Quantile Regression using Metaheuristic Algorithms

Mohammad Arshad Rahman

Department of Humanities and Social Sciences, Indian Institute of Technology Kanpur, 208016, India.
Office: Room No. 672, 6th Floor, Faculty Building. Phone: +91 969-564-7700. Fax: +91 512-259-7510.

Abstract

This paper demonstrates that metaheuristic algorithms can provide a useful general framework for estimating both linear and nonlinear econometric models. Two metaheuristic algorithms—firefly and accelerated particle swarm optimization—are employed in the context of several quantile regression models. The algorithms are stable and robust to the choice of starting values and the presence of various complications (e.g. non-differentiability, parameter restrictions, discontinuity, possible multimodality, etc.). Two comparative studies involving an autoregressive model and a conditional scale autoregressive conditional heteroscedasticity model, demonstrate the performance of metaheuristic algorithms relative to existing approaches. In addition to these examples, the paper also offers an application to consumption behavior in which the presence of constraints makes existing techniques difficult to implement, but metaheuristic algorithms are straightforward to apply. The findings indicate that, contrary to popular perception, marginal propensity to consume is highest in Quarter 3 for each of the sample years. However, pre- and post-recession comparisons reveal interesting asymmetries in consumption behavior.

Keywords: quantile regression; firefly algorithm; particle swarm optimization; conditional scale ARCH model; consumption study; economic recession.

---

1. Introduction

Quantile regression (Koenker and Bassett, 1978) allows estimation of the economic relationships between the dependent variable and the covariates at various points over the conditional distribution of the dependent variable given the covariates. The focus on several points

[1]Email address: marshad@iitk.ac.in. This paper forms a part of my dissertation and was largely written at University of California, Irvine. I am grateful to Ivan Jeliazkov and Dale Poirier for their continuous guidance and helpful comments.

(quantiles) in the conditional distribution of the dependent variable, rather than only on the conditional mean as in least squares estimation, provides a more comprehensive picture of the underlying relationships of interest. The methodology offers robustness against outliers of the response variable, higher efficiency than least squares over a wide class of non-Gaussian error distributions, and has desirable equivariance properties (Koenker and Bassett, 1978; Koenker, 2005). However, these benefits come at the cost of working with an objective function which is not differentiable and cannot be optimized by standard hill-climbing algorithms. Consequently, the estimator does not have a general closed-form solution and the objective function is minimized using linear programming (LP) techniques such as the simplex method and the interior point (IP) algorithm.

The LP techniques, although efficient and theoretically sound for linear models, cannot be used in a variety of situations. First, LP techniques are inapplicable to models that are non-linear in parameters. In this respect, a significant contribution was Koenker and Park (1996), where they proposed an algorithm based on interior point methods (henceforth referred to as the KP algorithm). However, their proposed algorithm requires local linearization of the objective function, suffers from initial value problems at higher and lower quantiles (shown later) and most importantly fails to respect bounds on model parameters. In addition, Koenker and Park (1996) suggest that the algorithm may also suffer due to rank deficiency of the model Jacobian. Second, LP techniques cannot be used when the objective function is stepwise or discontinuous (e.g. many discrete data models). Existing methods to estimate such models include maximum score estimator for binary response models (Manski, 1975), Powell's estimator for censored regression models (Powell, 1986) and an estimator based on smoothing device for count data models (Machado and Silva, 2005). Third, another problem arises when the model parameters are constrained, e.g. there are parameter bounds or other restrictions, which dramatically increases estimation complexity. For example, estimation of the conditional scale ARCH model (Koenker and Zhao, 1996) requires a multi-step plug-in approach which results in increased estimation costs and leads to inefficiency as discussed in Section 4.2.

The abundance of algorithms to cater different models and their difficulty and inadequacy to incorporate model specific information has long been recognized, but no effort has been made (at least to my knowledge) to develop or explore algorithms that can implement quantile regression on a wide range of models. This paper attempts to fill the gap by employing *metaheuristic algorithms*, a large and diverse class of algorithms that has been extensively employed in engineering (electrical, civil and mechanical), computer science and bio- and medical applications. In particular, the paper adapts and implements two metaheuristic algorithms—the firefly algorithm (FA) and the accelerated particle swarm optimization (APSO) algorithm—to quantile regression in order to deal with key difficulties of conventional techniques. Metaheuristic algorithms are nature inspired, typically derivative free, global optimization algorithms. Each of these terms emphasizes an associated advantage offered by metaheuristic algorithms. The algorithms are termed nature inspired since they are based on successful evolutionary behavior of natural systems and are generally very robust. The algorithms are typically derivative free and do not require the gradient or Hessian matrix, so the objective function need not be differentiable or even continuous. Lastly, they are better designed to optimize multimodal objective functions and can search large spaces of candidate solutions without assuming much about the problem being optimized. To add to above advantages, metaheuristic algorithms are known to be versatile, which in this paper is illustrated by estimation of both linear and nonlinear quantile regression models. Moreover, they are robust to initial value problem and can be easily modified (as done in the paper) to incorporate model information such as parameter restrictions. While it is theoretically possible to incorporate parameter restrictions with existing techniques, practical application becomes complex, even in simple cases.

Quantile regression models are estimated in two studies using the firefly algorithm, although any other metaheuristic algorithms including the PSO algorithm could have been employed with absolutely no difficulty, and results compared to that obtained from IP and KP techniques. In the first study, the 3-month US treasury bill rate (a common measure of short term interest rate) is modelled as a first order quantile autoregressive QAR(1) model. The objective is threefold: first, to compare QAR(1) model estimates obtained from the firefly and IP algorithm; second, to demonstrate the use of bounds in linear models; and third, to investigate the quantile crossing problem which may occur at low values of the interest rate, as observed during the recent recession. The model is estimated at 19 quantiles from 0.05 to 0.95 with increments of 0.05. Results show that estimates obtained from the firefly and IP algorithm are practically identical and the estimated conditional quantile functions do not cross each other for interest rates between 0.25 and 16.30 percent (maximum observed interest rate). However, nonmonotonicity issues occur for values below 0.25 observed during the recent recession. In the second study, the parameters of the conditional scale ARCH model (Koenker and Zhao, 1996) are jointly estimated using the firefly algorithm. Joint estimation is not possible using the simplex method or IP algorithm since the model is nonlinear in parameters. Hence, for comparison purposes, the conditional scale ARCH model was also estimated using the KP algorithm (Koenker and Park, 1996). Estimation suggests that the firefly algorithm can efficiently utilize model information to impose bounds on parameters and, unlike the KP algorithm, is more robust and does not suffer from (or rather less susceptible to) initial value problem.

The final section of the paper presents a small consumption model to compare consumption behavior during 2010, a year following the economic recession of December 2007 – June 2009, relative to 2005, a non-recession year. Attention was confined to marginal propensity to consume (MPC), because MPC can vary along consumption quantiles. The model was estimated using a version of the PSO algorithm known as the accelerated particle swarm optimization (APSO) algorithm (firefly algorithm can also be employed without any difficulty). In addition, the APSO algorithm was modified to emphasize the idea that any algorithm must incorporate the information MPC $\in (0, 1)$. Results show that MPC at different quantiles was higher during Q3 relative to other quarters for both 2005 and 2010. Such an occurrence is termed "summer effect" to denote increased propensity to spend by families on travel, shopping, amusement parks, etc., during summer vacation. When compared across years, MPC is higher during Q1-Q3 of 2005 for quantiles above the 30th quantile. This is consistent with the view that consumption is procyclical. During Q4, MPCs for 2010 are higher than MPCs for 2005 till the 35th quantile, remain almost the same until the 70th quantile, and thereafter is lower except at the 95th quantile. This implies that MPC actually increased in 2010 at lower consumption levels, remained the same for moderate consumption levels and decreased at higher consumption levels.

The remainder of the paper is organized as follows. Section 2 introduces the quantile regression problem as presented in Koenker and Bassett (1978). Section 3 presents the FA and PSO algorithms together with the pseudo code to solve the quantile regression problem. Section 4 uses the firefly algorithm in two comparative studies to illustrate some of the advantages of metaheuristic over conventional algorithms. Section 5 presents a consumption study where the regression quantiles are estimated using APSO algorithm, and Section 6 presents concluding remarks.

2. Quantile Regression

The τ-th quantile of a random variable Y is the value y such that the probability that Y will be less than y equals $\tau \in (0, 1)$. Formally, if $F(\cdot)$ denotes the cumulative distribution function of Y, the τ-th quantile is defined as

$$F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}.$$

The idea of quantiles is extended to regression analysis via quantile regression, where the aim is to estimate *conditional quantile functions* with $F(\cdot)$ being the conditional distribution function of the data given the covariates. Focusing on a set of quantiles, rather than solely on the conditional mean, produces a more complete picture of the underlying regression relationships. Regression quantiles are also known to have few desirable properties including equivariance to monotone transformation and robustness against outliers of the response variable, equivariance to reparameterization of design, comparable efficiency (to least squares estimator) for Gaussian error distribution and higher efficiency for a wide range of non-Gaussian error distributions. (Koenker and Bassett, 1978; Koenker, 2005).

However, the advantages offered by quantile regression comes at the cost of higher computational burden, relative to least squares, in minimizing the objective function—after all 'there's no such thing as a free lunch'. The quantile objective function is a sum of asymmetrically weighted absolute residuals, minimization of which yields regression quantiles. To formally explain the quantile objective function and the associated estimation problem, consider a linear model,

$$y = X\beta + \epsilon,$$

where y is a $n \times 1$ vector of responses, X is a $n \times k$ covariate matrix, β is a $n \times 1$ vector of unknown parameters and ϵ is a $n \times 1$ vector of unknown errors. The quantile regression problem can be formulated as minimizing (with respect to β) the following weighted loss function,

$$\min_{\beta \in \mathbf{R}^k} \left[\sum_{i:y_i \leq x_i'\beta} (1-\tau)\, |y_i - x_i'\beta| \;+ \sum_{i:y_i \geq x_i'\beta} \tau\, |y_i - x_i'\beta| \right], \tag{1}$$

where the solution vector $\hat{\beta}_\tau$ gives the τ-th regression quantile and the estimated conditional quantile function is obtained as $\hat{Y} = X\hat{\beta}_\tau$. Note that the objective function (1) is such that all observations above the estimated hyperplane $X\hat{\beta}_\tau$ are weighted by τ, and all observations below the estimated hyperplane are weighted by $(1-\tau)$. The objective function (1) can therefore be written as a sum of piecewise linear functions or check functions as follows,

$$\min_{\beta \in \mathbf{R}^k} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta),$$

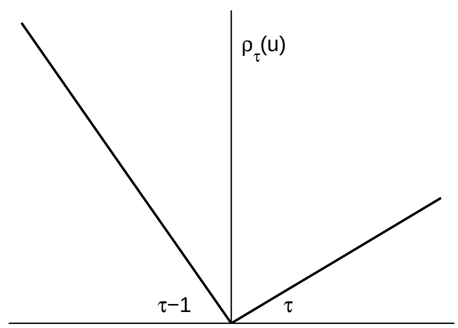


Figure 1: Quantile regression ρ function

where, the check function $\rho_\tau(u) = u(\tau - I(u < 0))$ and $I(\cdot)$ is an indicator function, which equals 1 if the condition inside the parenthesis is true and 0 otherwise. It is obvious from Figure 1 that the check function is not differentiable at the origin. Consequently, the estimator does not have a general closed form solution and one needs to use computational methods to optimize (minimize) the quantile objective function.

