymbol{x}, \alpha) - Q_Y(\tau|\boldsymbol{x}, \alpha)$, we obtain,

$$
\tau = \sum_{k=0}^{y-1} P(y_{it} = k|\boldsymbol{x}_{it}, \alpha_i) + (Q_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) - Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i))P(y_{it} = Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)|\boldsymbol{x}_{it}, \alpha_i),
$$

or alternatively, the conditional quantile function of the continuous variable can be expressed as,

$$
Q_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) + \frac{\tau - \sum_{k=0}^{Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)-1} P(y_{it} = k|\boldsymbol{x}_{it}, \alpha_i)}{P(y_{it} = Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)|\boldsymbol{x}_{it}, \alpha_i)}, \tag{2.6}
$$

which establishes a relationship between the conditional quantile function of the continuous random variable and the conditional quantile function of the count variable. Equation (2.6) is a panel data version of equation (1) in Machado and Santos Silva (2005) and a general version of the model described in Winkelmann (2008, p. 200).

The quantiles of the count variable are identified by the one-to-one relationship with the quantiles of the continuous variable. Note that because $u_{it} \in [0, 1)$, we have that,

$$
y_{it} - 1 \leq y_{it} - 1 + u_{it} < y_{it}, \tag{2.7}
$$

or, by equation (2.2),

$$
y_{it} - 1 \leq z_{it} - 1 < y_{it}. \tag{2.8}
$$

Because the conditional and unconditional quantile functions $Q(\tau) := F^{-1}(\tau)$ are non-decreasing, equation (2.8) implies $Q_{Y_{it}}(\tau) - 1 \leq Q_{Z_{it}}(\tau) - 1 < Q_{Y_{it}}(\tau)$ and

$$
Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) - 1 \leq Q_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) - 1 < Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i). \tag{2.9}
$$

Therefore, in our panel quantile count model, we have,

$$
Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = \lceil Q_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) - 1 \rceil, \tag{2.10}
$$

where $\lceil \cdot \rceil$ is the ceiling function defined as $\lceil y \rceil = \min\{x \in \mathbb{Z}|x \geq y\}$ where $\mathbb{Z}$ denotes the set of integers and y is a real number.

Remark 1. It is immediately apparent that the omission of unobserved heterogeneity in the conditional quantile function of the continuous variable z_{it}, say e.g. $Q_{Z_{it}}(\tau|\boldsymbol{x}_{it})$, might lead to failure of identification of the conditional quantile function of a panel count variable Y_{it} which is defined in equation (2.10).

The advantage of panel count data is that provides a framework for addressing unobserved heterogeneity α_i in the conditional quantile of the count variable $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)$. In cross-sectional models, it is known than the presence of individual heterogeneity distributed independently of the covariates implies over-dispersion, and therefore, it is possible to estimate consistently regression parameters even if the model is misspecified. Practitioners can apply for instance Poisson regression using robust standard errors. In this paper, we consistently estimate $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)$ without requiring the independence assumption.

Naturally $\beta_j(\tau) = 0$ implies no dependence between the j-th covariate $x_{j,it} \in \mathcal{R}$ and the conditional quantile function $Q_Y(\tau|\cdot)$ in equation (2.10). Note however that it is possible that $\beta_j(\tau) \neq 0$ and the conditional quantile $Q_Y(\tau|\cdot)$ is not influenced by changes in $x_{j,it}$. Moreover, it is possible that $\alpha_{i0} = 0$ and $\alpha_{h0} \neq 0$ and yet the conditional quantile function $Q_{Y_{it}}(\tau|\cdot) = Q_{Y_{ht}}(\tau|\cdot)$ at the same quantile τ when $x = x_{it} = x_{ht}$. On the other hand, if $\alpha_{i0} = 0$ and $\alpha_{h0} \neq 0$, the conditional quantile $Q_{Z_{it}}(\tau|\cdot)$ is expected to be different than $Q_{Z_{ht}}(\tau|\cdot)$ even when $x = x_{it} = x_{ht}$. This case is similar to the analysis of the effect of a change in $x_{j,it}$ on $Q_{Z_{it}}(\tau|\cdot)$ if $\beta_j(\tau) \neq 0$.

The following assumption is also considered:

ASSUMPTION 2. *The monotone transformation* $h(\cdot)$ *is known and the conditional quantile function of the continuous dependent variable is equal to,*

$$Q_{h(Z_{it},\tau)}(\tau|\boldsymbol{x}_{it}, \alpha_i) = \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau) + \alpha_i(\tau), \tag{2.11}$$

where $\tau \in (0,1)$ *and* $\alpha_i(\tau)$ *is a scalar individual effect for each* i. *Moreover, at least one* $\beta_j(\tau)$ *associated with the vector of continuous variables* $(\boldsymbol{x}'_{1,it}, \boldsymbol{x}'_{2,i}) \in \mathcal{X}^{p_1} \times \mathcal{X}^{p_2} \subset \mathbb{R}^{p_1+p_2}$ *is non-zero and there is* $1 \leq p_1 + p_2 \leq p - 1$ *satisfying that* $P((\boldsymbol{x}'_{1,it}, \boldsymbol{x}'_{2,i}) \in C) = 0$ *for any countable subset* $C \in \mathbb{R}^{p_1+p_2}$.

Assumption 2 restricts the panel conditional quantile functions to be single-index models of the form: $Q_Z(\tau|\boldsymbol{x}, \alpha) = h^{-1}(\boldsymbol{x}'\boldsymbol{\beta}(\tau) + \alpha(\tau), \tau)$, where $h^{-1}(\cdot)$ denotes the inverse of the monotone transformation $h(z_{it}, \tau)$. The quantile of the continuous variable z depends on an individual specific effect $\alpha(\tau)$ that is indexed by quantiles as in the panel quantile literature (e.g., Harding and

Lamarche 2009, Galvao et al. (2013), among others). In case that $\alpha_i(\tau) = \alpha_i$ for all τ's, the quantile function (2.11) can be written as $\boldsymbol{x}'\boldsymbol{\beta}(\tau)+\alpha$, where individual heterogeneity is a location shift in the sense of Koenker (2004). We discuss estimation of this alternative model in the following Section 2.2. The second part of Assumption 2 is important for the natural requirement that $\beta_j(\tau) \neq 0$ is equivalent to the conditional quantile $Q_Y(\tau|\cdot)$ depending on the j-th covariate $x_{j,it} \in \mathcal{R}$. Under this assumption, there are values of the independent variables, which includes continuous and discrete covariates, for which $\beta_j(\tau) \neq 0$ implies that $Q_Y(\tau|\cdot)$ depends on $x_{j,it}$. As in Manski (1985) and Machado and Santos Silva (2005), we require that at least one regressor is continuous and a large support assumption for the continuous variable holds.

2.2. **The proposed methods**

This paper investigates the estimation of panel quantile regression models for count data. The estimation of the slope parameter, the conditional quantile function of the panel count variable, and the marginal effect of covariates on the conditional quantile of the response variable can be estimated considering the following steps:

Step 1: Let $\boldsymbol{\alpha} := (\alpha_1, ..., \alpha_N)$. For $z_{it}^{(l)} = y_{it}+u_{it}^{(l)}$ with $l = 1, \ldots, m$, estimate $\boldsymbol{\theta}(\tau) = (\boldsymbol{\beta}(\tau)', \boldsymbol{\alpha}(\tau)')'$ by the following average jittering penalized estimator for a model with fixed effects,

$$\hat{\boldsymbol{\theta}}(\tau, \lambda) = \frac{1}{m}\sum_{l=1}^{m} \tilde{\boldsymbol{\theta}}_{(l)}(\tau, \lambda), \tag{2.12}$$

where λ is a tuning parameter and,

$$\tilde{\boldsymbol{\theta}}_{(l)}(\tau, \lambda) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau(h(z_{it}^{(l)}, \tau) - \boldsymbol{x}_{it}'\boldsymbol{\beta} - \alpha_i) + \lambda \sum_{i=1}^{N} \rho_\tau(\alpha_i), \tag{2.13}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the quantile regression loss function of Koenker and Bassett (1978). We denote the fixed effects estimator for count data as the limiting case $\hat{\boldsymbol{\theta}}(\tau) := \hat{\boldsymbol{\theta}}(\tau, \lambda)$ when $\lambda \to 0$.

The last term in definition (2.13) is introduced to improve the performance of the fixed effects estimator in Big data problems in which the dimensionality of the model, here proportional to N, can be large. It should be noticed that the penalty is different than the lasso-type penalty employed in panel data quantile regression (e.g., Koenker (2004), Harding and Lamarche (2014), among others). It represents a rather minor modification of existing routines and replaces the right hand side of the equality constraint in the dual formulation of the minimization problem by $(1 - \tau)(\boldsymbol{Z}'\iota_{NT} + \lambda\iota_N)$, where $\boldsymbol{Z}$ is a $NT \times N$ incidence matrix of individual effects and ι_N is a

$N \times 1$ vector of ones. In Big Data settings, this modification allows us to estimate and penalize individual specific distributional effects.

Under the assumption that an individual effect is a location-shift for the conditional distribution of the count variable, an alternative more efficient penalized estimator can be alternatively considered:

Step 1' (Optional): Let $\boldsymbol{\tau} = \{\tau_1, \tau_2, ..., \tau_J\}$ and J is the number of quantiles to be estimated. Estimate $\boldsymbol{\theta}(\boldsymbol{\tau}) = (\boldsymbol{\beta}(\boldsymbol{\tau})', \boldsymbol{\alpha}(\boldsymbol{\tau})')'$ by,

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\tau}, \lambda) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{j=1}^{J}\sum_{i=1}^{N}\sum_{t=1}^{T} \omega_j \rho_{\tau_j}(h(z_{it}, \tau_j) - \boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau_j) - \alpha_i) + \lambda \sum_{i=1}^{N} |\alpha_i|, \tag{2.14}$$

where ω_j is a relative weight given to the j-th quantile and $\sum_i |\alpha_i|$ is a penalty term. Koenker (2004) points out that the choice of the weights, ω_j, and the associated quantiles τ_j, is somewhat analogous to the choice of discretely weighted L-statistics, as for example in Mosteller (1946). An alternative less efficient, yet practical choice is to ignore the potential gains and estimate models with equal weights defined as $\omega_j = 1/J$.

Step 2: The τ-th quantile of the count variable, $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)$, can be estimated as,

$$\hat{Q}_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = \left\lceil \hat{Q}_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) - 1 \right\rceil \tag{2.15}$$

$$= \left\lceil h^{-1}\left(\boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}(\tau, \lambda) + \hat{\alpha}_i(\tau, \lambda), \tau\right) - 1 \right\rceil, \tag{2.16}$$

where as before $\lceil\cdot\rceil$ denotes the ceiling function.

Step 3: The partial effect of the independent variable x_j on the conditional quantile function of the count variable, $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)$, can be estimated as follows:

$$\Delta_j \hat{Q}_{Y_{it}}(\tau|\cdot, x_j^1, x_j^0) = \hat{Q}_{Y_{it}}(\tau|\cdot, x_j^1) - \hat{Q}_{Y_{it}}(\tau|\cdot, x_j^0), \tag{2.17}$$

where the quantile function is evaluated at fixed levels of the covariates and $\Delta x_j + x_j^0 = x_j^1$.

In conditional mean Poisson models, the slope parameter $\boldsymbol{\beta}$ can be consistently estimated for fixed T, as long as $N \to \infty$. Maximum likelihood (ML) estimation of a Poisson model with multiplicative fixed effects, after analytical expressions of α_i are concentrated in the likelihood function provided that the independent variables are strictly exogenous, do not exhibit the problem of incidental parameters. This result is true if both $\boldsymbol{\beta}$ and α_i are jointly estimated by ML. In fact, consistent estimates can be obtained by running standard Poisson cross-sectional regression on a transformed dependent variable on $\boldsymbol{x}_{it}$ and a series of incidence vectors for the individual effects. The situation might be different in a negative binomial model with fixed effects, although Greene (2004) suggests that incidental parameters might not impose large biases.

