Author's Accepted Manuscript

www.elsevier.com/locate/eer

Quantiles, corners, and the extensive margin of trade

José A.F. Machado, J.M.C. Santos Silva, Kehai Wei

PII: S0014-2921(16)30104-0
DOI: http://dx.doi.org/10.1016/j.euroecorev.2016.05.011
Reference: EER2880

To appear in: *European Economic Review*

Received date: 11 September 2015
Revised date: 14 March 2016
Accepted date: 31 May 2016

Cite this article as: José A.F. Machado, J.M.C. Santos Silva and Kehai Wei, Quantiles, corners, and the extensive margin of trade, *European Economi Review,* http://dx.doi.org/10.1016/j.euroecorev.2016.05.011

Quantiles, Corners, and the Extensive Margin of Trade*

José A. F. Machado
Faculty of Business and Management, Regent's University London

J. M. C. Santos Silva
School of Economics, University of Surrey

Kehai Wei
Department of Economics, University of Essex

March 14, 2016

Abstract: We develop a simple method for the estimation of quantile regressions for corner solutions data (i.e., fully observed non-negative data that have a mixed distribution with a mass-point at zero), focusing particular attention on the case where the domain of the variate of interest is bounded both from below and from above. We use the proposed method to study the determinants of the extensive margin of trade and find that most regressors have very different impacts on different parts of the distribution.

JEL classification code: C21, C29, C61, F14.

Key words: BFGS algorithm; Corner solutions; Local bandwidth; Non-linear quantile regression.

*We thank two anonymous referees and the Editor Theo Eicher for constructive comments and advice that helped to substantially improve the paper. We also thank Rodolphe Desbordes, José de Sousa, Daniel Dias, Silvana Tenreyro, Greg Wright, and especially Paulo Parente, for helpful comments and suggestions. The usual disclaimer applies. Machado is consultant for the Research Department of Banco de Portugal. Machado and Santos Silva also gratefully acknowledge the partial financial support from Fundação para a Ciência e Tecnologia.

Address for correspondence: João Santos Silva, School of Economics, University of Surrey, Guildford GU2 7XH, United Kingdom. Tel: +44 (0)1483686956. E-mail: jmcss@surrey.ac.uk.

1. INTRODUCTION

Empirical researchers are often faced with the need to model non-negative data that have a mixed distribution with a mass-point at zero. For example, data of this type can arise as a consequence of censoring, and methods to produce valid inference in this context are now well established (see, e.g., Wooldridge, 2002, and Cameron and Trivedi, 2005, for reviews). More frequently, however, data of this type are the result of the existence of the so-called corner solutions (see Wooldridge, 2002). In this case the variate of interest is fully observed but has a mixed distribution with a mass-point at zero, the lower bound of its support. Data with these characteristics are ubiquitous in economics, being often found in health economics (Duan, Manning, Morris and Newhouse, 1983), international economics (Santos Silva and Tenreyro, 2006), finance (La Porta, López-de-Silanes, and Zamarripa, 2003), and in many other areas.

Modelling corner solutions data poses particular problems and requires appropriate inference tools. Wooldridge (2002) and Cameron and Trivedi (2005) provide several examples of methods that have been used to model this type of data. These methods typically focus on the estimation of the conditional expectation of the variate of interest, although in some cases likelihood-based methods are also used to model its entire conditional distribution. In many practical situations, however, knowledge of the conditional expectation may not be enough to fully understand how the covariates affect the conditional distribution of interest, but the researcher may not be confident enough to apply more informative but also more demanding likelihood-based methods. Therefore, it is interesting to develop methods that can provide information on different features of the conditional distribution of this type of data without requiring strong distributional assumptions.

In this paper we develop a method to estimate conditional quantiles for corner solutions data, focusing particular attention on the case where the domain of the variate of interest is bounded both from below and from above.[1] Although we borrow form the lit-

[1] Papke and Wooldridge (1996) is the leading reference on the estimation of the conditional expecta-

erature on censored and non-linear quantile regression, the proposed approach is novel in that it is the first time that quantile regression for this type of data is considered. Specifically, we propose a simple non-linear specification of the conditional quantiles that is compatible with the characteristics of doubly-bounded corner solutions data, show that the parameters of interest can be estimated using a procedure that is easy to implement using standard software, consider the asymptotic properties of the estimator, and discuss its practical implementation. More generally, with small adaptations the proposed method can also be used to estimate both non-linear and censored quantile regressions, and it is more flexible and easier to implement than many of the alternative approaches currently available. The results of a set of simulation exercises suggest that the proposed methods perform well in practice and, therefore, are likely to be useful in a wide variety of empirical applications.[2]

We use the proposed methods to investigate the determinants of the extensive margin of trade, a problem that has attracted a great deal of attention since the seminal work of Hummels and Klenow (2005).[3] Specifically, we estimate quantile regression models for the extensive margin of trade defined as the number of sectors exporting from country j to country i in year t. This kind of data typically has a mass-point at zero because in most years there is a sizable proportion of countries that do not trade with many potential partners. Moreover, besides the natural lower bound at zero, this variable is

tion for doubly-bounded data. Estimation of quantile regression for doubly-bounded data with possible mass-points at the bounds was first considered by Machado and Santos Silva (2008). Liu and Bottai (2009) and Bottai, Cai, and McKeown (2010) considered the simpler case without mass-points. See Ramalho, Ramalho, and Murteira (2011) for a survey of other methods used to model doubly-bounded data with and without mass-points.

[2]A Stata (StataCorp., 2013) command implementing the methods proposed here is available from the authors on request.

[3]For example, Helpman, Melitz, and Rubinstein (2008), Chaney (2008), and Manova (2013), developed trade models that explicitly take into account the extensive margin. Examples of recent empirical studies in this area include, among many others, Dutt, Mihov, and Van Zandt (2013), Kehoe and Ruhl (2013), Santos Silva, Tenreyro, and Wei (2014), Feenstra and Ma (2014), and Eicher and Kuenzel (2016).

also bounded from above by the number of categories in the particular classification of economic activities that is used to define the sectors. The specific characteristics of these data imply that the conditional distribution of the variate of interest (the number of sectors) will often depend on the regressors in a complex way and, therefore, the informational gains provided by quantile regressions (e.g., Koenker, 2005) are likely to be particularly interesting in this context.

Our results suggest that there is a substantial degree of heterogeneity in the effect of the covariates on the conditional distribution of the extensive margin, with most regressors having a much larger impact on the upper tail of the distribution. This implies that changes in the regressors result in changes in the conditional distribution that generally are more pronounced in regions corresponding to pairs of countries that trade more than pairs with similar observable characteristics.

The remainder of the paper is organized as follows. Section 2 details our approach to the estimation of quantile regression for corner solutions data, and Section 3 presents the results of simulation experiments illustrating the performance of the proposed methods. In Section 4 we use the proposed method to study the determinants of the extensive margin of trade. Finally, section 5 contains some brief concluding remarks.

2. QUANTILE REGRESSION FOR CORNER SOLUTIONS DATA

2.1. Specification

Estimation of quantile regression for corner solutions data is complicated by the fact that the quantiles are not necessarily smooth functions of the regressors. For expository purposes, we will consider mainly the case where y, the variate of interest, has support on $[0, 1]$ and a mass-point at zero; other cases can be handled in a similar fashion.

For data with a mass-point at zero, there are conditional quantiles that become identically zero for some values of the covariates. Specifically, for $\tau \in (0, 1)$, the τ-th conditional quantile of y given x has the form

$$Q_y(\tau|x) = 1(\tau > \Pr(y=0|x))\, Q_y\left(\left. \frac{\tau - \Pr(y=0|x)}{1-\Pr(y=0|x)} \right| x, y>0 \right), \tag{1}$$

where $1(e)$ is the indicator function of event e and x is a vector of covariates.[4] Therefore, in general, $Q_y(\tau|x)$ is not strictly increasing in τ and x, and is not a smooth function, possibly having a "corner" at $\Pr(y=0|x)=\tau$. This suggests that in the case where y has support on $[0,1]$ and a mass-point at zero the conditional quantiles of y will have the form

$$Q_y(\tau|x) = \max\{0, h(x,\theta)\}, \tag{2}$$

where $h(x,\theta)$ is a function such that $h(x,\theta) < 1$.

The choice of $h(x,\theta)$ is naturally an empirical matter but, in the spirit of Papke and Wooldridge (1996), we suggest the following specification:

$$h(x,\theta) = (1+\gamma)\,\Lambda(x'\beta) - \gamma, \tag{3}$$

where $\Lambda(x'\beta)$ denotes a cumulative distribution function, γ is an unknown shape parameter, and $\theta = (\beta,\gamma)$.[5] The shape parameter γ adds some flexibility to the specification of $Q_y(\tau|x)$; for example, if $\Lambda(x'\beta)$ is the cumulative distribution function of a random variable symmetrically distributed around 0, the positive part of the quantiles will be s-shaped for $-1<\gamma<1$, and concave for $\gamma>1$.[6] In what follows, we will consider only the case where $\Lambda(x'\beta) = \exp(x'\beta)\,[1+\exp(x'\beta)]^{-1}$, but naturally this function can take much more flexible forms (see, e.g., Aranda-Ordaz, 1981).