The quantile regression objective function (1) has traditionally been minimized using LP techniques such as the simplex algorithm (Dantzig, 1963; Dantzig and Thapa, 1997, 2003; Barrodale and Roberts, 1973; Koenker and d'Orey, 1987), and the interior point algorithm (Karmarkar, 1984; Mehrotra, 1992), but each has certain limitations. The simplex algorithm is known to be computationally demanding for linear models because the number of iterations increases exponentially with sample size (Koenker, 2005). IP algorithm overcomes the computational difficulty but has other issues. The computational issues with simplex algorithm, IP algorithm and *smoothing algorithm* (Madsen and Nielsen, 1993; Chen, 2007) are well documented in Chen and Wei (2005). In addition, simplex and IP algorithms become increasingly complex when modified to incorporate bounds on parameters, cannot estimate nonlinear models or models with stepwise or discontinuous objective function. The nonlinear models, with continuous objective function, can be estimated using the KP algorithm (Koenker and Park, 1996) provided the objective function satisfies the local linearization requirement. Nonetheless, the KP algorithm, at least in its present state, cannot incorporate model information such as parameter restrictions and suffers from initial value problem at quantiles far from the median.

To overcome some of the above difficulties in estimating regression quantiles with conventional techniques, this paper borrows two nature inspired simulative algorithms, collectively known as "metaheuristic algorithms", and modifies them to estimate regression quantiles for linear and nonlinear models. They form the subject of the following section.

3. Metaheuristic Algorithms

The term *metaheuristic*, introduced by Glover (1986), is derived from two Greek words, "meta" and "heuriskein". The word "meta" means "upper level methodology" and heuristic (derived from *heuriskein* meaning *to discover*) means the art of discovering new strategies (rules) to solve problems (Talbi, 2009). Metaheuristic algorithms can thus be defined as upper level general methodologies which guide other heuristics to search for feasible solutions in any given optimization problem.

In common parlance, metaheuristic algorithms are defined as nature inspired, derivative free, global optimization algorithms. Each term in this definition emphasizes an associated advantage offered by metaheuristic algorithms as described in Section 1. Metaheuristic algorithms can also be classified into different types depending on number of solutions, search process and type of objective function. Talbi (2009) organized the algorithms under the following headings: single solution based metaheuristics, population based metaheuristics, metaheuristics for multi-objective optimization, hybrid metaheuristics and parallel metaheuristics. However, the underlying theme of all metaheuristic algorithms in general consists of three characteristics: intensification, diversification and selection. The intensification process searches locally and more intensively to find a better solution within an already explored region. In contrast, diversification tends to search unexplored feasible regions. Finally, during the selection process old solutions are replaced by new solutions that improve the objective function. To increase the speed of convergence, the selection process sometimes involve passing the best solution from one iteration (generation) to the next, a step known as *elitism*. Proper functioning of the three characteristics is important for any metaheuristic algorithm and consequently for the quality of the solution to an optimization problem.

The above mentioned characteristics i.e., intensification, diversification and selection, can be easily identified in the firefly and PSO algorithms, described in Sections 3.1 and 3.2, respectively. The same characteristics can also be found in Genetic Algorithm (Holland, 1975; Jong, 1975), Simulated Annealing (Kirkpatrick et al., 1983; Goffe et al., 1994) and Harmony Search Algorithm (Geem et al., 2001; Geem, 2007, 2009, 2010) which were also studied in the context of quantile regression, but have not been presented to keep the paper within reasonable length. These may be obtained from the author. Details on all the above five metaheuristic algorithms may be found in Talbi (2009) and Yang (2010).

The metaheuristic algorithms, including the above mentioned five algorithms, come with the need to specify certain parameters, which if inappropriate can make the algorithm difficult to converge. They are slightly more time consuming relative to deterministic methods for a single run of the algorithm. However, time difference diminishes when compared to multiple runs of conventional algorithms starting at different values. Lastly, metaheuristics are approximation techniques that give approximate solutions, but the accuracy of the approximations is very high and in most cases practically identical to deterministic methods. To sum up, the use of metaheuristic algorithms should be carefully considered. It would be unwise to use metaheuristics in problems where existing techniques are easily applicable and give efficient and robust estimators.

The following two subsections briefly describe the firefly algorithm (FA) and the particle swarm optimization (PSO) algorithm. The firefly algorithm is the most recent and hence was included as the first algorithm to be studied in the paper. Amongst the remaining algorithms, the PSO algorithm is ubiquitous in the engineering literature and thus was chosen as the second algorithm. The corresponding sections also present the pseudo code to minimize the quantile regression objective function.

3.1. Firefly Algorithm

The firefly algorithm (FA), inspired by the flashing behavior of fireflies to attract mating partners, was developed by Xin She Yang while at Cambridge University in 2007. Scientists have observed that the ability of a firefly to attract a partner is directly proportional to the intensity of light, with less bright firefly moving towards the brighter firefly. This idea of difference in light intensities and consequent movement of fireflies is utilized in FA to optimize (maximize or minimize) the objective function. The objective function is evaluated at several simulated points and solution vectors are moved towards better performing solutions. Solutions are further randomized to search the entire parameter space and eventually find the global optimum.

Firefly algorithm has seen rapid development since its formulation in 2007 and several versions of FA now exist in the literature. In its original form, as developed by Yang (2008, 2009, 2010), the algorithm is based on the following assumptions. First, all fireflies are unisex, so they will be attracted to each other without any consideration of gender. Second, attractiveness is proportional to their brightness. So for any two fireflies, the less bright firefly will move towards the brighter firefly, however, brightness may decrease depending on the distance between fireflies and the media through which light travels. Third, brightness of a firefly is negatively affected by the complexity of the objective function.

The optimization process in FA essentially depends on the following two elements, namely, light intensity (I) and attractiveness. The light intensity (or simply brightness of light) of a firefly is directly (inversely) proportional to the objective function for a maximization (minimization) problem. There may be variations in light intensity, since intensity at a source may be different compared to the intensity as perceived by an adjacent firefly. The variation depends on three factors, light intensity at the source, distance between fireflies and the media. A light intensity which is brighter at the source will appear brighter to adjacent fireflies, *ceteris paribus*. At the same time, brightness of the light diminishes as the distance between the source and perceiving fireflies increases. Lastly, brightness depends on the media through which light travels. A dense or foggy media will absorb more light relative to a clear media, and hence the same light would appear less bright. The intensity of light, assuming the Gaussian form can be expressed as follows,

$$I(d) = I_0 \exp(-\gamma d^2),$$

where, d is distance, γ is the light absorption coefficient $\in [0, \infty)$, and I_0 is intensity of light at source.

The second component or attractiveness of a firefly is formulated to depend positively on the intensity of light. Since light intensity is proportional to the objective function, attractiveness κ, of a firefly can be defined as

$$\kappa = \kappa_0 \exp(-\gamma d^2), \tag{2}$$

where, κ_0 is the attractiveness at source (i.e. when $d = 0$). To calculate attractiveness at some other location a measure of distance d is necessary. One obvious choice is the Cartesian distance, but other measures of distance may be used depending on the context of the problem. Assuming the Cartesian measure, the distance between two fireflies i and j located at β_i and β_j can be calculated as,

$$d_{i,j} = \Big(\sum_{m=1}^{k} (\beta_{i,m} - \beta_{j,m})^2 \Big)^{1/2},$$

where, $\beta_{i,m}$ is the m-th component of β_i for the i-th firefly. By the second assumption, a less bright firefly will always move towards a brighter firefly. Therefore, if firefly j is the brighter

Table 1: Pseudo Code: Quantile Regression using the Firefly Algorithm

- Specify the quantile (τ), quantile objective function, number of fireflies (N_f), range of parameters and maximum number of iterations.
- Initialize $\beta = (\beta_1, \cdots, \beta_k)$, where β_j ($j = 1, \cdots, k$), is a column vector of size (N_f) and compute the quantile objective function (or light intensity).
- Specify randomization parameter (α), light absorption coefficient (γ) and maximum number of iterations.

While t $\leq$ maximum iterations

  for i = 1: N_f

    for l = 1: N_f

    If $f_i > f_l$, (for minimization lower is better)

    update β: Move firefly i ($\beta_{i^{th}\ row}$) towards firefly l ($\beta_{l^{th}\ row}$),

    end if

    end for l

  end for i

  Check β^{t+1} to lie within the parameter range.

  Update the quantile function (light intensity) for new beta values β^{t+1}.

  Compute the best solution when t reaches maximum iterations.

End while.

firefly, then firefly i moves towards firefly j, according to the following movement equation,

$$\beta_i = \beta_i + \kappa_0 \exp(-\gamma d^2)(\beta_j - \beta_i) + \alpha \epsilon_i, \tag{3}$$

where, α is the randomization parameter and ϵ_i is a vector of random numbers, drawn either from a Gaussian or Uniform distribution. It is worth pointing here that a subtle modification is introduced in this paper by constraining the positions of all fireflies i.e., values to lie within the range of parameters. This can be easily seen from the pseudo code presented in Table 1. The movement equation (3) has two important terms on the right hand side, attraction given by the second term and randomization given by the third term. The parameter γ in the attraction term is important for determining the speed of convergence of FA, and for all practical purposes lies between 0.1 to 10. See Yang (2010) for a discussion on determining γ based on the characteristic length of the system being optimized.

The firefly algorithm leads to two special cases based on the limiting values of γ (Yang, 2010). In the limit, when $\gamma \to 0$, the attractiveness as given by equation (2) reduces to constant κ_0. This implies that light intensities do not decrease and perceiving fireflies can always find the brightest firefly anywhere in the domain. In the context of optimization, this means that the global optima can be reached in one single step. Therefore, if β_i on the RHS of equation (3) is replaced by the global optima, then FA reduces to a special case of accelerated particle swarm optimization (APSO) described in Section 3.2. At the other extreme, when $\gamma \to \infty$, $\kappa(d) \to \delta(d)$ i.e., attractiveness reduces to a Dirac delta function and the FA reduces to the completely random search method. This is analogous to a situation where all fireflies move randomly without any visibility. As mentioned in Yang (2010), since FA lies between APSO and the completely random search method, it is possible to adjust parameters γ and α in FA to outperform both APSO and the random search method.

3.2. Particle Swarm Optimization

The Particle Swarm Optimization (PSO), developed by Kennedy and Eberhart (1995) is a population based metaheuristic algorithm inspired by swarm intelligence. Swarms of organisms (agents or particles), such as schools of fish and flocks of birds, act in a coordinated manner without any central control to find the best available food reserve. This behavior is imitated in the PSO algorithm to find the global best solution to an optimization problem. In the basic model, the movement or velocity of each particle, as represented in equation (4), is updated towards the global optimum by using a particle's own best position and the swarm's best known position. As the search proceeds and better solutions are found, the best position of each particle and the swarm are continuously updated. The process is expected (and almost always does) to move the swarm towards the global best position and yield the solution to an optimization problem.