In panel quantile models, it is expected that incidental parameters might create biases when T is small, although Galvao, Lamarche and Lima (2013) shows that the performance of panel quantile methods rapidly improve under moderate T. (The analysis of incidental parameters in quantile regression is rigorously discussed in Kato, Galvao and Montes-Rojas (2012)). This paper however is primarily motivated by the analysis of Big data, a situation practitioners face in the use of click, tweet or scan data under the expectation that T is large.

With these caveats in mind, we will proceed as follows. Our derivation of large sample results in the next section relies under the assumption that both indexes go jointly to infinity. In Section 3, we investigate the finite sample performance of the approach in cases with a small number of time series observations T relative to the number of cross-sectional units N. Finally, in the application, we consider a panel data set for households making purchases over (on average) 44 months, reaching 72 months for a significant set of consumers.

2.3. **Large sample properties**

We analyze the large sample properties of the proposed approach. Let $Q_{Z_{it}}(\tau|\cdot) := Q_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = h^{-1}(\boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \alpha_i(\tau), \tau)$ and $Q_{Y_{it}}(\tau|\cdot) := Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = \lceil Q_{Z_{it}}(\tau|\cdot) - 1 \rceil$.

ASSUMPTION 3. *The discrete variable y_{it} has support on the set of non-negative integers, $\mathbb{N}_0$, and its conditional probability density function, $f_{Y_{it}}$, is bounded away from zero and infinity at the conditional quantile function $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)$.*

ASSUMPTION 4. *The variables α_i's are exchangeable, identically, and independently distributed with zero conditional quantile and distribution function G with continuous densities g uniformly bounded away from 0 and ∞, with bounded derivatives g' everywhere.*

ASSUMPTION 5. *There exist positive definite matrices $\boldsymbol{V}_0$, $\boldsymbol{V}_1$, and $\boldsymbol{D}$, such that*

$$\begin{aligned}
\boldsymbol{V}_0(\lambda) &= \lim_{T,N\to\infty} \frac{1}{TN} \tau(1-\tau) \left\{ \boldsymbol{X}'\boldsymbol{M}\boldsymbol{M}\boldsymbol{X} + \lambda^2 \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X} \right\}, \\
\boldsymbol{V}_1(\lambda) &= \lim_{T,N\to\infty} \frac{1}{TN} \left\{ \boldsymbol{X}'\boldsymbol{M}(\tau(1-\tau)\boldsymbol{I} - \boldsymbol{\Upsilon})\boldsymbol{M}\boldsymbol{X} + \lambda^2 \tau(1-\tau) \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{X} \right\}, \\
\boldsymbol{D}(\lambda) &= \lim_{T,N\to\infty} \frac{1}{TN} \left\{ \boldsymbol{X}'\boldsymbol{M}'\boldsymbol{\Phi}\boldsymbol{M}\boldsymbol{X} + \lambda \boldsymbol{X}'\boldsymbol{P}'\boldsymbol{\Psi}\boldsymbol{P}\boldsymbol{X} \right\},
\end{aligned}$$

where $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$, $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{\Phi}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{\Phi}$, $\boldsymbol{Z}$ *is an incidence matrix of dimension* $NT \times N$, $\boldsymbol{\Phi} = diag(f_{h(Z_{it},\tau)}(Q_{h(Z_{it},\tau)}(\tau|\cdot))$, $\boldsymbol{\Psi} = diag(g(0))$, *and*

$$\boldsymbol{\Upsilon} = diag\left(f_{Y_{it}}(Q_{Y_{it}}(\tau|\cdot))\left((Q_{Z_{it}}(\tau|\cdot) - Q_{Y_{it}}(\tau|\cdot))(1 - Q_{Z_{it}}(\tau|\cdot) + Q_{Y_{it}}(\tau|\cdot)\right)\right)\right).$$

ASSUMPTION 6. *The variable* $\boldsymbol{x}_{it} \in \mathcal{X}$ *has a bounded support satisfying* $\max||\boldsymbol{x}_{it}||/\sqrt{TN} \to 0$. *Moreover,* $\lambda_{\min}\left(E(\chi_M(\boldsymbol{x}_{it})\chi_M(\boldsymbol{x}_{it})')\right) > 0$ *and* $\lambda_{\min}\left(E(\chi_P(\boldsymbol{x}_{it})\chi_P(\boldsymbol{x}_{it})')\right) > 0$ *where* $\lambda_{\min}$ *is the smallest eigenvalue and* $\chi_M(\cdot)$ *and* $\chi_P(\cdot)$ *denote known transformations of the independent variables based on* $\boldsymbol{M}$ *and* $\boldsymbol{P}$.

Condition 3 is similar to Assumption A1 in Machado and Santos Silva (2005). It represents a slightly modified version of the standard quantile regression condition on the density $f_{Y_{it}}$ evaluated at a conditional quantile. The standard assumption on the continuity of the density conditional on independent variables guarantees a well-defined asymptotic behavior of the quantile regression estimator (see Koenker 2005). The zero quantile Assumption 4 is similar to the one used in Lamarche (2010) and it is possible to relax it as shown in Harding and Lamarche (2014) and in Corollary 2 below. The condition however is convenient to decompose the penalty term using Knight's identity (Koenker 2005). Assumption 5 is standard and is used to invoke the Lindeberg-Feller Central Limit Theorem. Lastly, Assumption 6 ensures that the conditional density is continuous almost for every realization of the independent variable $\boldsymbol{x}_{it}$ and it allows the finite-dimensional convergence of the objective function.

The following result is important for inference in panel count quantile regression models:

THEOREM 1. *Under conditions of 1-6, provided that there exists a constant c such that* $N^c/T \to 0$ *and that* $\lambda_T/\sqrt{T} \to \lambda \geq 0$, *the estimator for the slope coefficient over m-jittered samples is,*

$$\sqrt{NT}(\hat{\boldsymbol{\beta}}(\tau,\lambda) - \boldsymbol{\beta}(\tau)) \rightsquigarrow \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}(\lambda)^{-1}\boldsymbol{B}(\lambda)\boldsymbol{D}(\lambda)^{-1}),$$

where $\boldsymbol{B}(\lambda) = \boldsymbol{V}_0(\lambda)/m + (m-1)\boldsymbol{V}_1(\lambda)/m$.

Conditional on λ, it can be shown that the average-jittered estimator $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ is more efficient that the jittered estimator $\hat{\boldsymbol{\beta}}(\tau,\lambda)$ obtained by setting $m = 1$. The next result derives the covariance matrix of the proposed estimator when the number of jittered samples m tends to infinity.

COROLLARY 1. *Under the conditions of Theorem 1, when* $m \to \infty$, *the average-jittered estimator for the slope coefficient,* $\hat{\boldsymbol{\beta}}(\tau,\lambda)$, *is asymptotically normal with mean* $\boldsymbol{\beta}(\tau)$ *and covariance matrix,* $\boldsymbol{D}(\lambda)^{-1}\boldsymbol{V}_1(\lambda)\boldsymbol{D}(\lambda)^{-1}$.

Although shrinkage of the individual effects offers statistical and computational advantages in Big Data problems, it is possible to obtain an asymptotic distribution centered at zero without Assumption 4. The asymptotic distribution of the fixed effects quantile regression estimator for a count panel model is presented in the following result:

COROLLARY 2. *Under Assumptions 1, 2, 3, 5, and 6, when $\lambda_T/\sqrt{T} \to \lambda = 0$, provided that there exists a constant c such that $N^c/T \to 0$, the fixed effects estimator for the slope coefficient, $\hat{\boldsymbol{\beta}}(\tau)$, is asymptotically normal with mean $\boldsymbol{\beta}(\tau)$ and covariance matrix $\boldsymbol{D}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1}$, where $\boldsymbol{B} = \boldsymbol{V}_0/m + (m-1)\boldsymbol{V}_1/m$.*

2.4. Implementation issues and Sparsity-based Estimation

Sparsity has recently been argued to represent realistic features of economic problems and statistical models (see, e.g., Gabaix 2014, Belloni and Chernozhukov 2011, Belloni, Chen, Chernozhukov and Hansen 2012). Examples are consumers making online search and choice decisions and practitioners estimating models with a large number of independent variables. The approach developed in this paper overcomes the main problem of extending quantile methods for panel count models and this section discusses important implementation issues.

The distribution function of the count variable is naturally non-continuous and therefore the quantiles of the count variable, $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)$, cannot be modeled as a continuous function of the independent variable $\boldsymbol{x}_{it}$. We have defined a continuous variable $z_{it} = y_{it} + u_{it}$, which has quantile functions, $Q_{Z_{it}}(\tau|\cdot) = \tau/p_0$ if $\tau < p_0$ and $Q_{Z_{it}}(\tau|\cdot) = 1 + (\tau - p_0)/p_1$, if $p_0 < \tau < p_1$, where the probability $p_k = P(Y_{it} = k)$ for $k = 0, \ldots, K$. The quantile function of a continuous variable imposes a restriction that the quantile τ cannot be smaller than p_0. This condition can be imposed by considering a simple extension to panel data of the conditional quantile function introduced in Machado and Santos Silva (2005),

$$Q_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) = \tau + h^{-1}(\boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \alpha_i(\tau)). \tag{2.18}$$

In Big data problems, the number of groups or subjects included in equation (2.18), here denoted by α_i, can be very large. For instance, the application considered in Section 4 requires us to estimate a total number of 41,804 parameters. A sparse incidence matrix of individual effects, $\boldsymbol{Z}$, of dimension $NT \times N$ is large with $NT = 1,820,790$ and $N = 41,779$. Basic manipulations and calculations are handled in two stages. We first reduce the computational burden by storing data in the standard triplet form using the `Matrix` library developed by Bates and Maechler (2014) and the SparseM library in Koenker's (2014) R package quantreg. This is a relatively simple task as we only need to record the row and column numbers and value of the indicator for the individual effects or nonzero entries. We then use the sparse design matrices while employing panel quantile regression codes. As shown in Figure 2.1, the proposed algorithm leads to significant improvements of existing panel quantile methods and it appears to be particularly attractive to overcome the difficulties of high-dimensional panel quantile regression.[4]

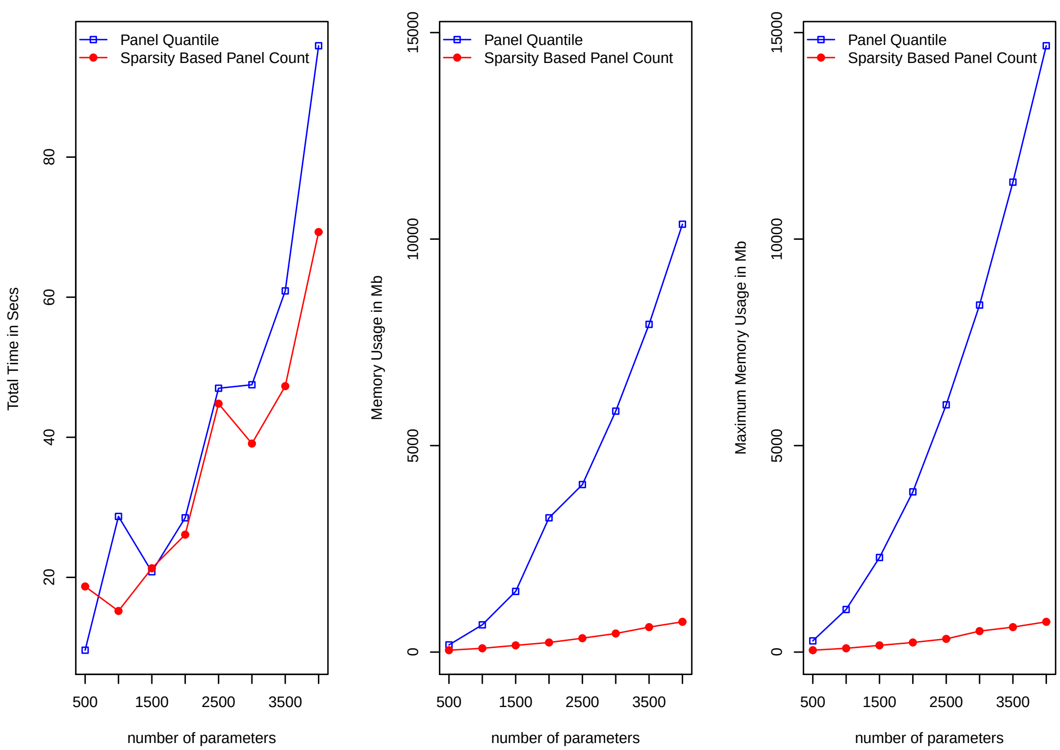

FIGURE 2.1. *Speed and computational cost of alternative sparse quantile panel approaches. Panel quantile is based on available algorithms and sparsity-based panel count denotes the implementation of the estimator proposed in this paper. Mb denotes Megabytes.*

[4]The figure presents results based on estimating a linear quantile regression model with N individual effects ranging from 500 to 4,000 and $T = 48$. The model also includes a slope parameter. To mimic the situation practitioners face when they analyze data, the model was estimated using R version 3.0.1 installed on a Dell Optiplex 7010 workstation which has a i5 Quad Core 3.4GHz processor and 16GB RAM.