The specification in (2) is reminiscent of the model considered by Powell (1984, 1986) for the case of zero-censored linear quantile regression. This similarity results from the fact that zero-censored data also have a mixed-distribution with a mass point at zero. However, it is important to note that in the case considered by Powell (1984, 1986) the mass-point is the result of censoring and the latent variate of interest is assumed to have

[4]See Buchinsky and Hahn (1998) for a related result in the context of censored quantile regression.

[5]As usual, θ depends on τ but we do not make that explicit to simplify the notation.

[6]Although we do not impose that restriction, (3) is only interesting for $\gamma > -1$.

an absolutely continuous distribution; this implies that $h(x,\theta)$ is always interpreted as an estimate of the relevant quantile, even when $h(x,\theta) < 0$. In contrast, in the case we are considering the distribution of y is not censored and the mass-point is an important feature of the distribution of the variate of interest, which is fully observable. In this case, the form of $h(x,\theta)$ is meaningless in regions where $h(x,\theta) < 0$.

Naturally, other specifications of $h(x,\theta)$ can be considered, and other types of data will require different specifications of this function. For example, Machado and Santos Silva (2008) consider the case where y has support on $[0,1]$ with a mass-point at 1, or mass-points both at 0 and 1. For data with support on $[0,\infty)$ we can specify $h(x,\theta) = \exp(x'\beta) - \gamma$, or simply $h(x,\theta) = x'\beta$, as in Powell (1984, 1986).

2.2. Estimation

The basic intuition for the estimator we will use is as follows; see Koenker and Bassett (1978) and Koenker (2005) for details. Under suitable regularity conditions,[7] the assumption that for a given $\tau \in (0,1)$ there exist a $p \times 1$ vector $\theta_0 = (\beta_0, \gamma_0)$ such that

$$Q_y(\tau|x) = \max\{0, h(x,\theta_0)\},$$

implies that θ_0 is the sole solution of

$$\min_{\theta\in\Theta} E\left[\rho_\tau\left(y - \max\{0, h(x,\theta)\}\right)\right], \tag{4}$$

where the parameter space $\Theta \subset \mathbb{R}^p$ is compact and $\rho_\tau(z) = z\left[\tau - 1(z<0)\right]$. Then, for an i.i.d. sample $\{y_i, x_i'\}_{i=1}^n$, the estimator will be defined by the analogue of (4), which is given by $\hat{\theta} = \arg\min_\theta S(\theta)$, with

$$S(\theta) = \frac{1}{n}\sum_{i=1}^n \rho_\tau\left(y_i - \max\{0, h(x_i,\theta)\}\right). \tag{5}$$

[7] In particular, identification requires that the matrices D_0 and D_1 in condition C3 below are non-singular and finite. In practice, identification depends on the curvature of $\Lambda(x'\beta)$: if the data are such that $h(x,\theta)$ is essentially linear, identification will be difficult. Moreover, $\max\{0, h(x,\theta)\}$ needs to be positive for a sufficiently large number of observations for θ to be identified.

Therefore, estimation of θ is a non-convex and non-linear quantile regression problem which has as a special case the usual censored quantile regression estimator (Powell, 1984, 1986) when $h(x, \theta)$ is linear and the mass-point at zero is the result of censoring.

Due to the non-convex and non-linear nature of the problem, minimization of (5) cannot be implemented using the linear programming methods typically used for quantile regression estimation, and a number of alternative approaches have been proposed for related problems.[8] Machado and Santos Silva (2008) proposed a method to minimize (5) based on a grid search over γ and on the repeated application of the elegant three-step estimator of Chernozhukov and Hong (2002). However, this algorithm does not work in the application we consider in Section 4 because in the initial sub-sample it is not possible to identify the coefficients of all the regressors, and therefore it is not possible to select the observations to use in the next step. More generally, popular algorithms based on estimation of quantile regressions in a sequence of subsamples (e.g., Buchinsky, 1994, and Chernozhukov and Hong, 2002) are likely to fail in applications where the dependent variable has many zeros and the model contains a large number of dummies.

To avoid this problem, here we use a much simpler approach motivated by Koenker's (2008) observation that the popular BRCENS algorithm for the estimation of censored quantile regression (Fitzenberger, 1997a) can be seen as the direct minimization of (5) using the steepest descent algorithm.[9] We, therefore, follow a similar approach and obtain $\hat{\theta}$ by direct minimization of (5) using the well-known BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm.

There are a number of reasons to use the BFGS method in place of the steepest descent algorithm used by Fitzenberger (1997a). First, the BFGS algorithm is known to generally perform better than the simpler steepest descent method (see, e.g., Judd, 1988). More importantly, the recent results of Lewis and Overton (2013) show that the BFGS algorithm works well even in non-smooth and non-convex problems such as the

[8] See, e.g., Buchinsky (1994), Koenker and Park (1996), Buchinsky and Hahn (1998), Chernozhukov and Hong (2002), and the survey by Koenker (2008).

[9] Naturally, Fitzenberger's (1997a) algorithm is for the case where $h(x, \theta) = x'\beta$.

one we consider here. Last but not the least, the BFGS approach is easily available to practitioners because it is implemented in popular software packages such as TSP (Hall and Cummins, 2009) and Stata (StataCorp., 2013).

Naturally, there is no guarantee that minimization of (5) using the BFGS algorithm will lead to the global minimum of the objective function and therefore it is important to choose a suitable set of starting values for θ. In the next section we describe a simple method to obtain starting values and report simulation results which suggest that the proposed approach works rather well in practice.[10]

2.3. Inference

The consistency of the estimator obtained by minimizing (5) follows directly from the usual results on the estimation of non-linear quantile regression (see Oberhofer, 1982, Koenker, 2005, and the references therein). The conditions needed to establish the asymptotic normality of the estimator can also easily be obtained from those given by Koenker (2005). Indeed, for the particular problem we are considering, we can essentially maintain conditions G1 and G2 in Koenker (2005, p. 124) and we only need to make a small modification to condition A1 (Koenker, 2005, p. 120). Below we spell out the necessary regularity conditions using the notation of this paper.

Let y_1, y_2,... be independent random variables with conditional distribution functions F_1, F_2,... with support on $[0, 1]$, and define $Q_{y_i}(\tau|x_i) = g_i(\theta)$, with $0 \leq g_i(\theta) < 1$. Additionally we assume that the following regularity conditions hold almost surely:

C1: The conditional distribution functions $\{F_i\}$ are absolutely continuous in the interval $(0, 1)$, with continuous conditional densities f_i uniformly bounded away form 0 and ∞ at any point with $g_i(\theta) > 0$.

C2: There exist constants k_1, k_2, and n_0, such that for θ_1, $\theta_2 \subset \Theta$ and $n > n_0$,

$$k_1 \|\theta_1 - \theta_2\| \leq \left(n^{-1} \sum\nolimits_{i=1}^{n} \left(g_i(\theta_1) - g_i(\theta_2)\right)^2\right)^{1/2} \leq k_2 \|\theta_1 - \theta_2\| .$$

[10] Bottai, Orsini, and Geraci (2015) also find that a gradient-based algorithm performs well in a closely related problem.

C3: There exist positive definite matrices D_0 and D_1 such that,

(a) $\lim_{n\to\infty} n^{-1} \sum_{\dot{g}_i \neq 0} (\tau - 1(y_i < g_i(\theta)))^2 \dot{g}_i \dot{g}_i' = D_0$,

(b) $\lim_{n\to\infty} n^{-1} \sum_{\dot{g}_i \neq 0} f_i(g_i(\theta)) \dot{g}_i \dot{g}_i' = D_1$,

(c) $\max_{i=1,\ldots,n} \|\dot{g}_i\| / \sqrt{n} \to 0$,

where $\dot{g}_i = \partial g_i(\theta) / \partial\theta|_{\theta=\theta_0}$.

Condition C1 is similar to condition A1 in Koenker (2005), the difference being that we assume F_i to be absolutely continuous only in the interval $(0, 1)$ and we require f_i to exist only for the points where $g_i(\theta) > 0$. This is done to allow for the fact that F_i has a mass-point at zero and has no practical consequences because (1) implies that $\dot{g}_i = 0$ when $g_i(\theta) = 0$, and therefore the form of D_1 in C3(c) is unaffected by the existence of the mass-point.[11] Conditions C2 and C3 parallel conditions G1 and G2 in Koenker (2005), the only difference being that in C3(a) we follow Chamberlain (1994), Kim and White (2003), and Angrist, Chernozhukov and Fernandez-Val (2004) and do not assume that $E\left[(\tau - 1(y_i < g_i(\theta)))^2 | x_i\right] = \tau(1-\tau)$.[12]

Under these conditions it is possible to show that (see, Powell, 1984, 1986, Koenker 2005, and the references therein)

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0, \Omega)$$

with $\Omega = D_1^{-1} D_0 D_1^{-1}$.