The PSO algorithm is relatively simple with the main ingredient being the velocity vector ν, which in the standard case takes the following form,

$$\nu_i^{t+1} = \nu_i^t + a\epsilon_1 \odot (\beta_i^* - \beta_i^t) + b\epsilon_2 \odot (\beta_g^* - \beta_i^t), \qquad (i = 1, \cdots, n) \tag{4}$$

where, β_i^t is the value (position) of the i^{th} row (particle i) of β at iteration t, β_i^* and β_g^* denotes a particle's own best and the global best position until iteration t, respectively. The parameters a and b are learning parameters or acceleration constants, ϵ_j $(j = 1, 2)$ are random vectors and the symbol $\odot$ denotes element wise multiplication. The updated velocity is then used to calculate the value of the parameter in period $t + 1$, as

$$\beta_i^{t+1} = \beta_i^t + \nu_i^{t+1}.$$

Table 2: Pseudo Code: Quantile Regression using the APSO algorithm

- Specify the quantile (τ), quantile objective function, range of parameters, swarm size (N_{ss}) and maximum number of iterations.
- Initialize $\beta = (\beta_1, \cdots, \beta_k)$, where β_j $(j = 1, \cdots, k)$ is a column vector of size N_{ss}. Compute the quantile objective function for each row of β.

While t $\leq$ maximum iterations

Generate a swarm of new β's as

$$\beta_{i,j}^{t+1} = a(\epsilon - 0.5) + b * \beta_{g,j}^{*} + (1 - b) * \beta_{i,j}^{t},$$

where a, b $\in$ (0, 1), $\epsilon \sim$ U(0, 1), i = (1, $\cdots$, N_{ss}) and $(j = 1, \cdots, k)$.
Ensure new beta values lie within the range of parameters.
for l = 1: N_{ss}

Compute the function for β_l^{t+1} and β_l^t.
If β_l^t reduces the function value, accept it
else reject it, **end if**

end for l
Get β and function values for the entire swarm. Compute swarm's best value and denote it as β_g^*, a row vector.

End while.

The velocity ν_i can take any value but is usually specified to lie in some range $[0, \nu_{max}]$.

The standard form of the velocity vector has been extended in several ways to increase convergence towards the global optimum. One such formulation yields the accelerated particle swarm optimization (APSO) algorithm (Yang, 2010), in which the velocity is updated as,

$$\nu_i^{t+1} = \nu_i^t + a\,(\epsilon - 0.5) + b\,(\beta_g^* - \beta_i^t),$$

where, $\epsilon \in [0, 1]$ is a random variable. Convergence can be further increased by removing the velocity completely and using a particle's own position and the swarm's best position to update the parameter in a single step as,

$$\beta_i^{t+1} = a\,(\epsilon - 0.5) + b\,\beta_g^* + (1 - b)\,\beta_i^t,$$

where, a, b lies between $(0, 1)$. This version of the APSO algorithm has been used to estimate the model in Section 5 with the modification that values are constrained to lie within the range of parameters. The modification, although simple can play an important role in certain applications. A pseudo code is presented in Table 2.

The PSO algorithm in general is very efficient and leads to quicker convergence relative to other simulative algorithms that require encoding and decoding of values such as genetic algorithm. This is partially true since particles or agents in PSO continually share information about the best available position. Besides, the PSO algorithm like other members of meta-heuristic algorithms does not require information on the gradient and hence it can optimize functions which may be partially linear, irregular, multimodal, noisy or non-stationary.

4. Comparative Studies

This section uses the firefly algorithm (FA) described in Section 3.1 to estimate quantile regression models in two studies and compares the estimates obtained from FA to those obtained from the IP algorithm (for linear model) or KP algorithm (for nonlinear model). Such a comparison highlights the advantages and flexibility offered by FA relative to IP and KP algorithms. The first study models the 3-month US treasury bill rate, a common measure of short term interest rate, as a first order quantile autoregressive model or QAR(1) model; and the second study jointly estimates the parameters of the conditional scale ARCH model (Koenker and Zhao, 1996) based on a simulated data.

4.1. QAR Model

Short term interest rate is an important instrument of monetary policy and many researchers have attempted to model interest rate. In this regard, a popular modelling choice has been the first order autoregressive model, which in the quantile regression setting gives rise to first order quantile autoregressive model. Most quantile autoregressive (QAR) models suffer from the quantile crossing problem, also known as comonotonicity problem, which typically occurs in the outlying region of the design space. Given that the short term interest rate (U.S. Federal Funds rate or the 3-month treasury bill rate) was at an all-time low during the recent recession, the QAR(1) model may suffer from quantile crossing problem. Therefore, the objective of this exercise is threefold: first, to compare estimates of the QAR(1) model obtained from the firefly and IP algorithms; second, to look into potential quantile crossing problem due to low values of the short term interest rate observed during the recent recession; and third, to demonstrate the use of bounds in linear models. The bounds, although not necessary for this model, are placed to emphasize the ease of doing so compared to considerable difficulty in the IP algorithm.

The QAR(1) model of interest rate used in this application takes the form

$$y_t|\tau = \beta_1(\tau) + \beta_2(\tau)y_{t-1}, \tag{5}$$

where, $\tau \in (0,1)$ denotes the quantile and y_t denotes the short term interest rate in period t. The model (5) allows both the intercept (β_1) and slope (β_2) parameter to be quantile dependent and hence is capable of altering the location, scale and shape of the conditional densities (Koenker, 2005). In the present study, short term interest rate is represented by the 3-month treasury bill rate. Data is obtained from FRED (Federal Reserve Bank of St. Louis) for the time period January 1975 to December 2011.

The firefly algorithm is employed to estimate the model (5) with the following parameter specifications: maximum number of iterations = 5,000, number of fireflies = 25, $\alpha = 0.2$, $\gamma = 1$, and, bounds on intercept and slope as $(-10, 10)$ and $(-2, 2)$, respectively. Table 3 presents estimates at 19 quantiles from 0.05 to 0.95 with increments of 0.05. Results show that intercepts are very close to zero and statistically insignificant at all quantiles.

Table 3: Parameter estimates, standard errors and estimated function values of the QAR(1) model using the firefly algorithm (FA). The standard errors are calculated based on 100 bootstrap samples.

Quantiles(τ)	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{f}$
0.05	−0.09	0.07	0.908	0.01	22.66
0.10	−0.03	0.45	0.933	0.07	32.94
0.15	−0.02	0.50	0.945	0.07	40.20
0.20	0.00	0.69	0.964	0.10	45.35
0.25	−0.01	0.54	0.974	0.08	48.85
0.30	−0.01	0.51	0.982	0.08	51.30
0.35	−0.01	0.40	0.987	0.06	53.01
0.40	−0.01	0.24	0.994	0.04	54.08
0.45	−0.01	0.15	0.998	0.03	54.38
0.50	0.00	0.17	1.002	0.03	54.15
0.55	0.00	0.26	1.009	0.04	53.35
0.60	0.00	0.29	1.014	0.04	51.91
0.65	0.00	0.40	1.018	0.06	49.85
0.70	0.00	0.62	1.025	0.08	47.17
0.75	0.01	0.56	1.029	0.08	43.86
0.80	0.02	0.39	1.032	0.06	39.86
0.85	0.02	0.51	1.045	0.07	34.87
0.90	0.01	0.48	1.064	0.08	28.19
0.95	0.00	0.05	1.125	0.12	18.24

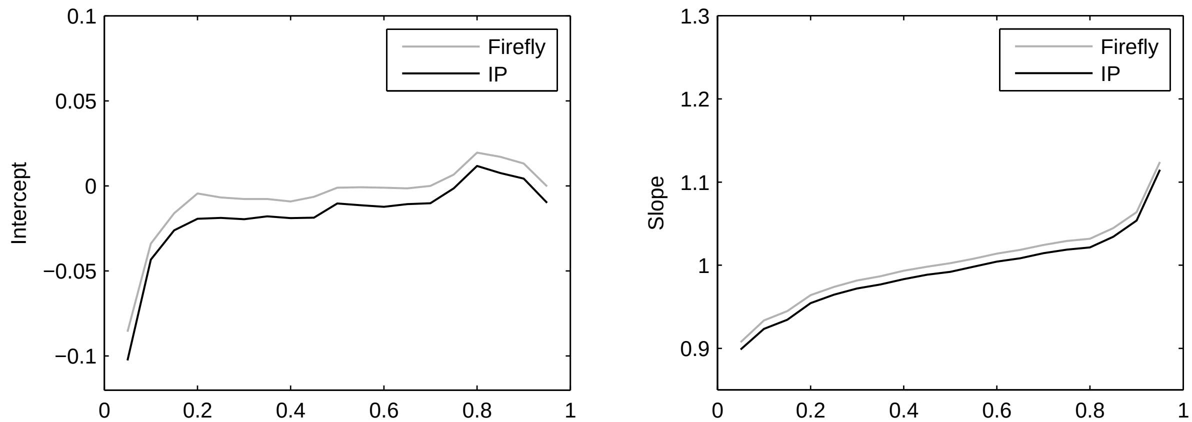


Figure 2: Parameter estimates of QAR(1) model using the firefly and interior point (IP) algorithm. The IP solutions are perturbed by subtracting 0.01 for better visualization.

On the other hand, slope estimates continuously rise as τ increases and are statistically significant at all quantiles. The estimated function value, shown in the last column, initially increases, reaches its maximum and then continually decreases. Standard errors based on 100 bootstrap samples are also presented in Table 3. Here, it is worth mentioning that quantile estimators do not satisfy the assumption of smooth function model because the quantile objective function is not first derivative continuous and hence *asymptotic refinements* are not possible without smoothing the objective function (Horowitz, 998b, 2001).

To examine the validity of the estimates obtained using FA, the QAR(1) model was also estimated using the IP algorithm. The estimates obtained from the two algorithms are plotted in Figure 2, where estimates from IP algorithm are perturbed by subtracting 0.01 for better visualization. Solutions are practically identical which provides evidence that FA can efficiently estimate the QAR(1) model and more generally, linear quantile regression models.