In terms of estimation of the model and the standard errors of the fixed effects estimator employed in Section 4, we proceed similarly than Machado and Santos Silva (2005) and Winkelmann (2006) adopting the logarithmic transformation,

$$h(Z_{it},\tau)=\begin{cases}\log(z_{it}-\tau) & \text{for } z_{it}>\tau\\ \log(\epsilon_{it}) & \text{for } z_{it}\leq\tau,\end{cases}$$

where $Q_{\log(Z_{it},\tau)}=\boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau)+\alpha_i(\tau)$. Let $\hat{Q}_{Z_{it}}:=\tau+\exp(\boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}(\tau)+\hat{\alpha}_i(\tau))$, $\xi_{it}(\tau):=\boldsymbol{x}'_{it}\boldsymbol{\beta}(\tau)+\alpha_i(\tau)$, and $f_{it}:=(\boldsymbol{\Phi})_{it}=f_{h(Z_{it},\tau)}(Q_{h(Z_{it},\tau)}(\tau|\cdot))$. The matrices $\boldsymbol{V}_0$ and $\boldsymbol{V}_1$ can be estimated considering the following estimators:

$$\begin{aligned}\hat{\boldsymbol{V}}_0 &= \frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}(\tau-I(h(Z_{it},\tau)\leq\hat{\xi}_{it}))^2(\boldsymbol{x}_{it}-\tilde{\boldsymbol{x}}_i)\hat{f}_{it}(\hat{\xi}_{it})(\boldsymbol{x}_{it}-\tilde{\boldsymbol{x}}_i)',\\ \hat{\boldsymbol{V}}_1 &= \frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big(\tau^2+(1-2\tau)I(y_{it}\leq\hat{Q}_{Z_{it}}-1)+(\hat{Q}_{Z_{it}}-y_{it})I(\hat{Q}_{Z_{it}}-1<y_{it}\leq\hat{Q}_{Z_{it}})\Big)\\ &\quad(\hat{Q}_{Z_{it}}-y_{it}-2\tau)(\boldsymbol{x}_{it}-\tilde{\boldsymbol{x}}_i)\hat{f}_{it}(\hat{\xi}_{it})(\boldsymbol{x}_{it}-\tilde{\boldsymbol{x}}_i)',\end{aligned}$$

where $\bar{f}_i=T^{-1}\sum_t\hat{f}_{it}$ and $\tilde{\boldsymbol{x}}_i=(T\bar{f}_i)^{-1}\sum_{t=1}^{T}\hat{f}_{it}\boldsymbol{x}_{it}$. It is important to note that Q_Z is simply estimated by using panel quantile methods on a transformed dependent variable, therefore the estimation of $f_{it}(\xi_{it}(\tau))$ simply require the use of standard quantile regression methods (see, e.g., Koenker (2005), Chapter 3). The nuisance parameters $f_{it}(\xi_{it}(\tau))$ can be estimated for i.i.d. and non-i.i.d. models considering residuals $\hat{u}_{it}(\tau):=h(Z_{it},\tau)-\boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}(\tau)-\hat{\alpha}_i(\tau)$. Lastly, the practitioner can specify $\lambda>0$ and the covariance matrix can be similarly estimated by adding a penalty term and estimating the density of the individual effects under the requirement that $\hat{\alpha}_i(\tau)$ consistently estimates $\alpha_i(\tau)$ as N and T tend to infinity.

Lastly, the approach followed to estimate $\boldsymbol{D}$ represents a natural extension of Machado and Santos Silva (2005) applied to our panel model with fixed effects. We consider,

$$\hat{\boldsymbol{D}}=\frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\exp(\hat{\xi}_{it}(\tau))I(F_{NT}(\hat{Q}_{Z_{it}})\leq z_{it}\leq F_{NT}(\hat{Q}_{Z_{it}})+1)(\boldsymbol{x}_{it}-\tilde{\boldsymbol{x}}_i)\hat{f}_{it}(\hat{\xi}_{it})(\boldsymbol{x}_{it}-\tilde{\boldsymbol{x}}_i)',$$

where for a properly defined bandwidth $c_{NT}\in(0,1/2)$ satisfying $c_{NT}=o(1)$ as $N,T\to\infty$ and $\underline{w}=w-\lfloor w\rfloor$ with $\lfloor.\rfloor$ defined as the floor function, and

$$F_{NT}(w)=\begin{cases}\lfloor w\rfloor-1/2(1+\underline{w}/c_{NT}) & \text{if } \underline{w}<c_{NT} \text{ and } w\geq 1,\\ \lfloor w\rfloor & \text{if } c_{NT}\leq\underline{w}<1-c_{NT} \text{ or } w>1,\\ \lfloor w\rfloor+1/2(1+(\underline{w}-1)/c_{NT}) & \text{if } \underline{w}>1-c_{NT}.\end{cases}$$

In the empirical section, we adopt $c_{NT}=0.5\ln(\ln(NT))/\sqrt{NT}$.

The estimation of $\boldsymbol{V}_0$, $\boldsymbol{V}_1$ and $\boldsymbol{D}$ relies on concentrating the Bahadur representation of the individual effects in the objective function but one can directly estimate the covariance matrix of the fixed effects estimator by replacing $\boldsymbol{MX}$ by $[\boldsymbol{X}\vdots\boldsymbol{Z}]$ in Assumption 5. Naturally, this procedure represents another important implementation issue. For instance, the covariance matrix for the count fixed effects estimator in the empirical section is $41,804 \times 41,804$ and to obtain the covariance matrix it is required to invert several matrices of large dimensions. The following proposition facilitates the estimation and construction of the covariance matrix. Let $\boldsymbol{L}$ be a $p+N \times p+N$ matrix, $\boldsymbol{L}_{11}$ be a $p \times p$ sub-matrix, $\boldsymbol{L}_{12}$ a $p \times N$ sub-matrix, $\boldsymbol{L}_{21}$ a $N \times p$ submatrix, and $\boldsymbol{L}_{22}$ be a $N \times N$ sparse diagonal matrix. This natural ordering is associated with ordering methods considered in other sparse problems in linear systems (see, e.g., Saad 2003 and Koenker 2004).

PROPOSITION 1. *Let $\boldsymbol{L}$ and $\boldsymbol{S}$ be matrices of dimension $p+N$. Then, the $p \times p$ covariance matrix of $\boldsymbol{\beta}(\tau)$ is $(\boldsymbol{L}^{-1}\boldsymbol{S}\boldsymbol{L}^{-1})_{11} = \boldsymbol{W}_{11}^{-1}\boldsymbol{H}_{11}\boldsymbol{W}_{11}^{-1}$ where $\boldsymbol{W}_{11} = \boldsymbol{L}_{11} - \boldsymbol{L}_{12}\boldsymbol{L}_{21}$ and $\boldsymbol{H}_{11} = \boldsymbol{S}_{11} - \boldsymbol{L}_{12}\boldsymbol{S}_{21} - \boldsymbol{L}_{21}\boldsymbol{S}_{12} + \boldsymbol{L}_{12}\boldsymbol{S}_{22}\boldsymbol{L}_{21}$.*

The previous result allows for efficient estimation of potentially large sparse covariance matrices and do not necessarily rely on concentrating the Bahadur representation of the individual effects as in Theorem 1 and Corollary 2. The proposal is to trivially reduce the computational burden by avoiding to invert a large $N \times N$ matrix. We illustrate the use of this procedure in Section 4.

3. Monte Carlo

In this section, we report the results of several simulation experiments designed to evaluate the performance of the method in finite samples. We generate the dependent variable using the following model:

$$\mu_{it} = \exp(\beta_0 + \beta_1 x_{it} + \beta_2 \alpha_i), \tag{3.1}$$

$$x_{it} = \pi_0 + \pi_1 \alpha_i + u_{it}, \tag{3.2}$$

$$u_{it} \sim \mathcal{N}(0, \sigma_u^2), \tag{3.3}$$

where $\sigma_u^2 = \beta_0 = 1$, $\pi_0 = 0$ and the parameter of interest $\beta_1 = 0.5$. Multiplicative unobserved heterogeneity is denoted by $v_i = \exp(\alpha_i)$ and it is assumed to be drawn from the Gamma distribution and the Gaussian distribution.

We consider different distributions for the count variable y_{it}. In Table 3.1, the counts are Poisson random variables with conditional mean μ_{it} and $v_i = \exp(\alpha_i) \sim \Gamma(1,1)$. The table also includes results for the case that counts y_{it}'s are negative binomial random variables with mean μ_{it} and variance $\mu_{it} + 0.5\mu_{it}^2$. The distribution of unobserved heterogeneity α_i in the negative binomial case is i.i.d. Normal with mean 0 and variance $\sigma_\alpha^2 = 1$. Therefore, the distribution of v_i is lognormal with mean $\exp(\sigma_\alpha^2/2)$ and variance $\exp(\sigma_\alpha^2/2)(\exp(\sigma_\alpha^2/2) - 1)$. It is known that Poisson regression models for the conditional mean are often inappropriate for empirical analysis due to the over-representation of zero counts in real world data, so in our simulation experiments, we include models with a relatively large proportion of zeros, or "zero inflation" models (Gurmu and Trivedi (1996)). In Table 3.2, the counts y_{it} are distributed as a zero-inflated Poisson with a proportion of zero inflation of 0.1 and $\exp(\alpha_i)$ continues to be distributed as $\Gamma(1, 1)$. We also consider that the counts y_{it}'s are negative binomial random variables with mean μ_{it} and variance $\mu_{it} + 0.5\mu_{it}^2$ with a proportion of zero inflation of 0.1. The distribution of unobserved heterogeneity α_i in the negative binomial case is i.i.d. Normal with mean 0 and variance equal to 1.

In Table 3.1, we consider $N = \{500, 1000\}$ and $T = \{5, 20\}$ and the following basic variations of the model:

- **Design 1:** Individual specific effects do not enter multiplicatively in the conditional mean function and the independent variable is not correlated with unobserved individual heterogeneity. When the counts are Poisson, this case produces an average sample mean and variance for the dependent variable that are approximately equal to 2.8. We assume $\beta_2 = 0$ and $\pi_1 = 0$.
- **Design 2:** Individual unobserved heterogeneity is a latent variable in the count model and it is not correlated with the independent variable. In the Poisson case, the average count mean of 2.8 is smaller than the variance of the count variable which is equal to 14.6. In this case, we assume $\beta_2 = 1$ and $\pi_1 = 0$.
- **Design 3:** We consider the case that latent individual heterogeneity is correlated with the independent variable and individual specific effects enter in the model for the count y_{it} (i.e., $\beta_2 = 1$ and $\pi_1 = 1$). As in Design 2, this case produces an average count value that is smaller than the variance of the count variable.

This design specification allows us to explore the impact of estimating models with random individual effects (Design 2 vs Design 1) and the impact of having individual effects which are correlated with other right hand side variables (Design 3 vs Design 2).