[11]It is interesting to notice that in the case of censored quantile regression the problems created by the existence of the mass-point at zero are side-stepped by imposing conditions on the distribution of the unobserved latent dependent variable (see Powell, 1984, 1986), something that is not possible in the case we are considering. Powell's (1984, 1986) approach is valid because, again, the density of the dependent variable is only needed for the observations where the quantile of the observable data is not identically zero.

[12]Following Parente and Santos Silva (2016), it is easy to modify these conditions to allow for clustering. That is done in the empirical application in Section 4.

2.4. Implementation issues

In practice, it is necessary to find an estimator for Ω. Following the pioneering work of Buchinsky (1995), one possibility is to estimate Ω using an appropriate bootstrap procedure. However, if the problem at hand is large, using the bootstrap may be impractical and therefore it is interesting to have an alternative way of estimating Ω.

Let $\widehat{g}_i = g_i\left(\hat{\theta}\right)$ and $\widehat{\dot{g}}_i = \partial g_i(\theta)/\partial\theta|_{\theta=\hat{\theta}}$. Then, following Powell (1984), Chamberlain (1994), Kim and White (2003), and Angrist, Chernozhukov and Fernandez-Val (2004), D_0 can be consistently estimated by

$$\widehat{D}_0 = n^{-1}\sum\nolimits_{\widehat{g}_i\neq 0}\left(\tau - 1\left(y_i < \widehat{g}_i\right)\right)^2 \widehat{\dot{g}}_i\widehat{\dot{g}}_i',$$

whereas, for an appropriately defined kernel $K(\cdot)$ and smoothing parameter δ_n,

$$\widehat{D}_1 = \frac{1}{n\delta_n}\sum\nolimits_{\widehat{g}_i\neq 0} K\left(\frac{y_i - \widehat{g}_i}{\delta_n}\right)\widehat{\dot{g}}_i\widehat{\dot{g}}_i'$$

is a consistent estimator of D_1. Therefore, $\widehat{D}_1^{-1}\widehat{D}_0\widehat{D}_1^{-1} \xrightarrow{p} \Omega$. In the reminder of this section we consider the choice of $K(\cdot)$ and δ_n.

To accommodate censored data, Powell (1984) uses a one-sided rectangular kernel in the estimation of $\widehat{D}_1$. In the case we consider here this approach is not valid because it does not account for the upper-bound in the support of y. As an alternative, we use the standard symmetric rectangular kernel with a variable bandwidth of the type considered by Dai and Sperlich (2010). This ensures that the kernel puts non-zero weight only on points where F_i is absolutely continuous and $f_i > 0$. In particular, in the spirit of Dai and Sperlich (2010), we set $\delta_n = \min\left\{\widehat{g}_i, 1-\widehat{g}_i, \delta_n^{Global}\right\}$, where δ_n^{Global} is defined using the method described in Koenker (2005, p. 81).[13] Specifically, using $\Phi(\cdot)$ and $\phi(\cdot)$ to denote the normal distribution and density functions, we set

$$\delta_n^{Global} = \kappa\left[\Phi^{-1}\left(\tau + d_n\right) - \Phi^{-1}\left(\tau - d_n\right)\right],$$

[13] Notice that this approach is valid when y has domain $[0,1]$, whether or not there are mass-points at any or at both boundaries.

where d_n is (see Koenker, 2005, p. 140)

$$d_n = n^{-1/3} \left(\Phi^{-1} \left(1 - \frac{0.05}{2} \right) \right)^{2/3} \left(\frac{1.5 \left(\phi \left(\Phi^{-1} \left(\tau \right) \right) \right)^2}{2 \left(\Phi^{-1} \left(\tau \right) \right)^2 + 1} \right)^{1/3},$$

and κ is a robust estimate of scale. Both in the simulations described in the next section and in the models for the extensive margin of trade presented in Section 4, we set $\kappa = \text{MAD}$, where MAD denotes the median absolute deviation of the τ-th quantile residuals $y_i - \widehat{g}_i$.

3. SIMULATION EVIDENCE

In this section we report the results of a simulation study providing evidence on the performance of the methods described in Section 2. Data for these simulations were generated as

$$y_i = \max \left\{ 0, g_i \left(\theta \right) + u_i \right\}, \quad i = 1, \ldots, n,$$

with

$$g_i \left(\theta \right) = \max \left\{ 0, \left(1 + \gamma \right) \frac{\exp \left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \right)}{1 + \exp \left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \right)} - \gamma \right\},$$

and

$$u_i = 0.1(1 - g_i \left(\theta \right))(\varepsilon_i - \Phi^{-1} \left(\tau \right)),$$

where x_{1i} and ε_i are obtained as independent draws from a standard normal distribution, and x_{2i} is obtained as independent draws from a Bernoulli distribution with $\Pr \left(x_{2i} = 1 \right) = 0.2$. That is, the disturbance u_i has an heteroskedastic normal distribution with $Q_{u_i} \left(\tau \right) = 0$, and therefore $Q_{y_i} \left(\tau | x_{1i}, x_{2i} \right) = g_i \left(\theta \right)$. The values of x_{1i}, x_{2i}, and ε_i are drawn independently for each of the 10000 replicas of each experiment, and in all cases we set $\beta_0 = \beta_1 = \beta_2 = -1$; we performed experiments with $n \in \{250, 1000, 4000, 16000\}$, $\tau \in \{0.25, 0.50, 0.75\}$, and $\gamma \in \{0.1, 0.3\}$.[14]

[14]The parameter γ controls the proportion of observations for which $g \left(\tau, x_i, \theta \right) = 0$; with this design, this proportion is approximately 15% for $\gamma = 0.1$ and 50% for $\gamma = 0.3$.

In each replica, θ was estimated by minimizing (5) using the BFGS algorithm as implemented in the `ml` command in Stata (StataCorp., 2013); the default options were used for the maximum number of iterations and for the convergence criteria.[15] In all cases the starting value of γ is set to zero, so that the initial values of $g_i(\theta)$ are always positive; the starting values of the other parameters were obtained as the estimates obtained in the τ-th quantile regression of $\ln(y_i/(1-y_i))$ on the regressors, for the observations with $y_i > 0$.

A first issue that has to be considered in this kind of problem is the ability of the algorithm to converge (see, e.g., Fitzenberger, 1997b). Table 1 reports the percentage of replicas where convergence was not achieved for each of the 24 cases considered. Remarkably, convergence was achieved in at least 99.91% of the replicas, with the algorithm being particularly successful for the larger samples. Naturally, if in a particular application it is not possible to obtain convergence with this approach, there are a number of aspects of the algorithm that can be modified. For example, a grid search over γ can be used to obtain better starting values, and we also had good results using the Davidon–Fletcher–Powell algorithm in the first few iterations before switching to the BFGS.

Table 1: Percentage of cases where convergence was not achieved

	$\gamma = 0.1$			$\gamma = 0.3$		
n	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
250	0.05	0.02	0.05	0.08	0.03	0.09
1000	0.02	0.00	0.02	0.02	0.04	0.06
4000	0.01	0.00	0.00	0.00	0.00	0.00
16000	0.00	0.00	0.00	0.00	0.00	0.00

[15] The simulations were performed in Stata 13 (StataCorp. 2013) but version control was used to run the code as in Stata 11. This is needed because from Stata 12 the convergence criteria for the BFGS algorithm uses the Hessian of the objective function.

Table 2 displays the biases and standard errors (in parentheses) of the estimates obtained in the replicas where convergence was achieved.[16] Overall, the biases are relatively small, even for $n = 250$ and for the cases where $\gamma = 0.3$ (in which the quantiles are equal to 0 for about 50% of the sample). Also, as could be expected, the standard errors go