The estimated conditional quantile functions for the model (5) are shown in Figure 3. Although, difficult to visualize from the figure, estimated conditional quantile functions do not cross each other for interest rate between 0.25 to 16.30 (16.30 is the maximum observed interest rate). However, for values below 0.25, where interest rate have been for most part of the recent recession, there does appear some nonmonotonicity issues. Such nonmonotonicity is not uncommon due to degenerate behavior of interest rate at low values and can be solved by adding a quadratic component in the model (Koenker, 2005). In addition, the nonmonotonicity problem can possibly be resolved using parallel implementation of metaheuristic methods, but that remains worthy of another article. In the present study, my principal

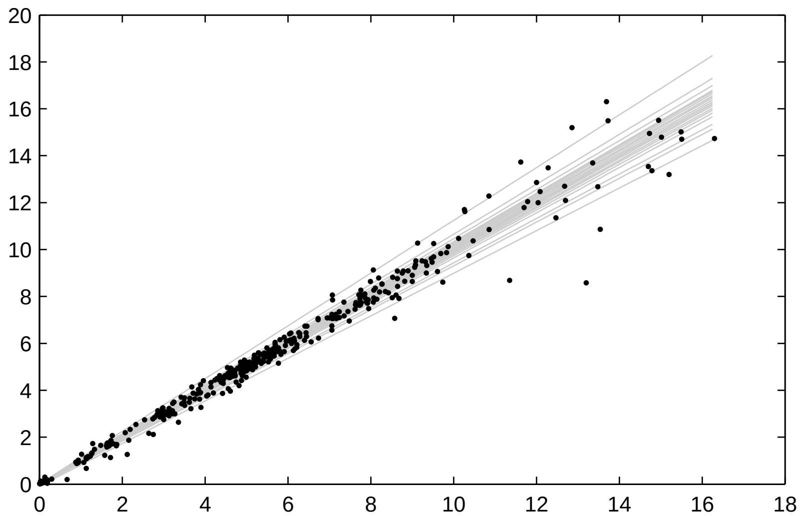


Figure 3: The AR(1) scatter plot of the U.S. 3-month treasury bill rate, superimposed with 19 equally spaced estimates of linear conditional quantile functions.

objective was to check the validity of the estimates from FA and hence nonmonotonicity problem is not explored further. Alternative methodologies that overcome nonmonotonicity issue may be found in (Takeuchi and Furuhashi, 2004) and (Chernozhukov et al., 2010).

4.2. The Conditional Scale ARCH Model

The autoregressive conditional heteroscedasticity (ARCH) model, introduced by Engle (1982), is a stochastic process y_t, where the autoregressive component is given by

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + \epsilon_t, \tag{6}$$

and the error term is modelled based on conditional variance as,

$$\epsilon_t = \sigma_t \nu_t = (\gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \cdots + \gamma_q \epsilon_{t-q}^2)^{1/2}\, u_t, \tag{7}$$

where, u_t is independently and identically distributed (i.i.d.) as N(0,1). The above model is typically denoted as ARCH (p, q) model, where p is the number of lags in the autoregressive structure and q is the number of lags in the error structure.

In a particular class of ARCH models, known as the conditional scale ARCH models (Koenker and Zhao, 1996), the autoregressive component given by equation (6) remains unaltered, but the error term is modelled as

$$\epsilon_t = (\gamma_0 + \gamma_1|\epsilon_{t-1}| + \cdots + \gamma_q|\epsilon_{t-q}|)\,\nu_t, \tag{8}$$

with $\gamma_0 > 0$, $(\gamma_1, \cdots, \gamma_q)' \in R_+^q$ and $\{\nu_t\}$ are i.i.d. random variables with zero mean and finite variance. In other words, the conditional scale σ_t is modelled as a linear function of lagged absolute residuals. This formulation is advantageous for non-Gaussian distributions because scale or standard deviation, is a more natural measure of dispersion compared to variance (Bickel and Lehmann, 1976). It also offers substantial advantages with respect to robustness of solutions as mentioned in articles including Bickel (1978); Carroll and Ruppert (1988) and Newey and Powell (1987). Given the advantages, many authors including Taylor (1987); Schwert (1989) and Nelson and Foster (1994) have also used the conditional scale ARCH model with a Gaussian distribution.

The conditional scale ARCH model given by equations (6) and (8), was studied in the quantile regression setting by Koenker and Zhao (1996). They derived the conditional quantile function to be the following expression,

$$Q_t\big(\tau|\mathcal{F}_{t-1}\big) = \alpha_0 + \sum_{i=1}^{p}\alpha_i y_{t-i} + \Big(\gamma_0 + \sum_{j=1}^{q}\gamma_j|\epsilon_{t-j}|\Big)F^{-1}(\tau), \tag{9}$$

where, quantile $\tau \in (0,1)$, $\mathcal{F}_{t-1}$ denotes information until period $t-1$ and F represents a distribution function for innovations $\{\nu_t\}$. To estimate the model, they proposed a two-step procedure which is analogous to the procedure used by Engle (1982). In the first step, the autoregressive parameters $\alpha = (\alpha_0, \alpha_1, \cdots, \alpha_p)'$ are estimated by $\sqrt{n}$-consistent estimators $\hat{\alpha}_n$, which under symmetric innovations can either be the least squares estimator, provided the moment conditions are satisfied (Koenker and Zhao (1996), p. 800), or the l_1 estimator provided the innovation density is continuous and strictly positive at the median. The $\sqrt{n}$-consistent estimators are then used to obtain residuals as $\hat{\epsilon}_t = y_t - X_t'\hat{\alpha}_n$, where $X_t = (1, y_{t-1}, \cdots, y_{t-p})'$. In the second step, the ARCH parameters $\gamma = (\gamma_0, \gamma_1. \cdots, \gamma_q)$ are estimated for a given value of τ by minimizing,

$$\hat{\gamma}(\tau, \hat{\alpha}_n) = \text{argmin}_\gamma \sum \rho_\tau(\hat{\epsilon}_t - \hat{Z}_t\gamma), \tag{10}$$

with constraints $\gamma_0 > 0$ and $\gamma_i \geq 0$ $(i = 1, \cdots, q)$, and where the vector $\hat{Z}_t$ consists of absolute values of the residuals i.e., $\hat{Z}_t = (1, |\hat{\epsilon}_{t-1}|, \cdots, |\hat{\epsilon}_{t-q}|)$. The method also holds for asymmetric distributions under the condition that $\sqrt{n}$-consistent estimators are obtained based on a quantile τ_0 such that $F^{-1}(\tau_0) = 0$.

The two-step estimation approach to conditional scale ARCH model, although conceptually straightforward, is unsatisfactory for two reasons. First, the two-step approach inflates the asymptotic variance of the second step estimators $\hat{\gamma}$'s, unless the innovation distribution is symmetric (Koenker and Zhao, 1996). Second, optimization of the second step objective function (10) subject to parameter restrictions $\gamma_0 > 0$ and $\gamma_i \geq 0$ $(i = 1, \cdots, q)$ can become quite detailed.

In order to overcome the problems associated with two-step estimation, Koenker and Zhao (1996) described a nonlinear quantile regression approach for jointly estimating the parameters (α, γ), but they did not estimate the model. The parameters (α, γ) can be jointly estimated using an algorithm based on interior point methods (Koenker and Park, 1996) or simply KP algorithm. However, the KP algorithm in its present form suffers from at least two drawbacks: it cannot incorporate parameter restrictions and is highly dependent on initial values. These two problems can be easily undone using metaheuristic algorithms, such as the firefly algorithm (FA). The FA can jointly estimate the parameters (α, γ) and can be easily modified to incorporate bounds on parameters, as done in the simulation study. In fact, metaheuristic algorithms actually become more potent under restrictions since they reduce the search space of parameters.

The following subsection presents a Monte Carlo simulation study on the conditional scale ARCH model, where the parameters (α, γ) are jointly estimated using the FA and KP algorithms. Estimation via KP algorithm uses the *nlrq* function from the *quantreg* package (R software), where the *nlrq* function is based on interior point ideas as described in Koenker and Park (1996). It should be noted that although results presented below are based on a single random sample, similar results were obtained for several other data samples generated from the conditional scale ARCH (1,1) model.

4.2.1. A Simulation Study

Consider a simple case of the conditional scale ARCH model,

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t, \tag{11}$$

$$\epsilon_t = (\gamma_0 + \gamma_1 |\epsilon_{t-1}|)\, \nu_t, \tag{12}$$

where, $|\alpha_1| < 1$, $\gamma_0 > 0$, $0 \leq \gamma_1 < 1$ and the innovations $\{\nu_t\}$ are i.i.d. as standard normal with cumulative distribution function Φ. For the simplified ARCH model, the conditional quantile function (9) reduces to,

$$Q_t(\tau|\mathcal{F}_{t-1}) = \alpha_0 + \alpha_1 y_{t-1} + \big(\gamma_0 + \gamma_1 |y_{t-1} - \alpha_0 - \alpha_1 y_{t-2}|\big)\Phi^{-1}(\tau). \tag{13}$$

Suppose, interest lies in the nonlinear quantile regression estimator,

$$\hat{\theta}_n(\tau) = \text{argmin}_\theta \sum \rho_\tau\big(y_t - \eta_t(\theta)\big)$$

where,

$$\eta_t(\theta) = \theta_0 + \theta_1 y_{t-1} + \theta_3 |y_{t-1} - \theta_2 - \theta_1 y_{t-2}|. \tag{14}$$

Koenker and Zhao (1996) show that the nonlinear estimator $\hat{\theta}_n(\tau)$ converges almost surely to $\theta(\tau) = (\theta_1, \theta_2, \theta_3, \theta_4) = \big(\alpha_0 + \gamma_0 F^{-1}(\tau),\, \alpha_1,\, \alpha_0,\, \gamma_1 F^{-1}(\tau)\big)$, where $\theta(\tau)$ uniquely minimize the expectation of the quantile objective function. Therefore, for any given value of the parameter vector (α, γ) and quantile τ, the minimum of the expected quantile objective function is uniquely known.

To perform a Monte Carlo simulation study, 1000 observations are generated from the conditional scale ARCH model (11 & 12), assuming $\nu_t \sim N(0, 1)$ and the following parameter specifications: $\alpha_0 = 3$, $\alpha_1 = 0.7$, $\gamma_0 = 2$, $\gamma_1 = 0.4$. The model is estimated using the firefly algorithm as described in Section 3.1, with number of fireflies $N_f = 20$, randomization parameter $\tilde{\alpha} = 0.2$ ($\sim$ is used over FA parameters to avoid confusion with ARCH parameters), light absorption coefficient $\tilde{\gamma} = 2$ and maximum number of iteration equalling 12,000.

In addition, parameter information on the conditional scale ARCH model was utilized to place bounds on $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$. This was done in reverse order since bounds placed on θ_3 are utilized to place bounds on θ_1. The conditional scale ARCH model requires parameter $\gamma_1 \in [0, 1)$, which implies that the support for $\theta_4 = \gamma_1 \Phi^{-1}(\tau)$ is as follows,

$$\theta_4 \in \begin{cases} \big[0, \Phi^{-1}(\tau)\big) & \text{if } \tau \geq 0.5 \\ \big[\Phi^{-1}(\tau), 0\big) & \text{otherwise.} \end{cases}$$

Parameter $\theta_3 = \alpha_0$ was assumed to lie between [-5, 5] and increasing the interval does not affect the results. Model specification also requires parameter $\theta_2 = \alpha_1$ to lie in the open interval (-1,1). Lastly, bounds on $\theta_1 = \alpha_0 + \gamma_0\, \Phi^{-1}(\tau)$ were specified as $[-5 + \Phi^{-1}(1-\omega), 5 + \Phi^{-1}(\omega)]$, where $\omega = 0.999$ and γ_0 was assumed to be 1.