All tables present the root mean square error (RMSE) for the slope parameter. We only report RMSE to the effect that the bias mirrors the RMSE closely and the variance of the estimators are

Sample Size		Quantile Regression Methods $\tau = 0.25$ quantile				$\tau = 0.50$ quantile			
N	T	QR	QC	QCFE	PQC	QR	QC	QCFE	PQC
		Poisson counts: Design 1							
500	5	0.095	0.052	0.043	0.043	0.004	0.046	0.042	0.042
500	20	0.095	0.052	0.045	0.046	0.002	0.038	0.034	0.035
1000	5	0.095	0.052	0.046	0.046	0.003	0.033	0.029	0.030
1000	20	0.095	0.052	0.046	0.047	0.002	0.028	0.026	0.026
N	T	Poisson counts: Design 2							
500	5	3.056	0.057	0.034	0.039	0.039	0.049	0.036	0.040
500	20	3.034	0.056	0.038	0.042	0.029	0.036	0.030	0.033
1000	5	3.054	0.056	0.039	0.043	0.032	0.027	0.026	0.029
1000	20	3.060	0.057	0.040	0.043	0.028	0.021	0.023	0.025
N	T	Poisson counts: Design 3							
500	5	2.760	0.623	0.013	0.081	0.494	0.607	0.014	0.050
500	20	2.743	0.625	0.015	0.065	0.488	0.583	0.012	0.043
1000	5	2.747	0.623	0.016	0.059	0.489	0.566	0.010	0.039
1000	20	2.759	0.623	0.016	0.058	0.491	0.553	0.009	0.035
N	T	Negative Binomial: Design 1							
500	5	1.042	0.021	0.031	0.025	0.027	0.019	0.028	0.023
500	20	0.782	0.021	0.031	0.025	0.015	0.013	0.022	0.017
1000	5	0.845	0.021	0.031	0.026	0.020	0.009	0.018	0.013
1000	20	0.767	0.022	0.032	0.026	0.011	0.007	0.015	0.010
N	T	Negative Binomial: Design 2							
500	5	2.960	0.011	0.018	0.015	0.037	0.010	0.029	0.024
500	20	2.969	0.012	0.027	0.023	0.021	0.004	0.025	0.019
1000	5	2.980	0.013	0.029	0.025	0.027	0.001	0.022	0.016
1000	20	2.975	0.013	0.029	0.025	0.018	0.002	0.020	0.014
N	T	Negative Binomial: Design 3							
500	5	2.892	0.552	0.020	0.234	0.723	0.549	0.005	0.179
500	20	2.893	0.553	0.007	0.198	0.713	0.542	0.005	0.173
1000	5	2.894	0.553	0.002	0.186	0.711	0.536	0.005	0.169
1000	20	2.898	0.554	0.003	0.184	0.712	0.532	0.006	0.166

TABLE 3.1. *Root mean squared error of a class of panel data estimators in the Poisson and Negative Binomial models. Unobserved heterogeneity is distributed as Gamma in the Poisson case and Gaussian in the Negative Binomial case. The evidence is based on 400 randomly generated samples considering 50 jittered samples.*

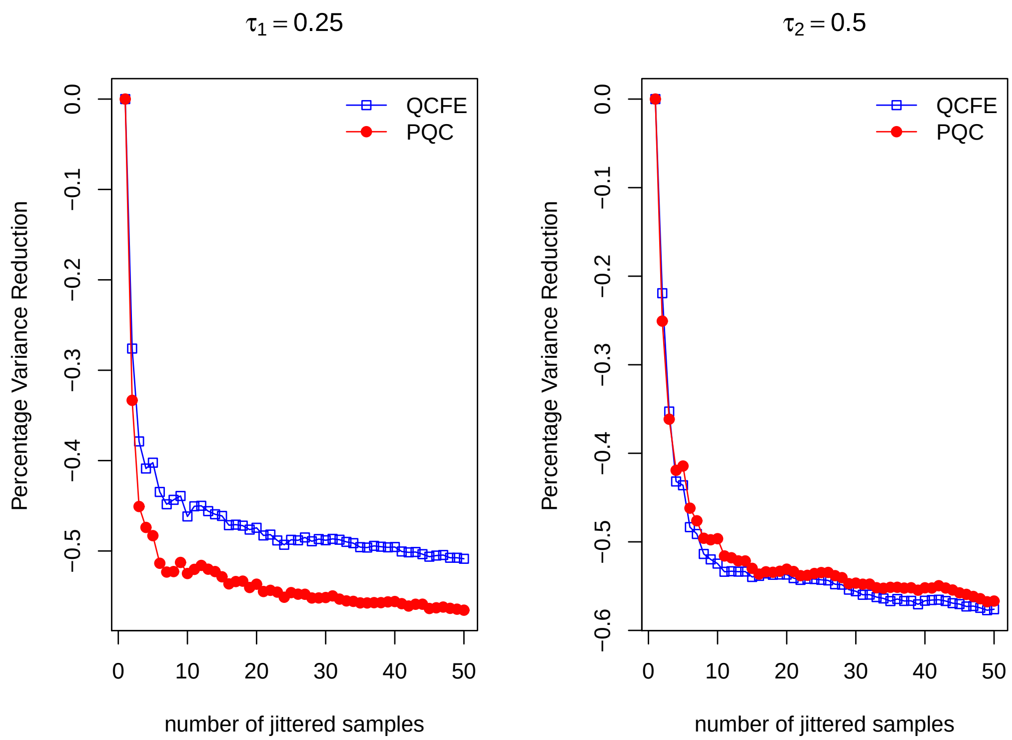

FIGURE 3.1. *Efficiency gains of the proposed panel count estimators.*

small in Big data problems. The methods are: Quantile regression (QR) as in Koenker and Bassett (1978), quantile for counts (QC) developed by Machado and Santos Silva (2005), average jittered estimator for a model with fixed effects (QCFE) which is defined in (2.12) - (2.13) by setting $\lambda = 0$, and average jittered estimator for a model with penalized fixed effects (PQC) which is defined in (2.12) - (2.13) by setting $\lambda > 0$. In the simulations, we select $\lambda = 1$ because it minimizes the variance of the PQR estimator in the negative binomial case with $\alpha_i \sim \mathcal{N}(0,1)$. Naturally, this choice is not expected to improve the performance of the QCFE estimator when $\exp(\alpha_i) \sim \Gamma(1,1)$. Moreover, the proposed approaches QCFE and PQC adopt the logarithmic transformation and are implemented using $m = 50$ jittered samples. Specifically, we obtain the dependent variable $\log(z_{it} - \tau)$ for $z_{it} > \tau$ and $\log(\epsilon_{it}) \leq \tau$ with $\epsilon_{it} = 10^{-5}$ and $z_{it} = y_{it} + u_{it}$ where u_{it} is distributed as $\mathcal{U}[0,1)$. Based on preliminary experiments, the choice of $m = 50$ provides a balanced compromise between computational cost in terms of time and efficiency gains. The main increase in precision

Sample Size		Quantile Regression Methods							
		$\tau = 0.25$ quantile				$\tau = 0.50$ quantile			
		QR	QC	QCFE	PQC	QR	QC	QCFE	PQC
N	T	Zero-inflated Poisson							
500	5	2.635	0.646	0.012	0.114	0.536	0.628	0.017	0.073
1000	5	2.647	0.643	0.017	0.085	0.529	0.587	0.012	0.056
N	T	Zero-inflated Negative Binomial							
500	5	2.602	0.477	0.072	0.202	0.901	0.483	0.037	0.156
1000	5	2.618	0.480	0.045	0.158	0.897	0.491	0.024	0.157

TABLE 3.2. *Root mean squared error of a class of panel data estimators in the case of zero-inflated models under Design 3. Unobserved heterogeneity is distributed as Gamma in the Poisson case and as Gaussian in the Negative Binomial case. The evidence is based on 400 randomly generated samples considering 50 jittered samples.*

can be achieved by selecting $m \geq 20$, as briefly illustrated in Figure 3.1 using simulations under Design 2. This evidence is consistent with the theoretical result introduced in Corollary 1.

First, let us consider the results for the Poisson model (Table 3.1). As expected the QR estimator performs in general worse than the other estimators across designs and is not suitable for the analysis of count data. The QC estimator performs reasonably well for Design 1 which does not include individual effects but its performance deteriorates in Design 3 as a result of its inability to correctly account for the presence of individual effects correlated with the independent variable. The QCFE estimator introduced in this paper performs very well across designs with a bias is no more than 4%-5% and very low RMSE too. As discussed before, in some practical situations it may be useful to consider the penalized version of this estimator too, which reduces the variance of the QCFE estimator by penalizing the estimation of the individual effects. In this data generating framework penalizing the model introduces few distortions in Design 3 and the results for PQC in Designs 1 and 2 are comparable in terms of bias and RMSE to those for the QCFE estimator.

In the lower block of Table 3.1, we explore the performance of the estimators further by using a negative binomial data generating process without zero inflation. The simulations continue to show that the QCFE estimator performs well. The RMSE is lower in this example than in the Poisson case due to the different distribution assumed for unobserved individual heterogeneity. In this case, the QC estimator performs well in Designs 1 and 2. Another very interesting development occurs when employing the penalized estimator PQC. The PQC estimator performs well for Designs 1 and

Sample Size		Quantile Count Model with Fixed Effects					
		$\tau = 0.25$			$\tau = 0.50$		
		1%	5%	10%	1%	5%	10%
N	T	Poisson counts: Design 2					
500	5	0.008	0.059	0.099	0.011	0.051	0.111
1000	5	0.009	0.053	0.104	0.015	0.063	0.120
N	T	Poisson counts: Design 3					
500	5	0.014	0.047	0.099	0.015	0.049	0.104
1000	5	0.010	0.048	0.100	0.015	0.060	0.104
N	T	Negative binomial: Design 2					
500	5	0.007	0.055	0.104	0.008	0.048	0.114
1000	5	0.015	0.057	0.111	0.010	0.057	0.106
N	T	Negative binomial: Design 3					
500	5	0.005	0.046	0.102	0.008	0.051	0.100
1000	5	0.011	0.057	0.107	0.006	0.052	0.105

TABLE 3.3. *Rejection probabilities for the Poisson and Negative Binomial Distributions at the median quantile. The evidence is based on 1000 randomly generated samples.*

2 improving the RMSE of the QCFE estimator, although its performance is substantially worse in Design 3. This shows that there are practical costs associated with the use of penalized estimators in models with dependence between α_i and x_{it} and the practitioner should keep these lessons in mind when employing panel count methods in applications.

Practitioners often face the problem of zero-inflated outcomes. Commonly encountered data often features an excess of zero observations. In the motivating examples for this paper, this is due to the fact that in many periods of interest households simply don't shop, click or tweet. Human activity tends to happen in bursts with significant periods where no activity is recorded. In order to evaluate the performance of our estimators in situations such as these we augment the previous Poisson generating process by setting 10% of the outcomes to zero. To save space, we consider $T = 5$ and we report results in Table 3.2 only for Design 3, which represents closely the situation we face in the empirical application in Section 4. The zero-inflated case is a more challenging setup and the associated costs in terms of bias and RMSE is visible in the case of QCFE estimator. Our proposed estimator continues to perform well and in most simulations the bias is less than 6% and it offers the best performance in terms of RMSE.

Lastly in Table 3.3 we compare the rejection probabilities for the quantile count model with fixed effects (QCFE) at 1%, 5%, 10% at the 0.25 and 0.5 quantiles for models based on the Poisson and Negative Binomial distributions. Of particular interest is the performance of the proposed estimator in models with individual unobserved heterogeneity (Designs 2 and 3). Given the relative novelty of the inferential approach proposed in this setting which differs substantially from that typically developed for a quantile regression model it is important to evaluate the rejection probabilities and determine the suitability of the inferential methods in practical settings. While the rejection probabilities suggest that the proposed approach works well in practice, at the 0.5 quantile in particular we observe a very robust performance of our methods.

Overall, the finite sample performance of the proposed methods for count data models with individual effects is very good in all the variants of the models considered in the simulations. When the degree of shrinkage is known and the degree of dependence between individual heterogeneity and the covariates is negligible, the PQC estimator appears to improve the performance of the QCFE estimator and can offer substantial efficiency gains in Big data applications. The QCFE estimator however has very low biases in all the variants of the models and it offers the best overall performance in the class of panel count models.