Table 2: Biases and standard errors

		$\gamma = 0.1$				$\gamma = 0.3$			
n	τ	β_0	β_1	β_2	γ	β_0	β_1	β_2	γ
250	0.25	−0.010	−0.007	−0.011	0.002	−0.065	−0.031	−0.032	0.005
		(0.188)	(0.080)	(0.140)	(0.057)	(0.398)	(0.130)	(0.190)	(0.195)
	0.50	−0.008	−0.006	−0.009	0.001	−0.051	−0.025	−0.028	0.005
		(0.169)	(0.072)	(0.126)	(0.049)	(0.365)	(0.119)	(0.176)	(0.171)
	0.75	−0.006	−0.006	−0.010	0.003	−0.098	−0.046	−0.051	−0.004
		(0.178)	(0.077)	(0.135)	(0.052)	(0.425)	(0.150)	(0.214)	(0.195)
1000	0.25	−0.001	−0.001	−0.002	0.001	−0.005	−0.004	−0.005	0.004
		(0.092)	(0.039)	(0.067)	(0.026)	(0.192)	(0.061)	(0.089)	(0.081)
	0.50	−0.001	−0.001	−0.002	0.001	−0.004	−0.003	−0.004	0.004
		(0.084)	(0.036)	(0.061)	(0.024)	(0.174)	(0.057)	(0.081)	(0.072)
	0.75	−0.002	−0.002	−0.003	0.001	−0.013	−0.007	−0.008	0.002
		(0.089)	(0.038)	(0.066)	(0.025)	(0.197)	(0.064)	(0.091)	(0.081)
4000	0.25	0.001	0.000	0.000	0.000	0.000	−0.001	0.000	0.001
		(0.044)	(0.019)	(0.033)	(0.012)	(0.092)	(0.030)	(0.043)	(0.037)
	0.50	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	0.001
		(0.041)	(0.017)	(0.030)	(0.011)	(0.083)	(0.027)	(0.039)	(0.034)
	0.75	−0.001	0.000	−0.001	0.000	−0.001	−0.001	−0.001	0.001
		(0.045)	(0.019)	(0.033)	(0.013)	(0.091)	(0.030)	(0.043)	(0.036)
16000	0.25	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		(0.022)	(0.009)	(0.016)	(0.006)	(0.045)	(0.015)	(0.021)	(0.018)
	0.50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
		(0.020)	(0.009)	(0.015)	(0.006)	(0.041)	(0.013)	(0.019)	(0.016)
	0.75	0.000	0.000	0.000	0.000	−0.001	0.000	−0.001	0.000
		(0.022)	(0.009)	(0.016)	(0.006)	(0.045)	(0.015)	(0.021)	(0.018)

[16] In some of the experiments with $\gamma = 0.3$ and $n \in \{250, 1000\}$, β_2 is not identified because x_1 is equal to 1 only when the estimated quantile is 0. In Tables 2 and 3, the results for β_2 are only for the cases where this parameter was identified.

down approximately by a factor of 2 when the sample size increases by a factor of 4. These results are particularly encouraging because they suggest that, although there is no guarantee that the proposed estimation algorithm converges to the global minimum of (5), that does not seem to have an adverse impact on the estimation results.

Finally, Table 3 reports the rejection frequencies at the 5% level of the null hypotheses that the parameters are equal to their true values. These tests were based on standard t-statistics, computed from estimates of the covariance matrix obtained using the method described in Subsection 2.4. The results in Table 3 show that the tests for the βs generally have good size,[17] even for the smaller sample and with $\gamma = 0.3$. In contrast, the t-tests for γ are oversized for $n \in \{250, 1000\}$, especially when $\gamma = 0.3$. For the larger samples, however, all the tests perform quite well. These results suggest that for relatively small samples it may be advisable to perform inference based on p-values computed by bootstrapping these test statistics; for moderately large samples, where the

Table 3: Rejection frequencies at the 5% level

		$\gamma = 0.1$				$\gamma = 0.3$			
n	τ	β_0	β_1	β_2	γ	β_0	β_1	β_2	γ
250	0.25	0.068	0.042	0.055	0.106	0.079	0.042	0.047	0.143
	0.50	0.057	0.033	0.047	0.085	0.056	0.027	0.041	0.116
	0.75	0.057	0.036	0.060	0.082	0.091	0.055	0.065	0.156
1000	0.25	0.065	0.050	0.057	0.080	0.057	0.044	0.058	0.079
	0.50	0.063	0.045	0.050	0.076	0.052	0.036	0.047	0.071
	0.75	0.059	0.046	0.057	0.071	0.067	0.051	0.062	0.089
4000	0.25	0.053	0.049	0.056	0.057	0.054	0.049	0.051	0.064
	0.50	0.055	0.046	0.054	0.063	0.047	0.043	0.048	0.053
	0.75	0.065	0.058	0.057	0.066	0.054	0.049	0.057	0.061
16000	0.25	0.051	0.051	0.048	0.055	0.054	0.052	0.048	0.054
	0.50	0.054	0.053	0.053	0.056	0.050	0.047	0.051	0.052
	0.75	0.057	0.055	0.052	0.058	0.054	0.051	0.056	0.057

[17]Following Cochran (1952), the test is considered to have good size when its actual size is between 4% and 6%. Given the number of replications performed, the test is considered to have good size if the rejection frequencies are between 0.036 and 0.065.

bootstrap may be too costly, the usual p-values computed from the asymptotic distribution are likely to be reasonably reliable.

Overall, the results of these simulations are quite encouraging and suggest that the proposed approach is likely to work well in a variety of cases, even when the percentage of corner solution observations is substantial.

4. QUANTILES FOR THE EXTENSIVE MARGIN OF TRADE

The last decade has seen a rapid increase in the attention devoted to the study of the extensive margin of trade.[18] There are several reasons for this.

First, it has been recognized that the extensive margin has had a substantial contribution to the expansion of international trade (see, e.g., Hummels and Klenow, 2005, and Kehoe and Ruhl, 2013). Second, it has been noted that the explanatory variables traditionally considered in trade models may have very different impacts on the extensive and intensive margins of trade (see, e.g., Lawless, 2010, Hillberry and Hummels, 2008, and Dutt, Mihov, and Van Zandt, 2013). Third, the increase in the variety of imported goods has been associated to increased welfare in the importing country (see, e.g., Romer, 1994, Broda and Weinstein, 2006, and Ardelean and Lugovskyy, 2010). Finally, and perhaps more importantly, diversification of exports has been linked to increases in productivity and more rapid growth (see, e.g., Feenstra et al., 1999, Feenstra and Kee, 2008, and Eicher and Kuenzel, 2016). In fact, in a model with heterogeneous firms à la Melitz (2003), exporting opportunities in new sectors will drive up factor prices forcing less productive firms to exit the market; this "natural selection" process leads to higher average productivity and more rapid growth. This link between diversification of exports and increased productivity and growth has been shown to be empirically relevant, at least for certain groups of countries (see, e.g., Feenstra and Kee, 2008, and Eicher and Kuenzel, 2016). Therefore, trade facilitating policies that contribute to the

[18]The origins of this literature can be traced back to the work of Krugman (1979) and Melitz (2003).

diversification of exports are seen as important tools in promoting growth and development (see, e.g., Dennis and Shepherd, 2011, and Feenstra and Ma, 2014).

The definition of the extensive margin of trade depends on the nature of the data that are available. Here we will consider the so-called sector margin, defined as the number of sectors exporting from origin country j to a destination country i in year t, with the sectors being defined by the 1996 revision of the Harmonized Commodity Description and Coding System at the 6-digit level, which has 5132 categories. This choice is motivated both by the availability of this kind of data and by the fact that at this level of aggregation the sector margin is reasonably informative about the diversity of the export flows.[19] Similar definitions of the extensive margin have been used, among others, by Hillberry and McDaniel (2002), Dennis and Shepherd (2007, 2011), Dutt, Mihov, and Van Zandt (2013), Santos Silva, Tenreyro, and Wei (2014), and Eicher and Kuenzel (2016).[20]

Santos Silva, Tenreyro, and Wei (2014) considered different models for the sector margin, but focused on the estimation of the conditional expectation of the variate of interest. However, because the data have a lower and an upper bound, and also a large percentage of zeros, the conditional expectation necessarily provides an incomplete picture of how the regressors affect the conditional distribution of interest. It is, therefore,

[19]There are two important points to make about the use of this definition of the extensive margin. First, strictly speaking, the dependent variable is discrete but given the large number of support points of its distribution this is immaterial. Second, naturally the results will depend on the level of aggregation of the data that are used. Specifically, the use of more aggregated data may hide important relations because quantiles of a coarser dependent variable will be less sensitive to changes in the regressors (this point is discussed in detail in Machado and Santos Silva, 2005). Therefore, the quantiles for the sector margin should be estimated using data that are as disaggregated as possible.

[20]Other definitions of the extensive margin have been used in empirical studies. For example, Hillberry and Hummels (2008) work at the shipment level, Eaton, Kortum, and Kramarz (2004), and Buono and Lalanne (2012) work at the firm level, and Helpman, Melitz, and Rubinstein (2008) consider data at the country level. See Lawless (2010) for a discussion of other definitions of the extensive margin.

interesting to see how trade frictions affect different regions of this conditional distribution, and quantile regression offers a flexible and practical way of doing that.

In this section we use data on sectoral exports from UN Comtrade for the years 1999-2001 to estimate quantile regression models for the number of sectors exporting from origin country j to a destination country i in year t. For estimation purposes we focus on the percentage of sectors exporting from j to i in t, denoted S_{ijt}, which is bounded between 0 and 1. In particular, in our sample, S_{ijt} varies between 0 and 0.99 with about 50% of the 137634 observations being equal to 0.[21]

Table 4 provides the definition of the regressors used, which were mainly obtained from CIA's World Factbook and CEPII; all the models also include yearly importer and exporter dummies, the multilateral resistance terms suggested by Anderson and van Wincoop (2003).[22] See Santos Silva, Tenreyro, and Wei (2014) for further details on the data, including their sources.