The conditional scale ARCH model (11 & 12) was estimated for $\tau \in (0.05, 0.95)$ with increments of 0.05. Table 4 presents the parameter values along with their estimates. Results

Table 4: The conditional scale ARCH model: The table presents parameter values, estimates and standard error estimates of $\theta(\tau) = (\theta_1, \theta_2, \theta_3, \theta_4)$, where, $\theta_1 = \alpha_0 + \gamma_0 F^{-1}(\tau)$, $\theta_2 = \alpha_1$, $\theta_3 = \alpha_0$ and $\theta_4 = \gamma_1 F^{-1}(\tau)$. Standard errors are obtained using 100 bootstrap samples. The last two columns give values of the quantile function and its estimates, respectively.

Quantiles (τ)	θ_1	$\hat{\theta}_1$	$se(\hat{\theta}_1)$	θ_2	$\hat{\theta}_2$	$se(\hat{\theta}_2)$	θ_3	$\hat{\theta}_3$	$se(\hat{\theta}_3)$	θ_4	$\hat{\theta}_4$	$se(\hat{\theta}_4)$	f	$\hat{f}$
0.05	−0.29	0.35	0.99	0.7	0.65	0.12	3.0	3.51	3.28	−0.66	−0.84	0.14	301	298
0.10	0.44	0.76	0.62	0.7	0.67	0.10	3.0	3.53	3.26	−0.51	−0.55	0.11	518	516
0.15	0.93	1.22	0.48	0.7	0.67	0.09	3.0	3.36	3.17	−0.41	−0.48	0.10	690	688
0.20	1.32	1.73	0.55	0.7	0.66	0.09	3.0	3.78	3.24	−0.34	−0.38	0.10	830	829
0.25	1.65	1.90	0.50	0.7	0.67	0.09	3.0	3.82	3.16	−0.27	−0.32	0.09	944	941
0.30	1.95	2.29	0.36	0.7	0.68	0.08	3.0	3.89	3.11	−0.21	−0.28	0.08	1,033	1,029
0.35	2.23	2.87	0.26	0.7	0.65	0.08	3.0	4.63	2.98	−0.15	−0.27	0.08	1,098	1,096
0.40	2.49	3.24	0.06	0.7	0.65	0.06	3.0	4.65	2.57	−0.10	−0.25	0.08	1,155	1,146
0.45	2.75	3.05	0.01	0.7	0.67	0.03	3.0	4.46	1.74	−0.05	−0.12	0.05	1,188	1,182
0.50	3.0	2.97	0.00	0.7	0.69	0.01	3.0	2.97	0.00	0	0	0	1,202	1,201
0.55	3.25	3.24	0.02	0.7	0.64	0.04	3.0	−1.77	1.74	0.05	0.11	0.02	1,194	1,189
0.60	3.51	3.78	0.01	0.7	0.64	0.03	3.0	0.97	0.50	0.10	0.15	0.01	1,164	1,161
0.65	3.77	4.06	0.01	0.7	0.66	0.05	3.0	1.97	0.21	0.15	0.16	0.02	1,113	1,110
0.70	4.05	4.23	0.01	0.7	0.68	0.05	3.0	2.26	0.43	0.21	0.20	0.02	1,043	1,039
0.75	4.35	4.49	0.05	0.7	0.67	0.05	3.0	2.44	0.33	0.27	0.29	0.02	952	948
0.80	4.68	4.84	0.12	0.7	0.66	0.05	3.0	2.49	0.19	0.34	0.35	0.02	837	836
0.85	5.07	5.21	0.17	0.7	0.67	0.05	3.0	2.92	0.31	0.41	0.51	0.03	695	693
0.90	5.56	5.52	0.03	0.7	0.69	0.14	3.0	2.89	1.41	0.51	0.55	0.02	518	518
0.95	6.29	6.11	0.01	0.7	0.71	0.12	3.0	2.90	1.00	0.66	0.64	0.02	303	302

Table 5: The conditional scale ARCH model: The table presents parameter and function values, and their estimates using the firefly algorithm and the KP algorithm (Koenker and Park, 1996). The estimates using the KP algorithm, denoted $\ddot{\theta}$'s , are based on *nlrq* function in *quantreg* package, R software.

Quantiles (τ)	θ_1	$\hat{\theta}_1$	$\ddot{\theta}_1$	θ_2	$\hat{\theta}_2$	$\ddot{\theta}_2$	θ_3	$\hat{\theta}_3$	$\ddot{\theta}_3$	θ_4	$\hat{\theta}_4$	$\ddot{\theta}_4$	f	$\hat{f}$	$\ddot{f}$
0.05	−0.29	0.35	0.57	0.7	0.65	0.61	3.0	3.51	3.95	−0.66	−0.84	−0.70	301	298	300
0.10	0.44	0.76	0.73	0.7	0.67	0.67	3.0	3.53	3.52	−0.51	−0.55	−0.55	518	516	517
0.15	0.93	1.22	1.23	0.7	0.67	0.71	3.0	3.36	3.32	−0.41	−0.48	−0.55	690	688	688
0.20	1.32	1.73	2.08	0.7	0.66	0.63	3.0	3.78	4.17	−0.34	−0.38	−0.41	830	829	829
0.25	1.65	1.90	1.82	0.7	0.67	0.68	3.0	3.75	3.75	−0.27	−0.32	−0.31	944	941	941
0.30	1.95	2.29	2.55	0.7	0.68	0.66	3.0	3.89	4.22	−0.21	−0.28	−0.31	1,033	1,029	1,029
0.35	2.23	2.87	2.82	0.7	0.65	0.64	3.0	4.63	4.60	−0.15	−0.27	−0.25	1,098	1,096	1,096
0.40	2.49	3.24	3.23	0.7	0.65	0.65	3.0	4.65	4.65	−0.10	−0.25	−0.25	1,155	1,146	1,146
0.45	2.75	3.05	2.61	0.7	0.67	0.64	3.0	4.46	−4.99	−0.05	−0.12	0.06	1,188	1,182	1,186
0.50	3.0	2.97	3.09	0.7	0.69	0.63	3.0	2.97	−0.24	0	0	0.11	1,202	1,201	1,197
0.55	3.25	3.24	3.38	0.7	0.64	0.63	3.0	−1.77	−0.38	0.05	0.11	0.13	1,194	1,189	1,189
0.60	3.51	3.78	3.87	0.7	0.64	0.63	3.0	0.97	0.73	0.10	0.15	0.15	1,164	1,161	1,161
0.65	3.77	4.06	4.09	0.7	0.66	0.66	3.0	1.97	1.79	0.15	0.16	0.15	1,113	1,110	1,110
0.70	4.05	4.23	4.27	0.7	0.68	0.67	3.0	2.26	2.29	0.21	0.20	0.20	1,043	1,039	1,039
0.75	4.35	4.49	4.47	0.7	0.67	0.67	3.0	2.44	2.43	0.27	0.29	0.30	952	948	948
0.80	4.68	4.84	4.80	0.7	0.66	0.67	3.0	2.49	2.48	0.34	0.35	0.36	837	836	835
0.85	5.07	5.21	5.25	0.7	0.67	0.66	3.0	2.92	2.98	0.41	0.51	0.50	695	693	693
0.90	5.56	5.52	5.69	0.7	0.69	0.67	3.0	2.89	2.98	0.51	0.55	0.57	518	518	517
0.95	6.29	6.11	6.12	0.7	0.71	0.71	3.0	2.90	2.91	0.66	0.64	0.64	303	302	302

show that the estimated values are very close to parameter values, especially for θ_2 and θ_4. For θ_1, the estimates closely follow the true values and increase with the quantiles. Lastly, estimates for θ_3 are all around 3 except at $\tau = 0.55$. As an observation, the estimated values are all greater (lower) than 3 for $\tau < 0.5$ ($\tau > 0.5$). Parameter estimation at the median ($\tau = 0.5$) deserve special attention. When $\tau = 0.5$, the conditional quantile function (13) only depends on autoregressive parameters since $\Phi^{-1}(0.5) = 0$. Hence, only autoregressive parameter estimates are presented at the median.

Table 4 also presents the standard errors of the estimates obtained from 100 bootstrap samples. As mentioned earlier, quantile estimators do not satisfy the assumption of smooth function model because the objective function is not first derivative continuous. Hence, *asymptotic refinements* are not possible without smoothing the objective function (Horowitz, 998b, 2001). In addition, bootstrap method does not consistently estimate the distribution of an estimator when the true parameter value is on the boundary of the parameter space (Andrews, 2001; Horowitz, 2001). In such cases, amongst other techniques, alternative resampling procedures such as *subsampling* is suggested in the literature (Wu, 1990; Politis and Romano, 1994; Andrews, 2001; Horowitz, 2001). However, for the conditional scale ARCH example the true parameter value lies inside the boundary and hence the standard errors computed from bootstrap do not suffer from inconsistency. Finally, the last two columns in Table 4 present the true and estimated values of the quantile objective functions. As seen from the two columns, the estimates are very close to the true function values.

To facilitate comparison, the conditional scale ARCH model was also estimated using the KP algorithm, implemented in the *nlrq* function of *quantreg* package (R software). Table 5 presents the estimated parameter and function values along with the true parameter and function values. It is clear from Table 5 that estimates obtained from the FA and KP algorithm are approximately same. However, there are two important drawbacks with the KP algorithm. First, it fails to respect bounds on the original parameter γ_1. Specifically, the estimated value for θ_4 were obtained as 0.06 at $\tau = 0.45$ and 0.13 at $\tau = 0.55$. The corresponding values for γ_1 are -0.48 and 1.06, which clearly lie outside the range $[0, 1)$. Second, the KP algorithm is very sensitive to initial values and occasionally fails to converge. In the present application, the following initial values were employed: $\theta_{s1} = (1.5, 0.5, 2, -0.3)$ for $\tau < 0.5$ except $\tau = 0.05$, $\theta_{s2} = (1.5, 0.5, 2, 0.3)$ for $\tau \geq 0.5$ and $\theta_{s3} = (-1, 0.5, 2, -0.3)$ for $\tau = 0.05$. If θ_{s1} is used as the initial value at $\tau = 0.05$, the algorithm fails and simply returns the initial value. Similarly, if the initial value θ_{s1} is replaced by $\theta_{s4} = (1.5, 0.1, 2, -0.3)$ estimates at $\tau = 0.10$ change from $(0.73, 0.67, 3.52, -0.55)$ to $(1.40, 0.44, -11.80, 0.00)$. Similar problems were observed at few other extreme quantiles.

In contrast, the modified FA incorporated the boundary information into the algorithm and did not encounter initial value problem across quantiles. The sole disadvantage of FA relates to time consumption and on average takes about 34 seconds (for 12,000 iterations) for a single run of the algorithm compared to 1.5 seconds of the KP algorithm. However, time difference diminishes significantly when compared to multiple runs of the KP algorithm to check for stability of solution to different starting values. The FA does not require restarting the algorithm since it simulates different values from the entire parameter space in its search for a global solution.

In summary, FA (a metaheuristic algorithm) can impose bounds on parameters and efficiently estimate quantile regression parameters for complicated models, such as the conditional scale ARCH model, with simply a change in the objective function. The FA, unlike KP algorithm, is also less susceptible to the initial value problem. Besides, many variants of FA and other metaheuristic algorithms do exist in the engineering literature and can be used to increase the efficiency and speed of the algorithm. These variants are promising and remain a source of possible future research with respect to quantile regression and to other estimation problems in general.