4. An Empirical Illustration

Modeling consumer search and choice behavior is central to many areas from economics and marketing to transportation research. While traditionally researchers focused on modeling trips to physical stores such as grocery stores (Bell, Ho and Tang 1998; Bawa and Gosh, 1999), most recently attention has shifted to include trips to virtual stores (Pozzi, 2012). Understanding the way in which consumers make choices over the number of trips to stores has important consequences for their search behavior over alternatives and ultimately the purchases they end up making. Therefore researchers try to model the number of trips in relation to customer demographics and local economic conditions. Understanding trip behavior has important consequences for firm decisions such as marketing and pricing strategies that enable a store to compete in the market place. Trip behavior is also important for deciding on store locations.

But attempts to model the number of shopping trips is not only relevant for firm actions but also impacts broader social planning decisions. The popularity of these models in transportation research reflects concerns related to traffic congestion and optimal city planning. In public health, researchers have recently started to explore the relationship between shopping trips and food purchases. Faced with increasing obesity rates increasing attention has been devoted to understanding purchasing environments and the way consumers make purchase decisions that are impacted by the availability of stores in their immediate proximity.[5] For example, it is now common to refer to areas where consumers lack access to healthy foods as "food deserts". These are areas associated with depressed economic conditions. According to some estimates over 29 million Americans live in food deserts and have to commute substantial distances to get to the nearest large supermarket or grocery store (Levi, Segal, St. Laurent, and Rayburn (2014)). Given that access to healthy food is not readily available for many consumers, grocery trip behavior is an important determinant of nutritional intake and eventual health outcomes. It complements other factors that are associated with the rise of obesity such as the decreasing cost of food, changing patterns of time allocation, and economic shocks (Sturm and Ruopeng (2014)).

In addition to demographic factors, researchers also focus on importance of local economic conditions as determinants of consumption outcomes. Dave and Kelly (2012) document the relationship between unemployment and the consumption of (healthy) foods. In other areas of consumption, wealth, local unemployment variation, and fluctuations in housing prices have also been investigated (see, e.g., Poterba 2000, Dynarski and Sheffrin 1987, Campbell and Cocco 2007).

The lack of suitable data has so far prevented a detailed investigation of the relative importance of demographics and local economic conditions in determining the number of shopping trips a household engages in during a given period of time. Given the broad implications of understanding shopping trips for both firms and social planners, the recent Great Recession provides an important source of variation for trying to disentangle these effects. In this section we document how the number of trips to grocery stores, relates to unemployment and housing prices as well as to household socioeconomic characteristics. Using detailed scanner data similar to the one previously employed in Burda, Harding and Hausman (2008, 2012) and Harding and Lovenheim (2014), we find evidence of a wealth effect operating through the house prices at the upper tail of the conditional trip distribution. We also find relatively weak demographic gradients at all quantiles. The

[5]Obesity is one of the major public health challenges of our time. Obesity has been associated with a variety of health conditions such as cancer, diabetes, and heart disease. Current health care costs associated with obesity are estimated to be between $147 billion and $210 billion per year (Levi, Segal, St. Laurent, and Rayburn (2014)). Over the last three decades obesity rates have more than doubled. Not only have obesity rates increased over time, but we have also witnessed substantial heterogeneity in obesity across a variety of socio-demographics. Today, 47.8% of African Americans are obese compared to 32.6% of Whites. Mississippi and West Virgia have obesity rates in excess of 35%, while the obesity rate in Colorado is only 21.3%. In 1980 obesity rates for all states were below 15%.

results also show that cross-sectional and panel results differ in quantitatively very meaningful ways. Once we account for unobserved heterogeneity the effect of observables is greatly diminished. This challenges the usefulness of using demographic variables to profile customers as a marketing device.

Variable	Mean	Std Dev	Quantiles				
			0.10	0.25	0.50	0.75	0.90
Number of shopping trips	8.691	5.720	3.000	5.000	7.000	11.000	16.000
Number of shopping days	6.791	3.995	2.000	4.000	6.000	9.000	12.000
Unemployment rate	6.662	2.809	3.800	4.500	5.800	8.400	10.700
Log of housing price	5.270	0.218	5.010	5.105	5.232	5.444	5.586
Unemployment	0.173	0.378	0.000	0.000	0.000	0.000	1.000
HH Income \$30k-\$45k	0.176	0.381	0.000	0.000	0.000	0.000	1.000
HH Income \$45k-\$70k	0.269	0.443	0.000	0.000	0.000	1.000	1.000
HH Income >\$70k	0.382	0.486	0.000	0.000	0.000	1.000	1.000
Kids under 12	0.195	0.396	0.000	0.000	0.000	0.000	1.000
Kids over 12	0.185	0.388	0.000	0.000	0.000	0.000	1.000
Married	0.610	0.488	0.000	0.000	1.000	1.000	1.000
2 household members	0.358	0.479	0.000	0.000	0.000	1.000	1.000
3 household members	0.164	0.370	0.000	0.000	0.000	0.000	1.000
4 household members	0.145	0.352	0.000	0.000	0.000	0.000	1.000
5 household members	0.057	0.231	0.000	0.000	0.000	0.000	0.000
6 or more members	0.030	0.170	0.000	0.000	0.000	0.000	0.000
Number of months	43.654	17.470	23.000	24.000	45.000	59.000	72.000
Number of households				41,779			
Number of observations				1,820,790			

Table 4.1. *Descriptive Statistics.*

4.1. **Data**

In our analysis we employ data from the Nielsen Homescan Panel over the period 2005-2010. The data is closely related to the sample used in Harding and Lovenheim (2014) which contains a more in-depth description of the different data elements. The data is collected by a large panel of households using home scanners. Enrolled households are provided with a device which records food purchases made in a store for at-home consumption at the Universal Product Code (UPC) level. For each transaction the database records detailed product and price information as well as the location and date of purchase. Household demographic characteristics are collected every year.

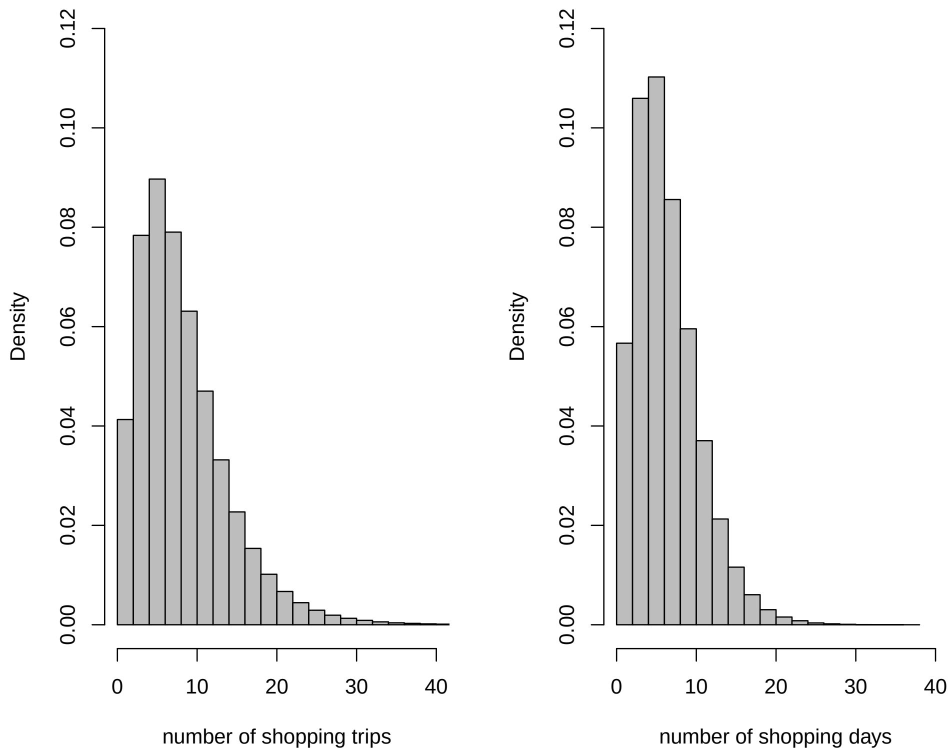

FIGURE 4.1. *The empirical distributions of number of shopping trips and number of shopping trips*

Although it is known that the Nielsen data offers certain sampling distortions (see Burda, Harding and Hausman (2012)), the detailed information on shopping trips offers a unique opportunity to understand search behaviors across the conditional distribution. A system of rewards and nudges is employed to induce a high participation and compliance rate. As a result the average participation in the panel is for over 43 months, with significant number of households staying in the panel for the entire 5 year period. Households are drawn from 52 large MSAs, ensuring a broad national coverage.

In this analysis we use an unbalanced panel of households trips to grocery stores over a maximum of 72 months, which is then augmented by the corresponding set of household demographic characteristics. We employ two measures of shopping trip behavior: number of trips in a month unique to each date and store name and number of trips in a month unique to each date. We interpret the former as the number of shopping trips to different stores in a given month and the latter as the number of shopping days. The total number of data entries used for estimation consists of 1,820,790 observations for a total of 41,779 households (Table 4.1). Based on our experience with working with the Nielsen panel we believe that these variables accurately capture the majority of the trips undertaken by the households to purchase food for at-home consumption. These measure do not include trips for food that is eaten outside of the house, e.g. in a restaurant. In our sample households engage in 8.7 store trips per month, over approximately 7 shopping days. The data shows a tremendous amount of variation in household trip behavior. At the 10th percentile, households shop only twice per month, while at the 90th percentile they go to grocery stores roughly every other day. Figure 4.1 illustrates the distribution of shopping trips and shopping days in the sample.

The households in our sample are characterized along a number of demographic dimensions, such as employment, household income, and family composition. In the sample, the (recorded) male or female household head is unemployed for over 17% of the sample, 61% of the households are married households, 24.6% of the households consist of single individuals, 17.3% of the households have household income <$30,000, while just over 38.2% of the households have income >$70,000, and over 32% of the households have children. These demographic patterns are consistent with those found in other papers using the Nielsen data, and reflect the extent to which the data collection strategy skews the sample towards higher income households. While in some studies, it is common to use demographic weights to rebalance the sample in order to match the demographic distribution found in Census data, we do not pursue this strategy within the context of this application.

Local economic conditions are measured using two variables, the local area unemployment rate and the housing price index. The Local Area Unemployment Rate is computed by the Bureau of Labor Statistics. The House Price Index (HPI) is computed by the Federal Housing Finance Authority and reflects house prices for single family homes. The Local Area Unemployment Rate is available at the county level, while the HPI is only available at the MSA level. For each household we know both the county and the MSA of their primary residence. The monthly values of the local unemployment rate and the HPI are thus assigned to each household based on their residence. Over the sampling period the local area unemployment rate varied both over time as a result of the Great Recession, and over geography as a result of events such as Hurricane Katrina with some MSAs experiencing unemployment rates in excess of 11%. Similar sources of variation are present

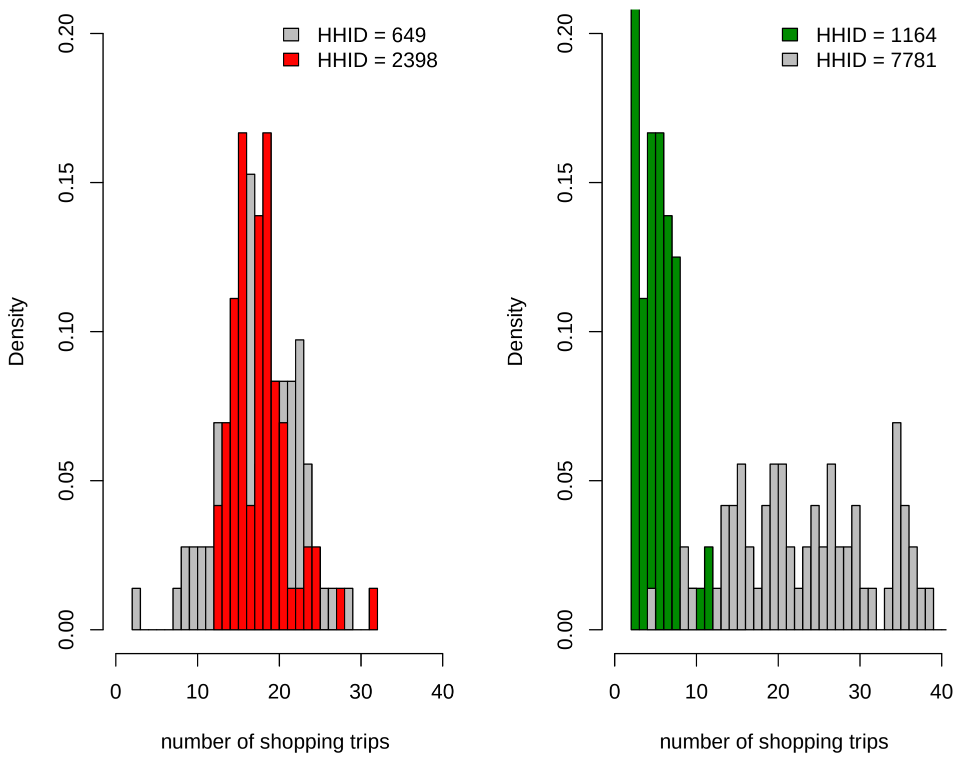

Figure 4.2. *Patterns of shopping trips for several households*

in the housing price data with a strong temporal dimension reflecting the collapse of the housing market during the sampling period.