Table 5 presents parameter estimates and corresponding standard errors for different quantiles of S_{ijt} given x_{ijt}, whose functional form is given by (2) and (3).[23] Additionally, the table reports the value of the objective function evaluated at the estimates, and an R^2 defined as the square of the correlation between S_{ijt} and the fitted values of $\mathrm{Q}_{S_{ijt}}(\tau|x_{ijt})$.

[21]Some of the observed zeros may be the result of underreporting and therefore may not correspond to zero trade flows. Quantile regression should be less sensitive to this problem than other methods commonly used in empirical work. Indeed, for observations with the τ-th conditional quantile above the true (unobserved) value of trade, the rounding to zero will not affect the location of the conditional quantile. Likewise, the existence of underreporting will not shift the conditional quantile for observations where the true probability mass at zero is at least equal to τ. Therefore, the rounding towards zero will only be a problem for observations where the τ-th conditional quantile is between zero and the true (unobserved) value of trade.

[22]Non-linear models with importer and exporter dummies can suffer from a subtle form of the incidental parameter problem (Hahn and Newey, 2004). When this happens, the consistency of the estimator is not affected, but standard hypotheses tests and confidence intervals may be invalid. We investigated this problem in numerous simulation experiments and found that in this context the standard t-test for the null hypothesis that the parameter is zero has the desired size.

[23]The standard errors are clustered by country pair, see Parente and Santos Silva (2016) for details.

Given the nonlinearity of the models, the estimates of β reported in Table 5 are not directly comparable across quantiles, but the estimates of γ for different values of τ provide interesting information. In particular, for $\tau = 0.90$ the estimated value of γ is close to zero, suggesting that $\mathrm{Q}_{S_{ijt}}(0.9|x_{ijt})$ is positive for most observations. Indeed, in the sample, the estimated value of $\mathrm{Q}_{S_{ijt}}(0.9|x_{ijt})$ equals zero for only 32% of the observations. The estimates of γ increase as τ goes down, reflecting the fact that lower quantiles become flat at zero for smaller values of $x'_{ijt}\beta$; the estimated values of $\mathrm{Q}_{S_{ijt}}(0.5|x_{ijt})$ and $\mathrm{Q}_{S_{ijt}}(0.1|x_{ijt})$ are equal to zero for about 72% and 87% of the observations, respectively.

Table 4: Definition of the regressors*

Log distance	Natural logarithm of distance between capitals (in kilometers)
Border	Dummy equal to 1 when the countries share a land border
Both islands	Dummy equal to 1 if neither country has land borders
Both landlocked	Dummy equal to 1 if both countries are landlocked
Colonial tie	Dummy equal to 1 either if the importer has ever colonized or been a colony of the exporter or if the two countries were once part of the same country
Common currency	Dummy equal to 1 if either both countries use the same currency or if the exchange rates between their currencies is fixed
RTA	Dummy equal to 1 if the countries are at least in one common regional trade agreement
Common language	Dummy equal to 1 if the countries share an official language
Both WTO	Dummy equal to 1 if the countries are members of the WTO
Religion	Sum of the products of the shares of the population in each of the partners that are Catholic, Muslim, or Protestant

* All models also include yearly importer and exporter dummies; see Santos Silva, Tenreyro, and Wei (2014) for further details on the data.

Table 5: Parameter estimates (and standard errors)

$\tau =$	0.10	0.30	0.50	0.70	0.90
Log distance	−0.6045	−0.6706	−0.8023	−0.9212	−0.9974
	(0.0210)	(0.0197)	(0.0233)	(0.0189)	(0.0281)
Border	0.2400	0.3004	0.6639	0.6637	0.8258
	(0.0798)	(0.0834)	(0.0799)	(0.0845)	(0.1631)
Both islands	0.1212	0.1153	0.1447	0.2252	0.5343
	(0.1159)	(0.0736)	(0.0764)	(0.0662)	(0.3462)
Both landlocked	−0.0697	0.0148	0.0276	0.1197	0.2583
	(0.2102)	(0.0781)	(0.0970)	(0.1548)	(0.0979)
Colonial tie	0.3977	0.5176	0.6832	0.8144	1.0810
	(0.0689)	(0.0479)	(0.0719)	(0.1030)	(0.0812)
Common currency	0.0014	−0.0608	−0.0653	0.0115	0.2346
	(0.0627)	(0.0563)	(0.0653)	(0.1441)	(0.0618)
RTA	0.1791	0.1694	0.1321	0.1025	0.0906
	(0.0391)	(0.0362)	(0.0383)	(0.0397)	(0.0499)
Common language	0.3099	0.3684	0.3901	0.4508	0.5446
	(0.0571)	(0.0336)	(0.0439)	(0.0463)	(0.0448)
Both WTO	0.0686	0.0465	0.2093	0.5099	0.4516
	(0.1047)	(0.0743)	(0.0965)	(0.1056)	(0.1455)
Religion	0.1184	0.1682	0.1604	0.2344	0.3400
	(0.0489)	(0.0488)	(0.0518)	(0.0424)	(0.0675)
γ	0.0265	0.0149	0.0061	0.0026	0.0006
	(0.0021)	(0.0009)	(0.0003)	(0.0002)	(0.0002)
Objective function	163.8245	391.6722	510.9052	503.3544	298.5422
R^2	0.8704	0.8998	0.9149	0.9182	0.8878

Table 6 presents the average across the entire sample of the partial effects on $Q_{5132 \times S_{ijt}}(\tau|x_{ijt})$ of each of the regressors; these partial effects allow us to compare the effects of the regressors on different quantiles.[24] As usual, for the continuous variables (Log distance and Religion) the partial effects are computed as the derivatives of

[24]Notice that, to facilitate their interpretations, the partial effects are measured in numbers of sectors.

the estimate of $Q_{5132 \times S_{ijt}}(\tau|x_{ijt})$ with respect to the regressors (notice that the derivative is with respect to log distance, not distance itself), while for the dummy variables the partial effect is defined as the difference between the estimate of $Q_{5132 \times S_{ijt}}(\tau|x_{ijt})$ with the dummy equal to 1 and with the dummy equal to 0.[25]

Before looking in more detail into the results in Table 6, it is important to be clear about their interpretation. First, it is important to keep in mind that we report average partial effects and that the actual partial effect for a given country-pair in a given year can be very different from the average values reported in Table 6. Second, these partial effects should not be confused with the total effects of the regressors in general equilibrium (see Egger et al., 2011, for an interesting discussion of this issue). Third, we emphasize that we are estimating conditional quantiles and not a structural model. This is important because some of the regressors we consider are sometimes treated as endogenous variables (see, e.g., Baier and Bergstrand, 2004 and 2007, Egger, Egger, and Greenaway, 2008, Egger et al., 2011, and Baier, Bergstrand, and Feng, 2014) but, by definition, all regressors are weakly exogenous with respect to the parameters of conditional quantiles and conditional expectations.[26] Therefore, our results have a precise and interesting interpretation, but should not be seen as causal. Finally, we believe that it is interesting to interpret these quantile regression results in the context of the "dark trade costs" discussed by Head and Mayer (2013), who emphasize the effect on trade of historical factors such as differences in tastes and the legacy of colonial relations and past conflicts.[27] In particular, we can view the dispersion of the values of S_{ijt} for a

[25]In interpreting these results, it is important to keep in mind that $Q_{S_{ijt}}(\tau|x_{ijt})$ is a function of τ. Therefore, in general, a variable with the same effect for all of the quantiles will have a proportional effect that declines with τ. Computing proportional effects in this context is not particularly informative because $Q_{S_{ijt}}(\tau|x_{ijt})$ can be equal to zero.

[26]See Goldberger (1991, pp. 338-41) for a discussion of this issue in the least squares context.

[27]More generally, we can consider as "dark costs" all the factors that affect trade and are not explicitly considered as regressors. For example, in the case of the sector margin, an important dark cost is the similarity of the pattern of specialization of the partners; this certainly contributes to the relatively low values of S_{ijt} that are observed when both partners border the Persian Gulf.

given value of regressors as resulting from variations in the unobservable "dark costs". Therefore, we can interpret the results for the upper quantiles as describing how the regressors affect the conditional distribution of S_{ijt} for pairs of countries with relatively small unobservable trade costs and, conversely, results for the lower quantiles as describing how the regressors affect the conditional distribution for pairs with relatively large dark costs.

The results in Table 6 show that the average impact of most regressors on the upper tail of the distribution is much stronger than their average impact on the lower tail; that is, the effect of the regressors tends to be stronger when the dark costs are smaller. The main exception to this pattern is the effect of RTA, which is remarkably stable across different parts of the distribution. The stability of this effect is somewhat surprising and it would be interesting to investigate why the pattern of the effect of RTA is so different from that of other regressors.[28]

One of the regressors with a very heterogeneous effect is Both WTO. In a recent paper, Dutt, Mihov, and Van Zandt (2013) studied the effect of WTO membership on the two margins of trade and found that it has a large positive impact on the extensive margin defined at the sector level (cf. Felbermayr and Kohler, 2010, and Buono and Lalanne, 2012). The results in Table 6 show that WTO membership shifts upwards the upper tail of the conditional distribution of the sector margin, but it does not seem to affect the lower tail of the distribution. The heterogeneous effect of WTO membership on the conditional distribution of S_{ijt} is in line with the results of Subramanian and Wei (2007), who found that the impact of WTO membership has been strong but very uneven.