5. Application: Consumption Behavior

The economic recession of December 2007 - June 2009, triggered by the subprime mortgage crisis, has altered the path of U.S. macroeconomic indicators such as output, employment, consumption and investment. Amongst other variables, the list possibly includes consumption behavior of individuals. Consumption behavior is extremely important for the U.S. economy since consumption comprises about 70% of gross domestic product (GDP). Hence, a change in consumption behavior can have important policy implications for the U.S. economy. This provides the motivation to study and compare consumption behavior, represented by the marginal propensity to consume, during 2010 (the year following recession) relative to 2005 (a non-recession year).

Marginal propensity to consume or MPC is defined as the ratio of change in consumption to a unit change in income. MPC by definition lies between 0 and 1 and is known to vary considerably in a cross section of population. Quantile regression can capture such variation by estimating MPC at different quantiles. The need to do quantile regression coupled with MPC $\in (0, 1)$ provides a reasonable argument to use a metaheuristic algorithm.

The study employs a small consumption model where consumption (in \$'000) for each consumer unit[2] (CU) is regressed on the following covariates: income (in \$'000), family size or number of members, number of earners and an urban dummy, for Q1 - Q4 of 2005 and 2010. The model is estimated *via* APSO algorithm (Section 3.2) at 19 quantiles $\in (0.05, 0.95)$ with 10,000 iterations for each quantile. Other parameters of the APSO algorithm are specified as follows: swarm size $N_{ss} = 40$, randomization parameter a = 0.4, learning parameter b = 0.5 and range of parameters specified as $(-5, 5)$, $(0, 1)$, $(0, 5)$, $(-3, 3)$ and $(-2, 2)$. While bounds on MPC is obtained from consumption theory, bounds on remaining parameters were based on economic intuition. The variable number of members and consumption is expected to have a positive relationship as consumption would typically increase with number of members in the consumption unit, hence the lower and upper bounds were specified as (0, 5). Similarly, number of earners and urban presence is expected to have a positive relationship with consumption. However, to allow for the possibility of a negative relationship, bounds were chosen to be $(-3, 3)$ and $(-2, 2)$, respectively. The bounds placed on the parameters can be increased, without affecting the estimates of the parameters (algorithm was tested with different bounds). However, some judgement needs to be used since parameter estimate greater (lower) than 20 (-20) only indicates that the variables are badly scaled. Further, if a researcher is not comfortable with bounds on parameters, other than MPC, they can be removed by modifying the algorithm to impose restrictions on a subset of parameters, a relatively straightforward procedure.

5.1. Data

The study utilizes a sub-sample of the Quarterly Interview Survey (QIS) data collected for the Bureau of Labor Statistics (BLS) by the U.S. Census Bureau. A sub-sample of the QIS data is utilized due to the following reason: BLS conducts interview using a rotating scheme, where individuals report their last 3 months of expenditure from month of the interview. For example, if an interview takes place in May, then expenditure for the last three months i.e. February, March and April is reported. Since my objective is to look at consumption

[2]A consumer unit consists of any of the following: (1) All members of a particular household who are related by blood, marriage, adoption, or other legal arrangements; (2) a person living alone or sharing a household with others or living as a roomer in a private home or lodging house or in permanent living quarters in a hotel or motel, but who is financially independent; or (3) two or more persons living together who use their incomes to make joint expenditure decisions. Financial independence is determined by spending behavior with regard to the three major expense categories: Housing, food, and other living expenses. To be considered financially independent, the respondent must provide at least two of the three major expenditure categories, either entirely or in part.

behavior within a calendar quarter, I use data corresponding to the interviews that took place in January, April, July and October. This allows me to collect data by quarters Q1 (January-March), Q2 (April-June), Q3 (July-September) and Q4 (October-December).

The dependent variable in the model, consumption, measured by total expenditure in the previous quarter is the BLS series TOTEXPPQ. The list of components in TOTEXPPQ are shown in Table A.9. Here, the limitations of consumption as defined by TOTEXPPQ deserve attention. The QIS collects data on major items of expense which respondents are expected to recall for 3 months or longer. In practice, the QIS collects detailed data on an estimated 60 to 70 percent of total household expenditures. In addition, global estimates are obtained for food and other selected items that account for an additional 20 to 25 percent of total expenditures. The QIS does not collect expenses for housekeeping supplies, personal care products, and nonprescription drugs, which contribute about 5 to 15 percent of total expenditures. Hence, as per QIS about 80 to 95 percent of total expenditures are covered in the interview survey. However, for the sample included in the study, consumption as percentage of quarterly total expenditures is around 75 percent. The partial account of total expenditure will later reflect in lower value of estimated MPC across quantiles.

Income as an independent variable is necessary to get an estimate of MPC and hence included in the model. Income measure used is average biannual income or last 12 months of income divided by 2, where last years income is the BLS series FINCATXM, obtained as FINCBTXM - TOTTXPDM. FINCBTXM is income received by all CU members before tax deductions (see Table A.10) and TOTTXPDM is the amount of personal taxes paid by CU in the past 12 months. All observations with biannual income less than \$500 or less than consumption are deleted. Biannual income is used as a measure of income in order to increase the sample size since consumption during Q4 is greater than average quarterly income for more than $\frac{3}{4}$th of the sample. Admittedly, the definition of income is not foolproof and a measure of permanent income would be desirable, however, given that QIS is a cross section survey with each quarter having different individuals, a measure of permanent income is not possible. To facilitate direct comparison across quarters, biannual income is also used as a measure of income for other quarters.

The remaining independent variables included in the model are family size, number of earners and a dummy for urban area. The variable family size, represented by number of members in CU, comprises the BLS series FAM_SIZE. Number of earners in CU is the BLS series NO_EARNR. The last covariate is a dummy for urban area (= 1, if urban and 0 otherwise) and corresponds to the BLS series BLS_URBN.

Table 6: Data Summary for Quarterly Consumption ($), Biannual Income ($), Number of Members and Number of Earners

	2005				2010			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Consumption ($)								
Mean	9,411	9,521	9,980	9,821	10,250	10,292	10,402	10,493
Std. Dev.	6,738	6,782	7,389	6,844	6,685	6,630	7,326	6,836
Minimum	505	767	583	512	509	706	602	515
Maximum	67,027	65,679	65,400	72,310	60,371	54,528	63,960	56,315
B. Income ($)								
Mean	25,625	26,134	25,968	26,579	28,078	28,413	27,459	27,971
Std. Dev.	18,099	18,822	18,666	19,620	19,430	20,181	19,480	20,032
Minimum	1,040	1,800	1,028	1,000	597	1,317	1,500	833
Maximum	153,384	160,000	146,088	235,303	135,165	126,972	126,180	126,387
No. of Members								
Mean	2.6	2.6	2.6	2.5	2.5	2.5	2.5	2.6
Std. Dev.	1.6	1.5	1.6	1.5	1.5	1.5	1.6	1.6
Minimum	1	1	1	1	1	1	1	1
Maximum	15	15	15	14	11	11	11	11
No. of Earners								
Mean	1.4	1.4	1.4	1.4	1.3	1.3	1.3	1.3
Std. Dev.	1	1	1	1	0.9	1	1	0.9
Minimum	0	0	0	0	0	0	0	0
Maximum	8	8	7	6	6	8	8	7

Table 7: Sample Size and Number of Observations in Urban Area

	2005				2010			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Sample Size	2,201	2,134	2,095	2,143	1,975	1,889	1,886	1,866
Obs. in Urban Area	2,036	1,984	1,964	2,002	1,857	1,776	1,787	1,766

Tables 6 and 7 present a summary of the variables. It is clear from Table 6 that mean consumption and mean biannual income increased in 2010 relative to 2005 for all quarters. However, the values are close and differences would be smaller when adjusted for inflation. The maximum values for consumption and biannual income are lower in 2010 relative to 2005 for all quarters. Table 6 also presents a summary for number of members and number of earners in CU. The maximum number of members in a CU is lower in 2010 relative to 2005 for all quarters. Similarly, the mean number of earners in a CU is lower in 2010 relative to 2005 for all quarters, possibly reflecting the unemployment situation following the recent recession. Table 7 presents the sample size by quarters and number of CUs in the samples that reside in urban area. As seen from the table, the samples mostly contain CUs that live in urban area.

5.2. Quantile Estimates

In order to get an initial idea about consumption behavior across quarters and years, the consumption model is estimated at the median ($\tau = 0.5$). Table 8 presents the median estimates and the standard errors (in parentheses) based on 100 bootstrap samples, that

Table 8: Parameter Estimates and Standard Errors (within parenthesis) at the Median

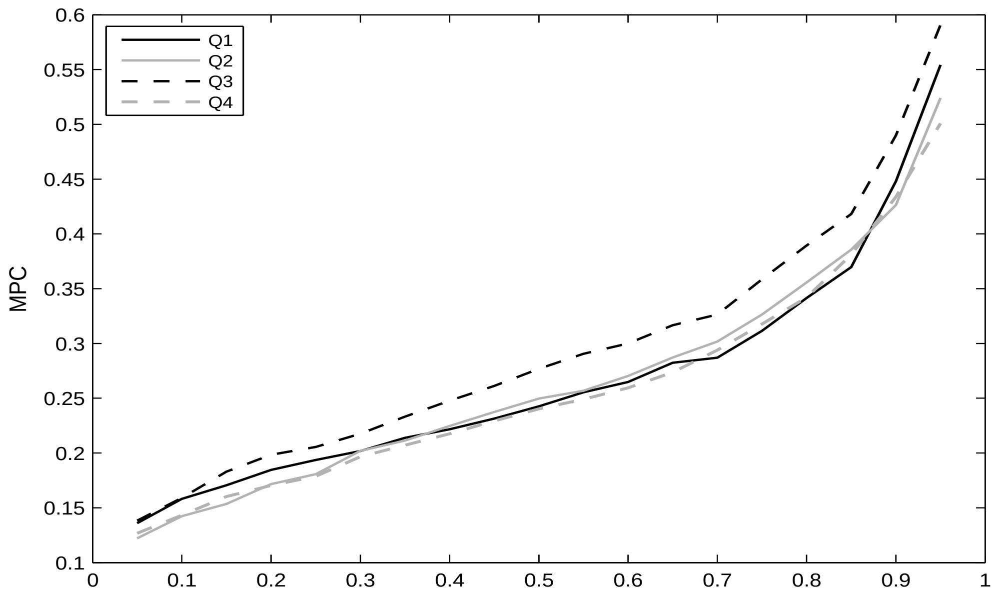
	2005				2010			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Intercept	1.478	1.625	0.659	1.441	1.578	2.072	1.313	2.034
	(0.287)	(0.278)	(0.346)	(0.252)	(0.288)	(0.274)	(0.374)	(0.245)
Biannual Income	0.243	0.250	0.277	0.240	0.234	0.233	0.259	0.238
	(0.012)	(0.009)	(0.009)	(0.009)	(0.007)	(0.008)	(0.009)	(0.009)
No. of Members	0.320	0.149	0.360	0.237	0.318	0.238	0.391	0.337
	(0.077)	(0.081)	(0.062)	(0.068)	(0.063)	(0.067)	(0.081)	(0.053)
No. of Earners	0.093	0.107	−0.026	0.199	0.152	0.060	−0.138	0.066
	(0.154)	(0.132)	(0.164)	(0.113)	(0.129)	(0.135)	(0.148)	(0.110)
Urban dummy	0.050	0.099	0.538	0.345	0.388	0.124	0.478	0.138
	(0.259)	(0.259)	(0.328)	(0.220)	(0.258)	(0.269)	(0.383)	(0.251)

Note: To get quarterly MPC double the above MPC estimates.