The empirical analysis illustrates the use and estimation of the proposed count panel quantile approach to search and choice behavior using detailed transaction data. The data shows that different households exhibit different trip patterns over the time. For instance, the count distributions of households included in the left panel of Figure 4.2 have a similar location parameter but a different scale parameter, while the panel in the right contrasts a household with a very small number of shopping trips per month over the duration of the panel against a household with a very heterogeneous shopping trip activity ranging from only 1 or 2 trips in some months to close to 40 shopping trips in some other months.

4.2. **Model specification**

We estimate the following panel count model:

$$\mu_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{d}_{it}'\boldsymbol{\gamma} + \alpha_i, \tag{4.1}$$

for $\max(t) = 72$ months and $i = 1, ..., 41, 779$ households. Note that we are estimating the model on an unbalanced panel of close to 2 million observations. The vector $\boldsymbol{x}_{it}$ includes the local unemployment rate and a local housing market price index. The vector $\boldsymbol{d}_{it}$ includes indicators for unemployment the head of the household, housing income, children under 12 years of age, children over 12 years of age, an indicator for marital status, and indicators for the number of household members. In our setting, it is possible that α_i and $\boldsymbol{d}_{it}$ are not independent, so we estimate the model by fixed effects methods.

We consider the following conditional expectation, commonly encountered in standard count models:

$$E(y_{it}|\mu_{it}) = \exp(\mu_{it}) = \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{d}_{it}'\boldsymbol{\gamma} + \alpha_i) \tag{4.2}$$

where y_{it} denotes the count variable of interest. In addition to estimating the parameters of interest, we also aim to estimate the conditional quantile function of the count variable defined as follows:

$$\begin{aligned} Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \boldsymbol{d}_{it}, \alpha_i) &= \lceil Q_{Z_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i) - 1 \rceil, \\ &= \lceil \tau - 1 + \exp(\boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{d}_{it}'\boldsymbol{\gamma} + \alpha_i) \rceil, \end{aligned}$$

where $Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \boldsymbol{d}_{it}, \alpha_i)$ denotes the conditional quantile function for the count variable and τ is the quantile of interest.

4.3. **An Empirical Analysis**

Given the observed heterogeneity in shopping behavior documented in Table 4.1 and Figures 4.1 and 4.2, it is of interest to relate these to the observed demographics and measures of the strength of the local economy. We compare the estimated coefficients and marginal effects for the cross-section and panel count models for the 0.1, 0.5 and 0.9 quantiles of the outcome distribution. Table 4.2 reports the estimated coefficients for the model of shopping trips, while Table 4.3 reports the coefficients for the model of the number of shopping days. Table 4.4 reports the estimated marginal effects for both outcomes of interest. Throughout we report the corresponding standard errors for coefficients and confidence intervals for the marginal effects.

Variable	Quantiles						Mean
	0.1		0.5		0.9		
	CQ	QCFE	CQ	QCFE	CQ	QCFE	
Unemployment rate	0.000	-0.002	0.004	-0.001	0.006	0.000	-0.001
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
Log of housing price	-0.046	0.002	0.006	-0.040	0.039	-0.060	-0.045
	(0.005)	(0.012)	(0.003)	(0.007)	(0.003)	(0.008)	(0.004)
Unemployment	0.069	0.015	0.075	0.021	0.055	0.018	0.017
	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)
HH Income $30k-$45k	0.019	0.003	-0.019	0.009	-0.058	0.005	0.005
	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
HH Income $45k-$70k	-0.020	-0.011	-0.063	0.001	-0.101	-0.005	-0.004
	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
HH Income >$70k	-0.075	-0.018	-0.126	-0.003	-0.176	-0.006	-0.010
	(0.003)	(0.006)	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)
Kids under 12	-0.088	0.021	-0.089	0.014	-0.091	0.005	0.007
	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Kids over 12	0.001	0.021	0.012	0.023	0.008	0.021	0.024
	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Married	0.103	0.044	0.086	0.036	0.051	0.013	0.039
	(0.003)	(0.008)	(0.002)	(0.004)	(0.002)	(0.005)	(0.003)
2 household members	0.191	0.053	0.200	0.049	0.178	0.032	0.046
	(0.003)	(0.006)	(0.002)	(0.003)	(0.002)	(0.004)	(0.002)
3 household members	0.185	0.063	0.227	0.057	0.230	0.041	0.059
	(0.004)	(0.007)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
4 household members	0.182	0.061	0.251	0.060	0.272	0.045	0.060
	(0.005)	(0.008)	(0.003)	(0.004)	(0.004)	(0.005)	(0.003)
5 household members	0.183	0.073	0.285	0.058	0.324	0.051	0.065
	(0.007)	(0.010)	(0.004)	(0.005)	(0.004)	(0.006)	(0.003)
6 or more members	0.184	0.080	0.331	0.064	0.407	0.055	0.079
	(0.008)	(0.013)	(0.005)	(0.007)	(0.005)	(0.008)	(0.004)
Bimonthly effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of parameters	25	41804	25	41804	25	41804	24
Number of observations	1820790	1820790	1820790	1820790	1820790	1820790	1820790

TABLE 4.2. *Cross-section and panel count results for the number of shopping trips. The table shows Quantile for counts (QC) and the average jittered estimator for a model with fixed effects (QCFE). Mean denotes ML estimation for a model with fixed effects. Standard errors are presented in parentheses.*

Variable	Quantiles						Mean
	0.1		0.5		0.9		
	CQ	QCFE	CQ	QCFE	CQ	QCFE	
Unemployment rate	-0.003	-0.002	0.000	-0.002	0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log of housing price	-0.052	0.002	0.001	-0.030	0.038	-0.045	-0.032
	(0.005)	(0.011)	(0.003)	(0.006)	(0.003)	(0.006)	(0.005)
Unemployment	0.062	0.015	0.070	0.018	0.047	0.017	0.016
	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
HH Income \$30k-\$45k	0.023	0.006	-0.007	0.009	-0.051	0.003	0.006
	(0.003)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
HH Income \$45k-\$70k	-0.004	-0.005	-0.035	0.003	-0.082	-0.004	-0.001
	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
HH Income >\$70k	-0.051	-0.015	-0.084	-0.002	-0.136	-0.006	-0.008
	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)	(0.003)	(0.003)
Kids under 12	-0.080	0.025	-0.066	0.017	-0.065	0.005	0.012
	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Kids over 12	0.002	0.019	0.020	0.023	0.016	0.021	0.024
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Married	0.099	0.031	0.076	0.024	0.046	0.008	0.028
	(0.003)	(0.008)	(0.002)	(0.004)	(0.002)	(0.004)	(0.003)
2 household members	0.151	0.046	0.159	0.048	0.139	0.033	0.042
	(0.003)	(0.005)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
3 household members	0.131	0.053	0.173	0.053	0.177	0.041	0.051
	(0.004)	(0.006)	(0.002)	(0.004)	(0.003)	(0.004)	(0.003)
4 household members	0.119	0.049	0.187	0.051	0.213	0.044	0.051
	(0.005)	(0.007)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
5 household members	0.105	0.052	0.205	0.048	0.244	0.049	0.052
	(0.006)	(0.009)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
6 or more members	0.104	0.065	0.233	0.055	0.301	0.046	0.062
	(0.008)	(0.011)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)
Bimonthly effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of parameters	25	41804	25	41804	25	41804	24
Number of observations	1820790	1820790	1820790	1820790	1820790	1820790	1820790

TABLE 4.3. *Cross-section and panel count results for number of shopping days. The table shows Quantile for counts (QC) and the average jittered estimator for a model with fixed effects (QCFE). Mean denotes ML estimation for a model with fixed effects. Standard errors are presented in parentheses.*

At a basic level one of the first questions for an applied researcher is the extent to which the observed variation can be explained by observable attributes of the household or whether the choices people made are driven largely by unobserved attributes (which themselves may be correlated with observables). The comparison between cross-sectional and panel estimators is informative as it allows us to distinguish between the effect due to changing demographics and the effect due to household characteristics not included in the model, but which may be correlated with the observables of interest. Our data allows us to address this question due to the relatively long period over which we observe the households, which means that we can observe changes in some household characteristics such family size and income, even though we do not observe changes in other characteristics such as education. This further highlights the need for the use of panel methods.

First let us explore the extent to which demographic gradients explain the heterogeneity in shopping behavior. We would expect family size and composition to be important drivers of shopping behavior. We estimate a very pronounced positive gradient for family size for both the number of trips and the number of shopping days using the cross-sectional methods. This gradient increases with the quantile of the conditional outcome distribution. The gradient is however substantially lower once we estimate the same model using panel methods indicating that by ignoring the unmeasured individual heterogeneity, cross-sectional methods tend to overestimate the effect of family size. In terms of the marginal effect however, we find no effect at the 0.1 quantile for either measure of shopping behavior, but we do find a weak effect of one additional trip but not shopping day for households in the upper quantiles of the conditional outcome distributions. The marginal effect is the same for all household sizes greater than 2. This indicates that children induce an additional shopping trip for households in the upper quantiles but that the number of trips is not determined by the number of children. Notice however, that the effect becomes more pronounced and statistically significant for households with 5 or more members.

The coefficients for married households indicates a small impact which is decreasing across conditional distribution. The marginal effect estimated from the panel model is zero for all quantiles. Similarly, once we control for household size the presence of children of different ages does not seem to impact the marginal effect in the panel models. The coefficients on household income appear to indicate a small negative income gradient on the number of shopping trips and shopping days. This appears to be rather substantial at the upper quantiles in the cross-sectional models to have a generally negative effect on shopping activity. We find that this effect is quite substantial in the cross-sectional models but is much smaller in the panel models. Broadly speaking higher household income is associated with lower shopping activity. While the effects are relatively small, high income households, defined as having income over $70,000 tend to spend fewer days shopping, reflecting the increase cost of time and the ability to make larger purchases. Notice however that the marginal effect across all quantiles is zero for the panel models.

Households where the self-reported head of the household is unemployed are also more likely to have a higher number of shopping trips and shopping days at the tails of the outcome distribution, in cross-sectional models. This may be interpreted as reflecting increased search activity and also a lower cost of time. In panel data models however the marginal effect is zero.

Once we control for household head unemployment status we find no effect of the local area unemployment variable for either measure of shopping intensity and across all quantiles of the conditional distribution. Housing prices on the other hand have a more pronounced negative effect, which is larger at the upper quantiles of the conditional distribution. This may be indicative of an overall wealth effect which dampens overall shopping intensity. The marginal effect for the housing price index equates -1 for both the shopping trip and shopping days outcome at the 0.9 quantile in the panel models.

It is noteworthy that most of the variables considered have no impact at the lower quantiles of either the number of shopping trips and the number of shopping days. The only exceptions seem to be variables related to family size and housing prices. Larger households engage in additional shopping trips.

At the same time the decrease in wealth appears to have induced an increase in the number of shopping trips and shopping days at the upper quantiles, which may be indicative of an increase in search behavior. From an economic point of view this latter result is related to the debate in Campbell and Cocco (2007) and Attanasio et. al. (2009). While this evidence does not conclusively show that wealth impacts consumption, it does show that at least at the upper quantiles of the conditional trip distribution, the fall in house prices (and associated wealth reduction) did increase the number of shopping trips and shopping days. We believe this to be indicative of increased search behavior that would be expected as a result of households facing tighter budget constraints and thus having an increased incentive to engage in economizing behavior.