Perhaps the most striking result in Table 6 relates to the effect of sharing a common currency. Indeed, our results indicate that sharing a common currency will not affect most of the conditional distribution of the sector margin, but nevertheless having a

[28] See Egger et al. (2011), Baier, Bergstrand, and Feng (2014), and Dai, Yotov, and Zylkin (2014) for recent results on the effects of free trade agreements.

common currency has a positive effect on the very top of the distribution.[29] De Sousa

Table 6: Average partial effects on $Q_{5132 \times S_{ijt}}(\tau|x_{ijt})$ (and standard errors)

$\tau =$	0.10	0.30	0.50	0.70	0.90
Log distance	−40.9041	−52.8695	−69.2165	−90.8424	−126.6567
	(2.5042)	(3.1713)	(2.0634)	(4.0717)	(19.4864)
Border	17.5356	25.8398	68.2004	77.1764	127.3760
	(7.2854)	(7.6543)	(9.7592)	(11.9511)	(35.7227)
Both islands	8.5229	9.4013	12.9777	23.4866	76.5765
	(8.5736)	(6.1678)	(7.1068)	(7.3798)	(56.5755)
Both landlocked	−4.6117	1.1720	2.4010	12.1682	34.8511
	(13.6249)	(6.2091)	(8.4942)	(16.3102)	(15.0411)
Colonial tie	30.4533	47.3377	70.4927	98.2676	176.8734
	(7.7672)	(4.8199)	(8.8614)	(15.2130)	(32.8242)
Common currency	0.0938	−4.7059	−5.5289	1.1389	31.5399
	(4.2464)	(4.2649)	(5.4377)	(14.2900)	(9.9439)
RTA	12.7087	13.9236	11.7481	10.3435	11.7386
	(3.8074)	(3.1143)	(3.5053)	(4.1272)	(6.7782)
Common language	22.6517	31.5320	36.3913	48.3441	75.9622
	(6.7223)	(3.2211)	(4.4689)	(5.7433)	(13.1991)
Both WTO	4.5932	3.6427	17.6716	48.3787	56.0211
	(6.6352)	(5.8051)	(8.0220)	(10.2206)	(18.4772)
Religion	8.0087	13.2624	13.8368	23.1154	43.1697
	(3.8131)	(3.9385)	(4.5007)	(4.3285)	(10.7032)

(2012) has noted that since 1999 the impact of currency unions on trade is minimal, and therefore it is not entirely surprising to find that sharing a common currency has little impact on the conditional distribution of the product margin. However, our results suggest that the common currency dummy has some effect on the upper tail of the

[29] A similar pattern is observed for Both landlocked.

distribution.[30] The fact that the effect is concentrated on the upper tail is in line with the findings of Costa-i-Font (2010) who reports that, for the particular case of the euro, the effect of the common currency on trade is stronger in regions that are more open to trade.[31] This set of results on the heterogeneous effect of sharing a common currency suggests that the impact of currency unions on trade is rather complex, and more research is needed to fully understand the channels through which this effect takes place.[32]

In summary, our results imply that observable trade frictions have very heterogeneous effects on the conditional quantiles of the extensive margin of trade defined at the sector level. In particular, we find that changes in the regressors result in changes in the conditional distribution that are generally more pronounced in regions corresponding to pairs of countries that trade more than pairs with similar observable characteristics. That is, the effects of changes in observable trade frictions are generally stronger in regions corresponding to pairs of countries with low levels of dark trade costs. The intensity of this pattern, however, varies considerably across regressors and further research is needed to explain these differences.

5. CONCLUDING REMARKS

Corner solutions data are often found in empirical applications and their analysis typically requires specialized tools. In particular, estimation of quantile regressions for this kind of data cannot be performed using standard specifications and corresponding estimators because the conditional quantiles of corner solutions data are identically zero for some observations.

[30]Santos Silva, Tenreyro, and Wei (2014) find that the common currency dummy also impacts the conditional expectation of the percentage of sectors exporting from j to i, and that this impact is particularly important in regions where the conditional mean is in the upper tail of the distribution.

[31]In their well-known study of the effects of the euro, Baldwin and Di Nino (2006) also find that sharing a common currency has an heterogeneous effect on the extensive margin of trade.

[32]See, e.g., Santos Silva and Tenreyro, (2010) and Bergin and Lin (2012) for more on the effect of currency unions on trade.

We developed a simple method for the estimation of quantile regressions for corner solutions data, focusing on the case where the variate of interest is doubly-bounded and has a mass-point at zero. The proposed estimator can be adapted to deal with other problems such as non-linear and censored quantile regression, and can be implemented using standard software.

We use the proposed method to study the determinants of the extensive margin of trade, defined as the number (or share) of sectors exporting from origin country j to a destination country i in year t. Our findings suggest that most regressors have very different impacts on different parts of the distribution. In particular, we find that for most regressors the impact on the upper tail is much larger than their impact on the lower tail. This implies that changes in the regressors generally result in changes in the conditional distribution of the number of exporting sectors that are more pronounced in regions corresponding to pairs of countries that trade more than pairs with similar observable characteristics.

REFERENCES

Anderson, J. and van Wincoop, E. (2003). "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review*, 93, 170-192.

Angrist, J.D.; Chernozhukov, V. and Fernandez-Val, I. (2006). "Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure," *Econometrica*, 74, 539-563.

Aranda-Ordaz, F.J. (1981). "On two families of transformations to additivity for binary response data," *Biometrika*, 68, 357-64.

Ardelean, A. and Lugovskyy V. (2010). "Domestic Productivity and Variety Gains From Trade," *Journal of International Economics*, 80, 280-291.

Baier, S.L. and Bergstrand, J.H. (2004). "Economic Determinants of Free Trade Agreements." *Journal of International Economics*, 64, 29-63.

Baier, S.L. and Bergstrand, J.H. (2007). "Do Free Trade Agreements Actually Increase Members' International Trade?," *Journal of International Economics*, 71, 72-95.

Baier, S.L., Bergstrand, J.H. and Feng, M. (2014). "Economic Integration Agreements and the Margins of International Trade," *Journal of International Economics*, 93, 339-350.

Baldwin, R.E. and Di Nino, V. (2006). "Euros and Zeros: The Common Currency Effect on Trade in New Goods," NBER, Working Paper No. 12673.

Bergin, P.R. and Lin, C.-Y. (2012). "The Dynamic Effects of a Currency Union on Trade," *Journal of International Economics*, 87, 191-204.

Bottai, M., Cai, B. and McKeown, R.E. (2010). "Logistic Quantile Regression for Bounded Outcomes," *Statistics in Medicine*, 29, 309-317.

Bottai, M., Orsini, N. and Geraci, M. (2015). "A Gradient Search Maximization Algorithm for the Asymmetric Laplace Likelihood," *Journal of Statistical Computation and Simulation*, 85, 1919-1925.

Broda, C. and Weinstein, D.E. (2006). "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, 121, 541-585.

Buchinsky, M. (1994). "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62, 405-458.

Buchinsky, M. (1995). "Estimating the Asymptotic Covariance Matrix for Quantile Regression Models a Monte Carlo Study," *Journal of Econometrics*, 68, 303-38.

Buchinsky, M. and Hahn, J. (1998). "An Alternative Estimator for the Censored Quantile Regression Model," *Econometrica*, 66, 653-671.

Buono, I. and Lalanne, G. (2012). "The Effect of the Uruguay Round on the Intensive and Extensive Margins of Trade," *Journal of International Economics*, 86, 269-283.

Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: Methods and Applications.* Cambridge: Cambridge University Press.

Chamberlain, G. (1994). "Quantile Regression, Censoring and the Structure of Wages," in Sims, C.A. (ed.), *Advances in Econometrics*, 171-210, Cambridge: Cambridge University Press.

Chaney, T. (2008). "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 98, 1707-1721.

Chernozhukov, V. and Hong, H. (2002). "Three-Step Censored Quantile Regression and Extramarital Affairs," *Journal of American Statistical Association*, 97, 872–882.

Cochran, W.G. (1952). "The χ^2 Test of Goodness of Fit," *Annals of Mathematical Statistics*, 23, 315-345.

Costa-i-Font, J. (2010). "Regional Single Currency Effects on Bilateral Trade with the European Union," LSE Research Online Documents on Economics 53292.

Dai, J. and Sperlich, S. (2010). "Simple and Effective Boundary Correction for Kernel Densities and Regression with an Application to the World Income and Engel Curve Estimation," *Computational Statistics & Data Analysis*, 54, 2487-2497.

Dai, M., Yotov, Y.V. and Zylkin, T. (2014). "On the Trade-Diversion Effects of Free Trade Agreements," *Economics Letters*, 122, 321-325.