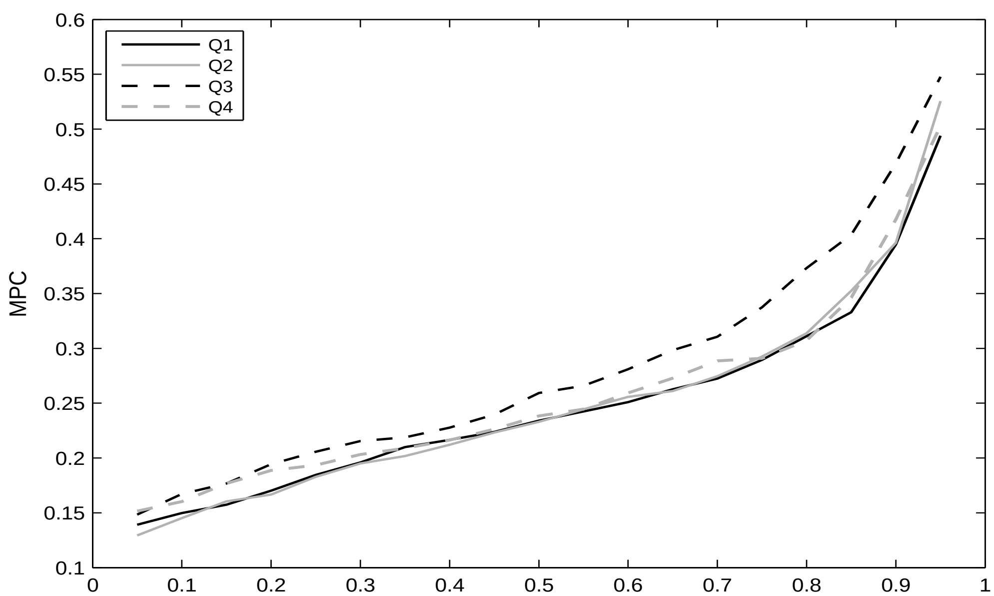
direct to some interesting results. MPC estimate is higher during Q3 for both 2005 and 2010 implying that median MPC is higher during Q3 of 2005 and 2010. Such a finding is surprising and contrary to the popular perception that individuals' are more willing to spend during Q4 (holiday season) relative to other three quarters. Number of members in CU is positively associated with consumption, and once again estimates during Q3 are higher relative to other quarters. The estimated coefficients for number of earners is positive for Q1, Q2 and Q4, but negative for Q3, and lastly, estimates for urban dummy are positive for all quarters of 2005 and 2010. However, the coefficients for the last two variables have high standard errors.

While the median MPC estimates display some interesting results, they give an incomplete picture of the distribution of consumption behavior. In order to get a complete picture, the consumption model is estimated at quantiles $\tau \in (0.05, 0.95)$ with increments of 0.05. Panels (a) and (b) of Figure 4 present a plot of the quantile estimates of MPC for the four quarters of 2005 and 2010. A look at Figure 4 reveals that MPC estimates monotonically increase along the quantiles followed by a sharp increase after the 80th quantile. This holds true for any given quarter of 2005 and 2010. It is therefore clear that MPC varies considerably along the quantiles. However, these variations will be overlooked if the analysis is solely restricted to the median or mean, an idea which has been strongly emphasized in the quantile regression literature.

The two panels of Figure 4 also show that the distribution of MPC as represented by its estimates at different quantiles varies across quarters for any given year. In 2005, the MPC estimates for Q3 are strictly greater than the corresponding estimates for the other three quarters, which are approximately same. Similarly, the MPC estimates for Q3 2010 are higher relative to other quarters at all quantiles, excluding the 5th quantile of Q4. The result i.e., higher MPC during Q3 relative to Q4 may surprise some of us, but we must note that Q3 largely overlaps with the summer vacation when families travel and spend more on food, transportation, vehicle rental lease, gasoline and admission fees to amusement parks. This increased expenditure during Q3 is possibly related to good weather and the associated feel good factor and hence labeled "summer effect". Moreover, most of the expenditures during Q3 are for self consumption which is in sharp contrast to expenditures during Q4, which are mainly for families and friends and sometimes done under pressure due to existing social norms. In this light, it is not difficult to comprehend that individuals' willingness to spend should be higher during Q3. To summarize quarter-wise comparison, higher MPC during Q3 relative to Q1, Q2 and Q4 is common to pre- and post recession of December



(a)



(b)

Figure 4: (a) Quantile plots of estimated MPCs Q1-Q4 in 2005, (b) Quantile plots of estimated MPCs Q1-Q4 in 2010.

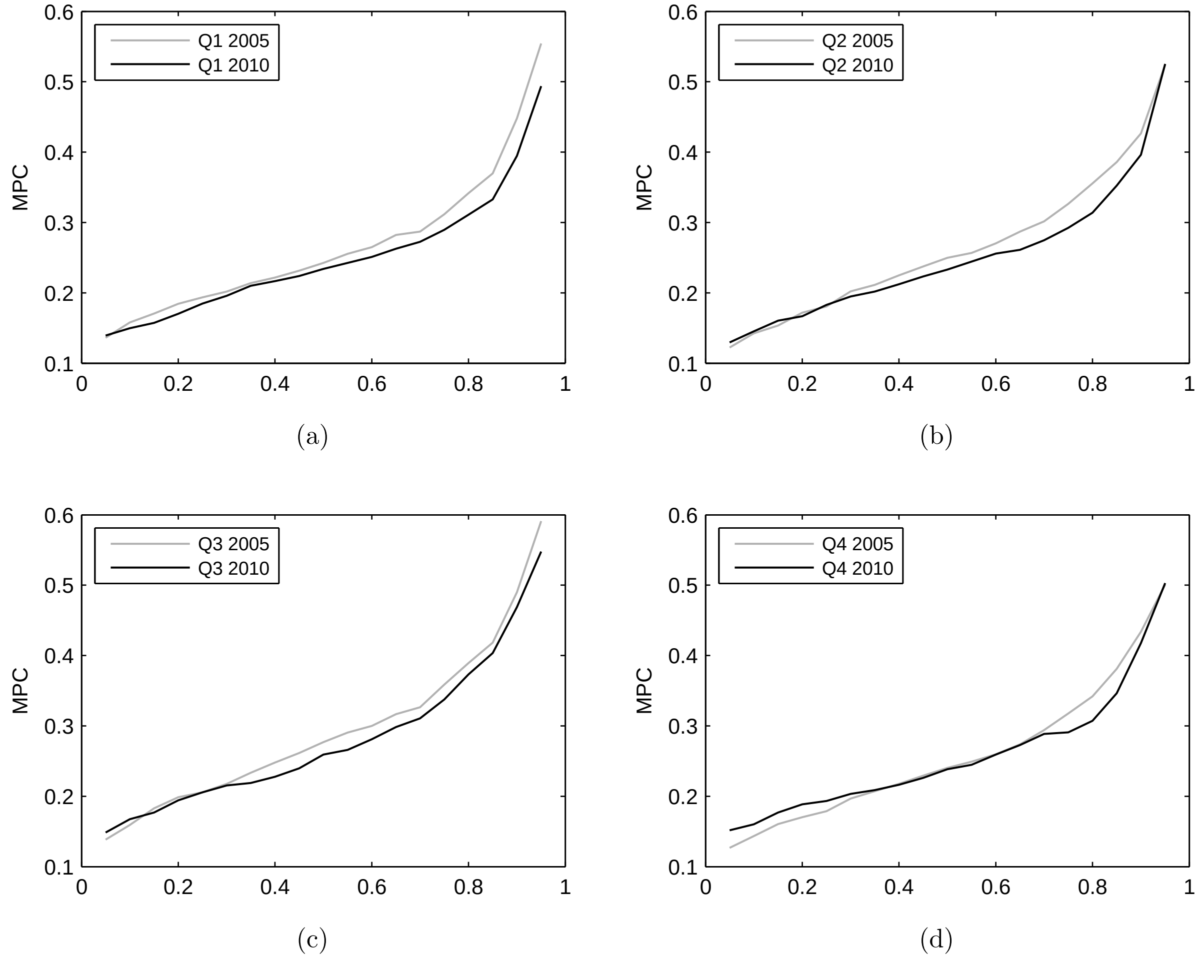


Figure 5: Quantile plots of estimated MPCs (a) Q1 (January - March), (b) Q2 (April - June), (c) Q3 (July - September), (d) Q4 (October - December).

2007 - June 2009.

Consumption and propensity to consume, apart from depending on individual characteristics, also depend on the state of the economy i.e., upswing and downswing of the business cycle. Several papers based on the Keynesian and Real Business Cycle (RBC) theories (Kydland and Prescott, 1982; John B. Long and Plosser, 1983; King and Rebelo, 2000) have found robust evidence in support of procyclical consumption. Given the recent recession of December 2007 - June 2009, it is of immense interest to examine consumption behavior during 2010, the year following the recession, in relation to 2005, a non-recession year. To examine the issue, Figure 5 plots MPC estimates of 2005 and 2010 at different quantiles by quarters. It can be seen that MPCs for 2010 are lower relative to 2005 for Q1, Q2 and Q3 at quantiles greater than the 30th consumption quantile. MPCs for CUs below the 30th consumption quantile (except the 5th quantile) are also lower during Q1 2010 relative to Q1 2005. However, for CUs below the 30th consumption quantile, there is no visible pattern in MPCs during Q2 and Q3. Therefore, my findings for Q1 - Q3 are more or less consistent with the widely held macroeconomic opinion that consumption behavior is procyclical.

The picture of consumption behavior as depicted by MPC estimates during Q4 is more complicated. MPC estimates for 2010 are greater than MPC estimates for 2005 until the 35th quantile, remain almost the same until the 70th quantile and thereafter are lower except for the 95th quantile. The increased marginal propensity to consume at lower consumption levels for 2010 can possibly be explained by Keynes' "animal spirit", in the form of higher willingness to consume following a prolong period of curtailed consumption. In contrast, the lower MPC estimates for 2010 at higher consumption quantiles reflect a more cautious consumption behavior as the recessionary effects was still existent during the 2010 holiday season.

In summary, quantile study of consumption behavior adds several layers of important information which can be useful for making policies targeted to boost consumption, especially, during and after periods of national economic stress. These layers of information will be overlooked if the analysis was solely focused on the mean or median.