Overall, our paper however highlights the importance of unobserved heterogeneity in determining the intensity with which people shop. Using cross-sectional methods over-estimates the impact of household characteristics on shopping intensity. Once we account for unobserved heterogeneity the impact of observable demographics is greatly diminished. It appears that profiling customers based

Variable	Quantiles					
	0.1		0.5		0.9	
	CQ	QCFE	CQ	QCFE	CQ	QCFE
	Count variable = Number of shopping trips					
Unemployment rate	0[0,0]	0[0,0]	1[1,1]	0[0,0]	0[0,0]	0[0,0]
Log of housing price	0[0,0]	0[0,0]	1[0,1]	0[0,0]	0[0,0]	-1[-1,0]
Unemployment	0[0,0]	0[0,0]	1[1,1]	0[0,0]	1[1,1]	0[0,0]
HH Income $30k-$45k	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-1[-1,-1]	0[0,0]
HH Income $45k-$70k	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-2[-2,-2]	0[0,0]
HH Income >$70k	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-3[-3,-3]	0[0,0]
Kids under 12	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-2[-2,-2]	0[0,0]
Kids over 12	0[0,0]	0[0,0]	1[1,1]	0[0,0]	0[0,0]	0[0,0]
Married	0[0,0]	0[0,0]	1[1,1]	0[0,0]	1[0,1]	0[0,0]
2 household members	0[0,0]	0[0,0]	2[2,2]	0[0,1]	3[3,3]	0[0,1]
3 household members	0[0,0]	0[0,0]	2[2,2]	1[0,1]	4[4,4]	1[0,1]
4 household members	0[0,0]	0[0,0]	3[2,3]	1[1,1]	4[4,5]	1[0,1]
5 household members	0[0,0]	0[0,0]	3[3,3]	1[0,1]	6[5,6]	1[1,1]
6 or more members	0[0,1]	0[0,0]	3[3,4]	1[1,1]	7[7,8]	1[1,1]
Bimonthly effects	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of parameters	25	41804	25	41804	25	41804
Number of observations	1820790	1820790	1820790	1820790	1820790	1820790
	Count variable = Number of shopping days					
Unemployment rate	0[0,0]	0[0,0]	0[0,0]	0[0,0]	0[0,0]	0[0,0]
Log of housing price	0[0,0]	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-1[-1,-1]
Unemployment	1[1,1]	0[0,0]	0[0,1]	0[0,0]	1[1,1]	0[0,0]
HH Income $30k-$45k	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-1[-1,-1]	0[0,0]
HH Income $45k-$70k	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-1[-1,-1]	0[0,0]
HH Income >$70k	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-2[-2,-2]	0[0,0]
Kids under 12	0[0,0]	0[0,0]	0[0,0]	0[0,0]	-1[-1,-1]	0[0,0]
Kids over 12	0[0,0]	0[0,0]	0[0,0]	0[0,0]	0[0,0]	0[0,0]
Married	1[1,1]	0[0,0]	1[0,1]	0[0,0]	1[0,1]	0[0,0]
2 household members	1[1,1]	0[0,0]	1[1,1]	0[0,0]	2[2,2]	0[0,0]
3 household members	1[1,1]	0[0,0]	1[1,1]	0[0,0]	2[2,2]	0[0,0]
4 household members	1[1,1]	0[0,0]	1[1,1]	0[0,0]	3[3,3]	0[0,0]
5 household members	1[1,1]	0[0,0]	1[1,1]	0[0,0]	3[3,3]	0[0,0]
6 or more members	1[1,1]	0[0,0]	2[2,2]	0[0,0]	4[4,4]	0[0,0]
Bimonthly effects	Yes	Yes	Yes	Yes	Yes	Yes
Year effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of parameters	25	41804	25	41804	25	41804
Number of observations	1820790	1820790	1820790	1820790	1820790	1820790

TABLE 4.4. *Marginal effects in the cross-section and panel count model of shopping trips. 95 percent confidence intervals are in brackets.*

on commonly used demographic variables and measures of the strength of the local economy is of limited value. In practice marketers may also have access to a much more extensive vector of demographic variables than we have used in this example. We cannot exclude the possibilities that additional variables or finer measures may not perform better. At the same time the limited use of some of the core demographic variables ought to give us pause and make us re-evaluate our priors about the use of demographic information to predict customer choice behavior. This appears to be consistent with anecdotal evidence from marketers who prefer to use past consumption patterns as predictors of future choices. This appears to indicate that the value of Big Data for predicting consumption and choices may come not from a richer set of individual demographic measures but rather from the time series of past individual transactions.

5. Conclusion

This paper introduces a new quantile regression estimator for panel count data. It overcomes the challenge of implementing quantile modeling on discrete data by "jittering" the discrete outcome by the addition of uniform random noise. This allows us to construct continuous response variables whose conditional quantiles have a one-to-one relationship with the count response variable conditional on observables and unobservables. The availability of Big Data allows us to estimate panel models where we allow for unobserved individual heterogeneity. At the same time realistic applications such as the one discussed in Section 4 imply the need to estimate tens of thousands of parameters, which imposes substantial computational challenges. We overcome these challenges by adopting a new sparsity based implementation. At the same time we also explore penalized versions of our proposed estimator.

The paper presents asymptotic results and also compares the finite sample performance of the proposed estimator with that of existing alternatives. We find that the proposed estimator performs very well under a variety of scenarios. We do however caution against the use of penalization to reduce the computational burden in cases with significant zero inflation and endogenous covariates.

In the empirical application, we explore the extent to which demographic variables and measures of the local economy can be used to explain the observed variation in the number of trips and the number of shopping days for households buying food at the store for consumption at home. We use a unique transaction level dataset of close to 2 million observations for over 40,000 US households over a period of up to 72 months. The results indicate that once we account for unobserved heterogeneity in a panel data framework the impact of the explanatory variables nearly vanishes.

Interestingly, we do observe a small negative wealth effect at the upper tail of the conditional trip density which operates through the house price channel and may indicate that the recent recession increased search behavior.

Appendix A. Proofs

Proof of Theorem 1. The proof is divided in three steps. We first need to show that existing quantile regression asymptotic results can be employed. This can be simply demonstrated following Machado and Santos Silva (2005)'s Theorem 1 which shows that the limiting objective function has a second-order Taylor expansion and then it is possible to perform inference on $Q_Z(\tau|\cdot)$. Second, we obtain the Bahadur representation of $\tilde{\boldsymbol{\beta}}$ by concentrating out the Bahadur representation of $\tilde{\alpha}$. Because standard panel quantile regression is employed in the first step, we use the results in Koenker (2004) and Lamarche (2010) to derive the function of interest for one jittered sample and to obtain its asymptotic distribution. Lastly, we derive the asymptotic covariance matrix of the jittered estimator $\hat{\boldsymbol{\beta}}$. The estimator $\tilde{\boldsymbol{\beta}}(\tau, \lambda)$ depends on τ and λ, but in what follows, we assume τ and λ to be fixed and suppress these dependencies for notational simplicity. We also supress for notational simplicity conditioning on $\boldsymbol{x}_{it}$ and α_{i0}.

Under the regularity conditions with N and T going jointly to ∞, by Theorem 1 in Koenker (2004, p. 82) and Theorem 1 in Lamarche (2010), we have that,

$$\sqrt{NT}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \boldsymbol{D}^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\psi_\tau(h(z_{it}) - \xi_{it}(\tau)) + \frac{\lambda_T}{\sqrt{T}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}_i\psi_\tau(\alpha_i)\right] + o_p(1)$$

where $\psi_\tau(u) = \tau - I(u \leq 0)$, $\tilde{\boldsymbol{x}}_i = (T\bar{f}_i)^{-1}\sum_{t=1}^{T} f_{it}\boldsymbol{x}_{it}$, $\bar{f}_i = \sum_{t=1}^{T} f_{it}(\xi(\tau))$, $f_{it} := f_{h(Z_{it},\tau)}$ and the panel data conditional quantile function is $\xi_{it}(\tau) = \boldsymbol{x}_{it}'\boldsymbol{\beta}(\tau) + \alpha_i(\tau)$.

Let $z_{it}^{(l)} = y_{it} + u_{it}^{(l)}$ for $l = 1, \ldots, m$. By the regularity conditions, the quantile regression estimator $\sqrt{NT}(\tilde{\boldsymbol{\beta}}^{(l)} - \boldsymbol{\beta}) \rightsquigarrow \boldsymbol{D}^{-1}(\mathbb{B}^{(l)} + \lambda\mathbb{C})$, where $\mathbb{B}^{(l)}$ is a zero mean Gaussian vector with covariance matrix $\tau(1-\tau)E\left((\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)'\right)$ and under Assumption 4, $\mathbb{C}$ is a zero mean Gaussian vector with covariance matrix $\tau(1-\tau)E\left(\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i'\right)$. It follows then that the asymptotic covariance matrix is then $\boldsymbol{D}^{-1}\boldsymbol{V}_0\boldsymbol{D}^{-1}$. It then remains to obtain $m(m-1)$ asymptotic covariance matrices of the estimators corresponding to m jittered samples.

It follows that,

$$\begin{aligned}
& E\left(\sqrt{NT}(\hat{\boldsymbol{\beta}}^{(l)} - \boldsymbol{\beta})\sqrt{NT}(\hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta})\right) \\
= \; & \boldsymbol{D}^{-1}E\Bigg[\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\psi_\tau(h(z_{it}^{(l)}) - \xi_{it}(\tau)) + \frac{\lambda_T}{\sqrt{T}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}_i\psi_\tau(\alpha_i)\right] \\
& \left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\psi_\tau(h(z_{it}^{(k)}) - \xi_{it}(\tau)) + \frac{\lambda_T}{\sqrt{T}}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}_i\psi_\tau(\alpha_i)\right]\Bigg]\boldsymbol{D}^{-1}.
\end{aligned}$$

Let $\vartheta_{itl} := (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\psi_\tau(h(z_{it}^{(l)}) - \xi_{it}(\tau))$ and $\zeta_{il} := \tilde{\boldsymbol{x}}_i\psi_{\tau_0}(\alpha_i)$. Under Assumptions 1 and 2,

$$E\left[(\vartheta_{itl} + \zeta_{il})(\vartheta_{itk} + \zeta_{ik})'\right] = E(\vartheta_{itl}\vartheta_{itk}') + E(\zeta_{il}\zeta_{ik}').$$

Conditional on $\boldsymbol{x}_{it}$ and α_i and using definitions, we write

$$\begin{aligned}
E(\vartheta_{itl}\vartheta_{itk}') &= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)E\left(\psi_\tau(h(z_{it}^{(l)})\psi_\tau(h(z_{it}^{(k)})\right)(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)E\left((\tau - I(h(z_{it}^{(l)}) \leq \xi_{it}(\tau)))(\tau - I(h(z_{it}^{(k)}) \leq \xi_{it}(\tau)))\right)(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)E\Bigg((\tau^2 - \tau I(h(z_{it}^{(l)}) \leq \xi_{it}(\tau)) - \tau I(h(z_{it}^{(k)}) \leq \xi_{it}(\tau)) \\
&\quad + I(h(z_{it}^{(l)}) \leq \xi_{it}(\tau))I(h(z_{it}^{(k)}) \leq \xi_{it}(\tau))\Bigg)(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\left(E\left(I(z_{it}^{(l)} \leq h^{-1}(\xi_{it}(\tau)))I(z_{it}^{(k)} \leq h^{-1}(\xi_{it}(\tau)))\right) - \tau^2\right)(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\left(E\left(I(u_{it}^{(l)} \leq h^{-1}(\xi_{it}(\tau)) - y_{it})I(u_{it}^{(k)} \leq h^{-1}(\xi_{it}(\tau)) - y_{it})\right) - \tau^2\right)(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)',
\end{aligned}$$