Dennis, A. and Shepherd, B. (2007). "Trade Costs, Barriers to Entry, and Export Diversification in Developing Countries," The World Bank Policy Research Working Paper No. 4368, Washington, D.C.

Dennis, A. and Shepherd, B. (2011). "Trade Facilitation and Export Diversification," *The World Economy*, 34, 101-122.

De Sousa, J. (2012). "The Currency Union Effect on Trade is Decreasing Over Time," *Economics Letters*, 117, 917-920.

Duan N., Manning, W.G. Jr, Morris, C.N. and Newhouse J.P. (1983). "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economic Statistics*, 1, 115-126.

Dutt, P., Mihov, I., and Van Zandt, T. (2013). "The Effect of WTO on the Extensive and the Intensive Margins of Trade," *Journal of International Economics*, 91, 204-219.

Eaton, J., Kortum, S. and Kramarz, F. (2004). "Dissecting Trade: Firms, Industries, and Export Destinations," *American Economic Review*, 94, 150-154

Eicher, T.S. and Kuenzel, D.J. (2016). "The Elusive Effects of Trade on Growth: Export Diversity and Economic Take-Off," *Canadian Journal of Economics*, forthcoming.

Egger, H., Egger, P.H. and Greenaway, D. (2008). "The Trade Structure Effects of Endogenous Regional Trade Agreements." *Journal of International Economics*, 74, 278-98.

Egger, P.H., Larch, M., Staub, K. and Winkelmann, R. (2011). "The Trade Effects of Endogenous Preferential Trade Agreements," *American Economic Journal: Economic Policy*, 3, 113-143.

Fitzenberger, B. (1997a). "A Guide to Censored Quantile Regressions," in Maddala, G.S. and Rao, C.R. (eds.), *Handbook of Statistics, Volume 15: Robust Inference*, 405-437, New York, NY: North-Holland.

Fitzenberger, B. (1997b). "Computational Aspects of Censored Quantile Regression," in *Lecture Notes-Monograph Series, 31: L_1-Statistical Procedures and Related Topics*, 171-186, Hayward, CA: Institute of Mathematical Statistics.

Feenstra, R.C. and Kee, H.-L. (2008). "Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity", *Journal of International Economics*, 74, 500-514.

Feenstra, R.C. and Ma, H. (2014). "Trade Facilitation and the Extensive Margin of Exports," *Japanese Economic Review*, 65, 158–177.

Feenstra, R.C., Madani, D., Yang,T.-H. and Liang, C.-Y. (1999). "Testing Endogenous Growth in South Korea and Taiwan," *Journal of Development Economics*, 60, 317-341.

Felbermayr, G. and Kohler, W. (2010). "Modelling the Extensive Margin of World Trade: New Evidence on GATT and WTO Membership," *The World Economy*, 33, 1430-1469.

Goldberger, A.S. (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.

Hahn, J. and Newey, W. (2004). "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319.

Hall, B.H. and Cummins, C. (2009). *TSP 5.1 User's Guide*, Palo Alto, CA: TSP International.

Head, K. and Mayer, T. (2013). "What Separates Us? Sources of Resistance to Globalization," *Canadian Journal of Economics*, 46, 1196-1231.

Helpman, E., Melitz, M. and Rubinstein, Y. (2008). "Estimating Trade Flows: Trading Partners and Trading Volumes," *The Quarterly Journal of Economics*, 123, 441-487.

Hillberry, R. and Hummels, D. (2008). "Trade Responses to Geographic Frictions: A Decomposition Using Micro-Data," *European Economic Review*, 52, 527-550.

Hillberry, R. and McDaniel, C. (2002). "A Decomposition of North American Trade Growth since NAFTA," Working Papers 15866, United States International Trade Commission, Office of Economics.

Hummels, D. and Klenow, P.J. (2005). "The Variety and Quality of a Nation's Exports," *American Economic Review*, 95, 704-723.

Judd, K.L. (1998). *Numerical Methods in Economics*. Cambridge, MA: The MIT Press.

Kehoe, T.J and Ruhl, K.J. (2013). "How Important Is the New Goods Margin in International Trade?," *Journal of Political Economy*, 121, 358-392.

Kim, T.-H. and White, H. (2003). "Estimation, Inference and Specification Testing and Possibly Misspecified Quantile Regression," *Advances in Econometrics*, 17, 107-132.

Koenker, R. (2005). *Quantile Regression*, New York, NY: Cambridge University Press.

Koenker, R. (2008). "Censored Quantile Regression Redux," *Journal of Statistical Software*, 27, http://www.jstatsoft.org/v27/i06.

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

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

Krugman, P. (1979), "Increasing Returns, Monopolistic Competition, and International Trade," *Journal of International Economics*, 9, 469-479.

La Porta, R., Lopez-de-Silanes, F. and Zamarripa, G. (2003). "Related Lending," *Quarterly Journal of Economics*, 118, 231-268.

Lawless, M. (2010). "Deconstructing Gravity: Trade Costs and Extensive and Intensive Margins," *Canadian Journal of Economics*, 43, 1149-1172.

Lewis, A.S. and Overton, M.L. (2013). "Nonsmooth Optimization via Quasi-Newton Methods," *Mathematical Programming Series A*, 141, 135-163.

Liu, Y. and Bottai, M. (2009). "Mixed-Effects Models for Conditional Quantiles with Longitudinal Data," *The International Journal of Biostatistics*, 5, Article 28.

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

Machado, J.A.F. and Santos Silva, J.M.C. (2008). *Quantiles for Fractions and Other Mixed Data*, Department of Economics, University of Essex, Discussion Paper No 656.

Manova, K. (2013). "Credit Constraints, Heterogeneous Firms, and International Trade," *The Review of Economic Studies*, 80, 711-744.

Melitz, M.J. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695-725.

Oberhofer, W. (1982). "The Consistency of Nonlinear Regression Minimizing the L_1-Norm," *Annals of Statistics*, 10, 316-319.

Papke, L.E. and J.M. Wooldridge (1996). "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates," *Journal of Applied Econometrics*, 11, 619-632.

Parente, P.M.D.C. and Santos Silva, J.M.C. (2016), Quantile Regression with Clustered Data, *Journal of Econometric Methods*, 5, 1-15.

Powell, J.L. (1984). "Least Absolute Deviation Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303-325.

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

Ramalho, E.A., Ramalho, J.J.S. and Murteira, J.M.R. (2011). "Alternative estimating and testing empirical strategies for fractional regression models," *Journal of Economic Surveys*, 25, 19-68.

Romer, P. (1994). "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions," *Journal of Development Economics*, 43, 5-38.

Santos Silva, J.M.C. and Tenreyro, S. (2006), "The Log of Gravity," *The Review of Economics and Statistics*, 88, 641-658.

Santos Silva, J.M.C. and Tenreyro, S. (2010), "Currency Unions in Prospect and Retrospect," *Annual Review of Economics*, 2, 51-74.

Santos Silva, J.M.C., Tenreyro, S. and Wei, K. (2014), "Estimating the Extensive Margin of Trade," *Journal of International Economics*, 93, 67-75.

StataCorp. (2013). *Stata Release 13. Statistical Software.* College Station, TX: StataCorp LP.

Subramanian, A. and Wei, S.-J. (2007). "The WTO Promotes Trade, Strongly but Unevenly," *Journal of International Economics*, 72, 151-175.

Wooldridge, J.M. (2002). *Econometric analysis of cross section and panel data*, Cambridge, MA: MIT Press.

Further Results on Quantiles and Corners*

José A. F. Machado
Faculty of Business and Management, Regent's University London

J. M. C. Santos Silva
School of Economics, University of Surrey

Kehai Wei
Department of Economics, University of Essex

December 10, 2015

Abstract

In this note we illustrate the application of the method proposed by Machado, Santos Silva, and Wei (2015) using an example where the dependent variable has a lower bound at zero but no upper bound.

Cameron and Trivedi (2009) use data from the 2001 wave of the Medical Expenditure Panel Survey to illustrate the estimation of different models for corner-solutions data. This dataset contains 3328 observations on employed individuals aged between 21 and 64 who are covered by private health insurance, and is a subset of the data used by Deb, Munkin and Trivedi (2006). Here, these data are used to illustrate the application of the method proposed by Machado, Santos Silva, and Wei (2015) in the case where the dependent variable has a lower bound at zero but no upper bound.

The variate of interest in the models considered by Cameron and Trivedi (2009) are the ambulatory expenditures (AmbExp).[1] As is typical with this kind of data, AmbExp has a very skewed distribution, and it is equal to 0 for about 16% of the observations. As in Cameron and Trivedi (2009), we will consider six covariates: Age, age in years

*Address for correspondence: João Santos Silva, School of Economics, University of Surrey, Guildford GU2 7XH, United Kingdom. Tel: +44 (0)1483686956. E-mail: jmcss@surrey.ac.uk.