6. Conclusion

The paper studies the applicability of nature inspired metaheuristic algorithms to quantile regression. Two of these algorithms, the firefly algorithm and accelerated particle swarm optimization algorithm, were examined in detail and compared to existing estimation techniques for linear and nonlinear quantile regression models. Relative to conventional techniques of estimating quantile regression, the metaheuristic algorithms are shown to have several advantages. First, they are versatile and can be easily adapted to meet the special challenges of quantile regression. Second, they can be easily adapted to linear and non-linear models, or to models that incorporate additional information in the form of restrictions or bounds on parameters. Third, they are more robust than standard algorithms and less susceptible to initial value problem. Fourth, they can efficiently estimate complicated models because the search for a global solution does not require restarting the algorithm at different starting points.

The paper presents two studies to illustrate the advantages of metaheuristic algorithms relative to conventional techniques such as IP (for linear models) and KP (for nonlinear models) algorithms. In the first study, the short term interest rate is modelled as a first order quantile autoregressive process. The model is linear in parameters and is estimated using the firefly and IP algorithms. The illustration shows that estimates from algorithms agree, bounds in linear models can be easily accommodated, and the quantile-crossing problem occurs at low values of the interest rate observed during the recent recession. In the second study, the parameters of the conditional scale ARCH model are jointly estimated using the firefly algorithm and the KP algorithm. Joint estimation is not possible using the IP algorithm since the model is nonlinear in parameters. Results based on simulated data suggest that the firefly algorithm can efficiently utilize model information and, unlike the KP algorithm, is not prone to initial value problem.

The paper also presents an application to consumption behavior that studies the marginal propensity to consume during pre- and post recession years of 2005 and 2010. The consumption model was estimated using the APSO algorithm that incorporates the restriction MPC $\in (0, 1)$, which is difficult to implement in existing techniques. The results exhibit interesting asymmetries. Consumption behavior as represented by MPC at different quantiles was higher during Q3 relative to other quarters for both 2005 and 2010. Such an occurrence was termed "summer effect" to denote increased Q3 spending by families on travel, shopping, amusement parks, etc., during summer vacation. When compared across years, MPC was higher during Q1-Q3 of 2005 for quantiles above the 30th quantile. This is consistent with the widely held macroeconomic view that consumption is procyclical. At quantiles below the 30th quantile, MPCs were higher during Q1 of 2005 except the 5th quantile, but no such pattern was visible for Q2 and Q3. However, the distribution of MPC during Q4 is more entangled. In particular, MPCs for 2010 were higher than that of 2005 below the 35th quantile, remain approximately same between the 35th and 70th quantile, and thereafter was lower except at the 95th quantile. The Q4 asymmetry implies that consumption behavior as represented by MPC actually increased in 2010 at lower consumption levels, remain mostly unaltered at moderate consumption levels and decreased at higher consumption levels.

The two comparative studies and the consumption application presented in this paper shows that metaheuristic algorithms represent a valuable alternative to standard LP and KP techniques. Furthermore, powerful new hybrid algorithms can be formulated by combining existing and new metaheuristic algorithms with classical optimization methods. Finally, a promising area of new development is the applicability in parallel computing methods, which would not only speed up the computations, but also allow for conditional estimation of multiple quantile regressions subject to cross-quantile restrictions. This remains an interesting avenue for future research.

Appendix A. Tables

Table A.9: Components of TOTEXPPQ or total expenditure in previous quarter

Variable	Explanation
FOODPQ	Expenditure on food.
ALCBEVPQ	Expenditure on alcoholic beverages.
HOUSPQ	Expenditure on housing.
APPARPQ	Expenditure on apparel and services.
TRANSPQ	Expenditure on transportation.
HEALTHPQ	Expenditure on health care.
ENTERTPQ	Expenditure on entertainment.
PERSCAPQ	Expenditure on personal care.
READPQ	Expenditure on reading.
EDUCAPQ	Expenditure on education.
TOBACCPQ	Expenditure on tobacco and smoking supplies.
MISCPQ	Miscellaneous expenditures.
CASHCOPQ	Cash contributions.
PERINSOQ	Personal insurance expenditure.

Table A.10: Components of FINCBTXM or Income before Tax

Variable	Explanation
FSALARYXM	Amount of wage and salary income
FNONFRMM	Income or loss from nonfarm business, partnership or professional practice.
FFRMINCM	Income or loss from own farm.
FRRETIRM	Social Security and Railroad Retirement income prior to deductions for medical insurance and Medicare.
FSSIXM	Amount of Supplemental Security Income from all sources.
UNEMPLXM	Income from unemployment compensation.
	Continued on next page

Table A.10 – continued from previous page

Variable	Explanation
COMPENSM	Income from worker's compensation or veteran's benefits, including education benefits, but excluding military retirement.
WELFAREM	Income from public assistance or welfare, including money received from job training grants.
INTEARNM	Income from interest on saving accounts or bonds.
FININCM	Regular income from dividends, royalties, estates or trusts.
PENSIONM	Total income from pensions or annuities from private companies, military, Government, IRA or Keogh.
INCLOSAM	Net income or loss received from roomers or boarders.
INCLOSBM	Net income or loss received as payments from other rental units.
ALIOTHXM	Total income from regular contributions to alimony and other sources such as from persons outside CU.
CHDOTHXM	Total income from child support payments in other than a lump sum amount.
OTHRINCM	Total amount of other money income, including money received from cash scholarships and fellowships, stipends not based on working, or from the care of other foster children.
FOODSMPM	Value of all food stamps and electronic benefits received.

References

Andrews, D. W. (2001), “Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space,” *Econometrica*, 68, 399–405.

Barrodale, I. and Roberts, F. (1973), “Improved Algorithm for Discrete L_1 Linear Approximation,” *SIAM Journal of Numerical Analysis*, 10, 839–848.

Bickel, P. (1978), “Using Residuals Robustly,” *Annals of Statistics*, 6, 266–291.

Bickel, P. and Lehmann, E. (1976), “Descriptive Statistics for Nonparametric Models. III: Dispersion,” *Annals of Statistics*, 4, 1139–1158.

Carroll, R. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman and Hall, London.

Chen, C. (2007), “A finite Smoothing Algorithm for Quantile Regression,” *Journal of Computational and Graphical Statistics*, 16, 136–164.

Chen, C. and Wei, Y. (2005), “Computational Issues for Quantile Regression,” *Sankhya: The Indian Journal of Statistics*, 67, 399–417.

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2010), “Quantile and Probability Curves without Crossing,” *Econometrica*, 78, 1093–1125.

Dantzig, G. B. (1963), *Linear Programming and Extensions*, Princeton University Press, Princeton.

Dantzig, G. B. and Thapa, M. N. (1997), *Linear Programming 1: Introduction*, Springer, New York.

Dantzig, G. B. and Thapa, M. N. (2003), *Linear Programming 2: Theory and Extensions*, Springer, New York.

Engle, R. F. (1982), “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation,” *Econometrica*, 50, 987–1007.

Geem, Z., Kim, J., and Loganathan, G. (2001), “A New Heuristic Optimization: Harmony Search,” *Simulation*, 76, 60–68.

Geem, Z. W. (2007), “Novel Derivative of Harmony Search Algorithm for Discrete Design Variables,” *Applied Mathematics and Computation*, 199, 223–230.

Geem, Z. W. (2009), *Music-Inspired Harmony Search Algorithm: Theory and Applications*, Springer, Berlin.

Geem, Z. W. (2010), *Recent Advances in Harmony Search Algorithm (Studies in Computational Intelligence)*, Springer, Berlin.

Glover, F. (1986), “Future Paths for Integer Programming and Links to Artificial Intelligence,” *Computers and Operations Research*, 13, 533–549.

Goffe, W. L., Ferrier, G., and Rogers, J. (1994), “Global Optimization of Statistical Functions with Simulated Annealing,” *Journal of Econometrics*, 60, 65–99.

Holland, H. (1975), *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.

Horowitz, J. L. (1998b), “Bootstrap Methods for Median Regression Models,” *Econometrica*, 66, 1327–1351.

Horowitz, J. L. (2001), “The Bootstrap,” *Handbook of Econometrics*, 5, 3159–3228.

John B. Long, J. and Plosser, C. (1983), “Real Business Cycles,” *Journal of Political Economy*, 91, 36–69.

Jong, K. (1975), “Analysis of the Behavior of a Class of Genetic Adaptive Systems,” Ph.D. thesis, University of Michigan, Ann Arbor.

Karmarkar, N. (1984), “A New Polynomial Time Algorithm for Linear Programming,” *Combinatorica*, 4, 373–395.

Kennedy, J. and Eberhart, R. (1995), “Particle Swarm Optimization,” *Proceedings of the IEEE International Conference of Neural Networks*, 4, 1942–1948.

King, R. G. and Rebelo, S. T. (2000), “Resuscitating Real Business Cycles,” *NBER Working Paper No. 7534*, 50.

Kirkpatrick, S., Gellat, C., and Vecchi, M. (1983), “Optimization by Simulated Annealing,” *Science*, 220, 671–680.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press, Cambridge.

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

Koenker, R. and d'Orey, V. (1987), "Computing Regression Quantiles," *Applied Statistics*, 36, 383–393.

Koenker, R. and Park, B. J. (1996), "An Interior Point Algorithm for Nonlinear Quantile Regression," *Journal of Econometrics*, 71, 265–283.

Koenker, R. and Zhao, Q. (1996), "Conditional Quantile Estimation and Inference for ARCH models," *Econometric Theory*, 12, 793–813.

Kydland, F. E. and Prescott, E. C. (1982), "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370.

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

Madsen, K. and Nielsen, H. B. (1993), "A Finite Smoothing Algorithm for Linear L_1 Estimation," *SIAM Journal of Optimization*, 3, 223–235.

Manski, C. F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205–228.

Mehrotra, S. (1992), "On the Implementation of Primal-Dual Interior Point Methods," *SIAM Journal of Optimization*, 2, 575–601.

Nelson, D. and Foster, D. (1994), "Asymptotic Filtering Theory for univariate ARCH models," *Econometrica*, 62, 1–42.

Newey, W. and Powell, J. (1987), "Asymmetric Least Squares Estimation and Testing," *Econometrica*, 55, 186–199.

Politis, D. and Romano, J. (1994), "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22, 2031–2050.

Powell, J. L. (1986), "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143–155.

Schwert, G. (1989), "Why does Stock Market Volatility Change over Time?" *Journal of Finance*, 44, 1115–1154.

Takeuchi, I. and Furuhashi, T. (2004), "Non-crossing Quantile Regression by SVM," *Proceedings of the International Joint Conference on Neural Networks*, 2004, 401–406.

Talbi, E. (2009), *Metaheuristics: From Design To Implementation*, John Wiley & Sons, New Jersey.

Taylor, S. (1987), *Modeling Financial Time Series*, John Wiley & Sons, New York.

Wu, C. (1990), "On the Asymptotic Properties of the Jackknife Histogram," *Annals of Statistics*, 18, 1438–1452.

Yang, X. (2008), *Nature-Inspired Metaheuristic Algorithms*, Luniver Press.

Yang, X. (2009), "Firefly Algorithms for Multimodal Optimization," *Stochastic Algorithms: Foundations and Applications, SAGA 2009, Lecture Notes in Computer Sciences*, 5792, 169–178.

Yang, X. (2010), *Engineering Optimization: An Introduction with Metaheuristic Applications*, John Wiley & Sons Inc, New Jersey.