Let $u_{it}(\tau) = h^{-1}(\xi_{it}(\tau)) - y_{it}$. Because the u's are i.i.d. from a uniform distribution, using the law of iterated expectations, we have that,

$$\begin{aligned}
E(\vartheta_{itl}\vartheta_{itk}') &= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\left[E\left(E\left(I(u_{it}^{(l)} \leq u_{it}(\tau))I(u_{it}^{(k)} \leq u_{it}(\tau))\right)|y_{it}\right) - \tau^2\right](\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\left[E\left(F_U(u_{it}(\tau))^2\right) - \tau^2\right](\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\Bigg[E\Bigg(I(y_{it} \leq h^{-1}(\xi_{it}(\tau)) - 1) + (h^{-1}(\xi_{it}(\tau)) - y_{it})^2 \\
&\quad I(h^{-1}(\xi_{it}(\tau)) - 1 < y_{it} \leq h^{-1}(\xi_{it}(\tau))\Bigg) - \tau^2\Bigg](\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\Big[P(y_{it} \leq h^{-1}(\xi_{it}(\tau)) - 1) + (h^{-1}(\xi_{it}(\tau)) - Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i))^2 \\
&\quad P(y_{it} = Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)) - \tau^2\Big](\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\Big[\tau - P(y_{it} = Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i))(h^{-1}(\xi_{it}(\tau)) - Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)) \\
&\quad (1 - h^{-1}(\xi_{it}(\tau)) - Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)) - \tau^2\Big](\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' \\
&= (\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)\Big[\tau(1 - \tau) - \big(F(Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i))(h^{-1}(\xi_{it}(\tau)) - Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i)) \\
&\quad (1 - h^{-1}(\xi_{it}(\tau)) + Q_{Y_{it}}(\tau|\boldsymbol{x}_{it}, \alpha_i))\big)\Big](\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)'.
\end{aligned}$$

On the other hand,

$$\begin{aligned} E(\zeta_{il}\zeta_{ik}') &= \lambda^2 E\left(\tilde{\boldsymbol{x}}_i\psi_\tau(\alpha_i)\psi_\tau(\alpha_i)\tilde{\boldsymbol{x}}_i'\right) = \lambda\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i' E\left((\tau - I(\alpha_i \le 0))(\tau - I(\alpha_i \le 0))\right) \\ &= \lambda^2\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i' E\left(\tau^2 - \tau I(\alpha_i \le 0) - \tau I(\alpha_i \le 0) + I(\alpha_i \le 0)\right) \\ &= \lambda^2\tau(1-\tau)\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i'. \end{aligned}$$

The last equality is obtained by Assumption 4 because $Q_\alpha(\tau|\cdot) = 0$. It follows then,

$$\begin{aligned} & E\left(\sqrt{NT}(\hat{\boldsymbol{\beta}}^{(l)} - \boldsymbol{\beta})\sqrt{NT}(\hat{\boldsymbol{\beta}}^{(k)} - \boldsymbol{\beta})\right) \\ = \; & \boldsymbol{D}^{-1}\left(\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)(\tau(1-\tau) - \Upsilon_{it})(\boldsymbol{x}_{it} - \tilde{\boldsymbol{x}}_i)' + \frac{\lambda_T}{\sqrt{T}}\frac{\tau(1-\tau)}{N}\sum_{i=1}^N\tilde{\boldsymbol{x}}_i\tilde{\boldsymbol{x}}_i'\right)\boldsymbol{D}^{-1}, \end{aligned}$$

with $\Upsilon_{it} = f_{Y_{it}}(Q_{Z_{it}}(\tau|\cdot) - Q_{Y_{it}}(\tau|\cdot))(1 - Q_{Z_{it}}(\tau|\cdot) + Q_{Y_{it}}(\tau|\cdot))$. □

Proof of Corollary 1. The proof follows immediately from Theorem 1 and Machado and Santos Silva's (2005) Theorem 5. □

Proof of Corollary 2. The results follows immediately by considering the limiting case of the penalized estimator when $\lambda \to 0$. □

Proof of Proposition 1. It follows that,

$$(\boldsymbol{L}^{-1}\boldsymbol{S}\boldsymbol{L}^{-1})_{11} = \begin{bmatrix} \boldsymbol{W}_{11}^{-1} & -\boldsymbol{W}_{11}^{-1}\boldsymbol{L}_{12} \end{bmatrix}\begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{bmatrix}\begin{bmatrix} \boldsymbol{W}_{11}^{-1} \\ -\boldsymbol{L}_{21}\boldsymbol{W}_{11}^{-1} \end{bmatrix},$$

where $\boldsymbol{W}_{11} = \boldsymbol{L}_{11} - \boldsymbol{L}_{12}\boldsymbol{L}_{22}^{-1}\boldsymbol{L}_{21} = \boldsymbol{L}_{11} - \boldsymbol{L}_{12}\boldsymbol{L}_{21}$. Then,

$$\begin{aligned} (\boldsymbol{L}^{-1}\boldsymbol{S}\boldsymbol{L}^{-1})_{11} &= \boldsymbol{W}_{11}^{-1}\left[\boldsymbol{S}_{11} - \boldsymbol{L}_{12}\boldsymbol{S}_{21} - \boldsymbol{S}_{12}\boldsymbol{L}_{21} + \boldsymbol{L}_{12}\boldsymbol{S}_{22}\boldsymbol{L}_{21}\right]\boldsymbol{W}_{11}^{-1} \\ &= \boldsymbol{W}_{11}^{-1}\boldsymbol{H}_{11}\boldsymbol{W}_{11}^{-1}. \end{aligned}$$

□

References

Anscombe, F. J. (1948): "The Validity of Comparative Experiments," *Journal of the Royal Statistical Society. Series A (General)*, 111(3), pp. 181–211.

Attanasio, O., H. R. Blow, L., and A. Leicester (2009): "Booms and Busts: Consumption, House Prices and Expectations," *Economica*, IFS WP 05/24.

Bates, D., and M. Maechler (2014): "Matrix: Sparse and Dense Matrix Classes and Methods," R package version 1.1-4, `www.r-project.org`.

BAWA, K., AND A. GOSH (1999): "A model of household grocery shopping behavior," *Marketing Letters*, 10(2), 149–160.

BELL, D., H. HO, TECK-HUA, AND C. TANG (1998): "Determining where to shop: Fixed and variable costs of shopping," *Journal of Marketing Research*, (3), 352–369.

BELLONI, A., D. CHEN, V. CHERNOZHUKOV, AND C. HANSEN (2012): "Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain," *Econometrica*, 80(6), 2369–2429.

BELLONI, A., AND V. CHERNOZHUKOV (2011): "L1-penalized quantile regression in high-dimensional sparse models," *The Annals of Statistics*, 39(1), 82–130.

BLUNDELL, R., R. GRIFFITH, AND F. WINDMEIJER (2002): "Individual effects and dynamics in count data models," *Journal of Econometrics*, 108(1), 113–131.

BROWNING, M., A. DEATON, AND M. IRISH (1985): "A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle," *Econometrica*, 53(3), pp. 503–544.

BURDA, M., M. HARDING, AND J. HAUSMAN (2008): "A Bayesian mixed logit-probit model for multinomial choice," *Journal of Econometrics*, 147(2), 232–246.

——— (2012): "A Poisson mixture model of discrete choice," *Journal of Econometrics*, 166(2), 184–203.

CAMERON, C., AND P. TRIVEDI (2013): *Regression Analysis of Count Data*. Cambridge University Press, 2 edn.

CAMPBELL, J., AND J. COCCO (2007): "How do house prices affect consumption? Evidence from micro data," *Journal of Monetary Economics*, 54, 591–621.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, J. HAHN, AND W. NEWEY (2013): "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81, 535–580.

DEATON, A. (1985): "Panel data from time series of cross-sections," *Journal of Econometrics*, 30, 109 – 126.

DYNARSKI, M., AND S. M. SHEFFRIN (1987): "Consumption and Unemployment," *The Quarterly Journal of Economics*, 102(2), 411–28.

GABAIX, X. (2014): "A Sparsity-Based Model of Bounded Rationality," *The Quarterly Journal of Economics*, 129(4), 1661–1710.

GALVAO, A. F. (2011): "Quantile regression for dynamic panel data with fixed effects," *Journal of Econometrics*, 164(1), 142 – 157.

GALVAO, A. F., C. LAMARCHE, AND L. R. LIMA (2013): "Estimation of Censored Quantile Regression for Panel Data With Fixed Effects," *Journal of the American Statistical Association*, 108(503), 1075–1089.

GREENE, W. (2004): "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects," *Econometrics Journal*, 7(1), 98–119.

GURMU, S., P. RILSTONE, AND S. STERN (1999): "Semiparametric estimation of count regression models," *Journal of Econometrics*, 88(1), 123 – 150.

GURMU, S., AND P. K. TRIVEDI (1996): "Excess Zeros in Count Models for Recreational Trips," *Journal of Business & Economic Statistics*, 14(4), 469–77.

HARDING, M., AND C. LAMARCHE (2009): "A Quantile Regression Approach for Estimating Panel Data Models Using Instrumental Variables," *Economics Letters*, 104, 133–135.

Harding, M., and C. Lamarche (2014): "Estimating and testing a quantile regression model with interactive effects," *Journal of Econometrics*, 178, Part 1(0), 101 – 113.

Harding, M., and M. Lovenheim (2014): "The Effect of Product and Nutrient-Specific Taxes on Shopping Behavior and Nutrition: Evidence from Scanner Data," working paper, NBER.

Hausman, J., B. H. Hall, and Z. Griliches (1984): "Econometric Models for Count Data with an Application to the Patents-R & D Relationship," *Econometrica*, 52(4), pp. 909–938.

Hong, H. G., and X. He (2010): "Prediction of Functional Status for the Elderly Based on a New Ordinal Regression Model," *Journal of the American Statistical Association*, 105(491), 930–941.

Kato, K., A. F. Galvao, and G. Montes-Rojas (2012): "Asymptotics for Panel Quantile Regression Models with Individual Effects," *Journal of Econometrics*, 170, 76–91.

Koenker, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

——— (2005): *Quantile Regression*. Cambridge University Press.

——— (2014): "Quantreg," R package version 5.02, `www.r-project.org`.

Koenker, R., and G. Bassett (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Lamarche, C. (2010): "Robust Penalized Quantile Regression Estimation for Panel Data," *Journal of Econometrics*, 157, 396–408.

Levi, J., L. Segal, R. St. Laurent, and J. Rayburn (2014): "The State of Obesity 2014," Discussion paper, Trust for America's Heath and the Robert Wood Johnson Foundation.

Machado, J. A., and J. M. C. S. Silva (2005): "Quantiles for Counts," *Journal of the American Statistical Association*, 100, 1226–1237.

Manski, C. F. (1985): "Semiparametric analysis of discrete response : Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27(3), 313–333.

Mosteller, F. (1946): "On Some Useful 'Inefficient' Statistics," *Annals of Mathematical Statistics*, 17, 377–408.

Pearson, E. S. (1950): "On Questions Raised by the Combination of Tests Based on Discontinuous Distributions," *Biometrika*, 37, pp. 383–398.

Poterba, J. M. (2000): "Stock Market Wealth and Consumption," *Journal of Economic Perspectives*, 14(2), 99–118.

Pozzi, A. (2012): "Shopping cost and brand exploration in online grocery," *American Economic Journal: Microeconomics*, 4(3), 96–120.

Saad, Y. (2003): *Iterative Methods for Sparse Linear Systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2nd edn.

Stevens, W. L. (1950): "Fiducial Limits of the Parameter of a Discontinuous Distribution," *Biometrika*, 37, 117–129.

Sturm, R., and A. Ruopeng (2014): "Obesity and Economic Environments," *CA: A Cancer Journal for Clinicians*.

Taddy, M. (2013): "Multinomial Inverse Regression for Text Analysis," *Journal of the American Statistical Association*, 108(503), 755–770.

Taddy, M. (2014): "Distributed Multinomial Regression," *ArXiv e-prints*.

Winkelmann, R. (2006): "Reforming health care: Evidence from quantile regressions for counts," *Journal of Health Economics*, 25(1), 131–145.

——— (2008): *Econometric Analysis of Count Data*. Wiley.

Wooldridge, J. (1999): "Distribution-free estimation of some nonlinear panel data models," *Journal of Econometrics*, 90(1), 77 – 97.