[1]To facilitate the estimation in Stata (2013) we use AmbExp in thousands of dollars.

divided by 10; FEMALE, a dummy variable equal to 1 for females, being zero otherwise; EDUC, years of schooling of decision maker; BLHISP, a dummy variable equal to 1 for blacks or hispanics, being zero otherwise; TOTCHR, number of chronic diseases; and INS, a dummy variable equal to 1 if the individual has either a preferred provider organization (PPO) or a health maintenance organization (HMO) type insurance, being zero for individuals with less restrictive fee-for-service (FFS) plans. Further information on the data, including descriptive statistics, is provided in Cameron and Trivedi (2009) and in Deb, Munkin and Trivedi (2006).

Cameron and Trivedi (2009) estimate several models for AMBEXP, including a sample selection model (Heckman, 1979) of the form

$$\Pr\left(\text{AMBEXP} > 0|x\right) = \Pr\left(x'\delta + e_1 > 0|x\right),$$
$$\ln\left(\text{AMBEXP}\right) = x'\lambda + e_2, \quad \text{for AMBEXP} > 0$$

where x denotes the vector of covariates and e_1 and e_2 are random disturbances assumed to follow a bivariate normal distribution with correlation ρ and $\text{Var}\left(e_1\right) = 1$ and $\text{Var}\left(e_2\right) = \sigma^2$. Models of this type have often been used to describe medical expenditures (see Duan et al., 1983, or Jones, 2000, for a survey), but they rely heavily on strong distributional assumptions. It is, therefore, interesting to consider the estimation of conditional quantiles of AMBEXP, which can provide information on the impact of the covariates on different features of the conditional distribution of interest using only relatively mild assumptions.

To model the quantiles of AMBEXP we follow Machado, Santos Silva, and Wei (2015) and specify

$$Q_{\text{AMBEXP}}\left(\theta|x\right) = \max\left\{0, \exp\left(x'\beta\right) - \gamma\right\}, \qquad (1)$$

where, as in Cameron and Trivedi (2009), $x'\beta$ has the form

$$x'\beta = \beta_0 + \beta_1\text{AGE} + \beta_2\text{FEMALE} + \beta_3\text{EDUC} + \beta_4\text{BLHISP} + \beta_5\text{TOTCHR} + \beta_6\text{INS}.$$

Table 1 displays the estimated parameters and corresponding standard errors for all the models.[2] As a benchmark, Table 1 also includes the estimates obtained using the

[2]As in Machado, Santos Silva, and Wei (2015), estimation was performed using the BFGS algorithm as implemented in the `ml` command in Stata (StataCorp., 2013); version control was used to run the code

Table 1: Parameter estimates

	Quantile regression			Sample Selection	
	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	1st part	2nd part
Intercept	-0.338	-0.139	1.179	-0.724	-1.870
	(0.384)	(0.313)	(0.593)	(0.192)	(0.226)
AGE	0.045	0.077	0.065	0.098	0.212
	(0.014)	(0.018)	(0.027)	(0.027)	(0.023)
FEMALE	0.163	0.181	0.109	0.644	0.350
	(0.041)	(0.043)	(0.047)	(0.060)	(0.060)
EDUC	0.017	0.019	0.010	0.070	0.019
	(0.005)	(0.006)	(0.005)	(0.011)	(0.011)
BLHISP	-0.101	-0.124	-0.079	-0.373	-0.220
	(0.033)	(0.031)	(0.037)	(0.062)	(0.059)
TOTCHR	0.247	0.312	0.194	0.795	0.541
	(0.042)	(0.032)	(0.068)	(0.071)	(0.039)
INS	0.030	0.012	-0.030	0.182	-0.030
	(0.024)	(0.021)	(0.019)	(0.063)	(0.051)
γ	1.071	1.259	3.885	—	—
	(0.338)	(0.305)	(2.076)	—	—
σ	—	—	—	—	1.271
	—	—	—	—	(0.018)
ρ	—	—	—	-0.124	
	—	—	—	(0.144)	
Objective function	1042.188	1772.311	1963.860	-5838.397	
R^2	0.153	0.157	0.163	0.139	

Standard errors in parenthesis: misspecification robust for quantile regressions; Hessian based for the sample selection model.

selection model considered by Cameron and Trivedi (2009, p. 559). The most noteworthy feature of the results in Table 1 is that the estimates of γ increase with θ. In spite of this, and because the intercept also increases with θ, the proportion of observations for which $Q_{\text{AmbExp}}(\theta|x)$ is equal to zero decreases as θ increases. Indeed, for $\theta \in \{0.25, 0.50, 0.75\}$,

with the default convergence criteria for Stata 11. This is needed because from Stata 12 the convergence criteria for the BFGS algorithm uses the Hessian of the objective function.

Table 2: Average partial effects

	Quantiles			Mean
	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	
AGE	0.050	0.150	0.365	0.387
	(0.015)	(0.023)	(0.069)	(0.043)
FEMALE	0.183	0.353	0.603	0.736
	(0.024)	(0.042)	(0.078)	(0.077)
EDUC	0.019	0.037	0.054	0.049
	(0.006)	(0.009)	(0.013)	(0.018)
BLHISP	−0.106	−0.235	−0.430	−0.444
	(0.028)	(0.040)	(0.073)	(0.087)
TOTCHR	0.274	0.607	1.069	1.112
	(0.053)	(0.119)	(0.100)	(0.077)
INS	0.034	0.023	−0.162	−0.010
	(0.026)	(0.043)	(0.073)	(0.088)

Standard errors in parenthesis: misspecification robust for quantile regressions; Hessian based for the sample selection model.

the estimated values of $Q_{\text{AMBEXP}}(\theta|x)$ are equal to zero for about 22.9, 4.9, and 1.1 percent of the observations, respectively.

As for the effects of the covariates, the results in Table 1 are not very informative because all the models have different functional forms and therefore the coefficient estimates are not directly comparable. To overcome this problem, Table 2 presents the average across the entire sample of the partial effects of each of the regressors on $\mathrm{Q}_{\text{AMBEXP}}(\tau|x_{ij})$, and on the conditional mean implicit in Heckman's sample-selection model (see Cameron and Trivedi, 2009, p. 563, for details).[3]

The results in Table 2 show that the partial effects of the regressors vary widely across the different conditional quantiles, generally increasing (in absolute value) with θ. Therefore, looking only at the partial effects on the conditional mean, which in this

[3]The partial effects for binary an non-binary variables are computed as usual (see, e.g., Wooldridge, 2002, p. 15).

example are generally close to those obtained for $\theta = 0.75$, may give a very incomplete picture of how the covariates affect the conditional distribution of the AmbExp.

As in Deb, Munkin and Trivedi (2006), it is particularly interesting to consider the effect of the type of insurance plan on AmbExp.[4] The results in Table 2 show that Ins has a small and statistically insignificant effect on the conditional mean and on the first two conditional quartiles, but has a sizable and statistically significant effect on the third conditional quartile. These results suggest that having the more restrictive HMO or PPO insurance plans has little effect on most of the distribution, but has a strong negative effect on its upper tail, which however is not enough to generate a significant effect on the mean implied by the sample selection model. These findings are also very different from what could be inferred from the sample selection results presented in Table 1, which suggest that Ins is only significant in going from a zero to a positive expenditure. Therefore, the selection model and the implied conditional mean mask the very different effects that changes in Ins have on different areas of the conditional distribution of AmbExp.

Although perhaps less striking, the results for other covariates also confirm that focusing on the partial effects on the conditional mean provides an incomplete, and even somewhat misleading, picture of the effects of the covariates on the conditional distribution of interest.

Finally, we note that for this dataset the linear specification used by Powell (1984, 1986) for censored quantile regression leads to higher values of the objective function and fails to reveal that Ins has a significant effect in the third quartile. These results suggest that in the context of corner-solutions data, the proposed non-linear model can have important advantages over a specification with constant marginal effects.

[4] Deb, Munkin and Trivedi (2006) consider the possible endogeneity of Ins but do not reject the null hypothesis of exogeneity.

REFERENCES

Cameron, A.C. and Trivedi, P.K. (2009). *Microeconometrics using Stata*, Revised Edition. College Station (TX): Stata Press.

Deb, P., Munkin, M.K. and Trivedi, P.K. (2006). "Bayesian analysis of the two-part model with endogeneity: application to health care expenditure," *Journal of Applied Econometrics*, 21, 1081-1099.

Duan N., Manning, W.G. Jr, Morris, C.N. and Newhouse J.P. (1983). "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economic Statistics*, 1, 115-126.

Heckman, J.J. (1979). "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-161.

Jones, A.M. (2000). "Health Econometrics," in Newhouse, J.P. and Culyer, A.J. (eds.) *Handbook of health economics*, Vol. 1A, Ch. 6, 265-344, Amsterdam: Elsevier.

Machado, J.A.F., Santos Silva, J.M.C. and Wei, K. (2015). Quantiles, Corners, and the Extensive Margin of Trade, *mimeo*.

StataCorp. (2013). *Stata Release 13. Statistical Software.* College Station (TX): StataCorp LP.

Wooldridge, J.M. (2002). *Econometric analysis of cross section and panel data*, Cambridge, MA: MIT Press.