# Asymptotics for Estimation of Truncated Infinite-Dimensional Quantile Regressions

version of: 25. 5. 2006

Serguei Zernov<sup>1</sup> , Victoria Zinde-Walsh<sup>2</sup> and John W. Galbraith<sup>3</sup>

Department of Economics McGill University 855 Sherbrooke St. West Montreal, Quebec H3A 2T7 CANADA

### Abstract

Many processes can be represented in a simple form as infinite-order linear series. In such cases, an approximate model is often derived as a truncation of the infinite-order process, for estimation on the finite sample. The literature contains a number of asymptotic distributional results for least squares estimation of such finite truncations, but for quantile estimation, only results for finite-order processes are available at a level of generality that accommodates time series processes. Here we establish consistency and asymptotic normality for conditional quantile estimation of truncations of such infinite-order linear models, with the truncation order increasing in sample size. The proofs use the generalized functions approach and allow for a wide range of time series processes may be represented as an AR( $\infty$ ) or an MA( $\infty$ ); here we use a simulation to illustrate the degree of conformity of finite-sample results with the asymptotics, in case of a truncated AR representation of a moving average.

Key words: generalized function,  $L_1$ -norm, LAD, quantile regression

\* The authors thank the *Fonds québécois de la recherche sur la société et la culture* (FQRSC) and the Social Sciences and Humanities Research Council of Canada (SSHRC) for support of this research.

<sup>&</sup>lt;sup>1</sup>Also affiliated with Toronto-Dominion Asset Management.

<sup>&</sup>lt;sup>2</sup>Also affiliated with the Centre Inter-Universitaire de Recherche en Économie Quantitative (CIREQ).

<sup>&</sup>lt;sup>3</sup>Also affiliated with CIREQ and the Centre Inter-Universitaire de Recherche en Analyse des Organisations (CIRANO).

### 1. Introduction

Many processes can be represented as infinite-order linear series or as infiniteorder approximations involving other processes. An example important in econometrics is the time series context in which such infinite-order series may have an  $AR(\infty)$ ,  $MA(\infty)$  or  $ARCH(\infty)$  form. The autoregressive model in particular has been very widely applied, despite the fact that a finite-order autoregression may not be a plausible representation of the true process. Instead, the justification for its use lies in the fact that when a low-order parametric model describing a process precisely is not known or is not convenient for estimation, the process may nonetheless be well characterized by a finite truncation of an infinite-order representation, where the truncation order increases in sample size. The fundamental asymptotic results on LS estimation of such truncated autoregressions date to Berk (1974) and Lewis and Reinsel (1985). Numerous treatments of econometric time series problems have used such results, for example Braun and Mittnik (1993), Galbraith and Zinde-Walsh (1994, 1997), Lütkepohl and Poskitt (1996), Saikkonen and Lütkepohl (1996), Lütkepohl and Saikkonen (1997), Inoue and Kilian (2002), Gonçalves and Kilian (2005), and many others. Of course, infinite-order processes occur in many other circumstances, and the results of the present paper apply much more widely than to these time series cases.

Another important literature, originating with Koenker and Bassett (1978), addresses asymptotic properties of quantile estimation (of which Least Absolute Deviations, LAD, is a special case) of regression models. The primary technical challenge in this literature, shared by the present paper, arises from the nondifferentiability of the LAD or quantile criterion function. Although now fairly extensive, this literature in general treats the order of the estimated model as being finite, or treats the process as having i.i.d. errors, either of which may be inadequate for the treatment of many cases of interest in econometrics.

The present paper lies at the intersection of these two literatures and provides results which link the two and extend each of the classes of result. In particular, we present a general result on quantile estimation for finite truncations of infinitedimensional processes. This result implies consistency and asymptotic normality of estimates from truncations of a wide range of processes, including but not limited to the time series examples just given.

The next section of the paper describes key results in the two literatures to which the present paper is related. Section 3 describes the general result on consistent estimation of conditional quantiles in a process which may be represented as an infinite-order approximation to another process, truncated to finite order. Section 4 provides a simulation example in which a finite truncation of an  $AR(\infty)$  representation of an MA process is estimated by LAD regression.

The proof of Theorem 1 is given in the Appendix.

#### 2. Existing asymptotic results for OLS and quantile estimation

A number of asymptotic results for OLS estimation of truncations of infinitedimensional regressions are available. The critical requirements for consistent estimation are (i) an increase in the number of regressors as sample size increases, and (ii) a model in which the contribution of the infinite non-included part can be made small relative to that of the included part. Such requirements can be shown to be met in time series models such as the  $AR(\infty)$ , examined by Berk (1974), who provided the assumptions under which a rate of increase in the number of regressors satisfying (ii) can be determined. Berk's results and their applications and extensions have involved mainly processes considered in  $L_2$  Hilbert spaces (e.g. the space of stationary stochastic processes) and their  $L_2$ -norm approximations (usually by lagged values of the process, as in AR approximations). (Below we address a similar question concerning approximation in  $L_1$  space, corresponding with the absolute value criterion in LAD estimation for median regression, and with criteria based on quantiles for quantile regression.)

Berk studied a case in which a stationary process meeting weak regularity conditions is modelled as an autoregression. Because estimation is of a truncation of an infinite-order process, consistency requires that the order of truncation (that is, the order of the approximating autoregression) increase without bound as the sample size  $T \to \infty$ . The rate at which this order, k, must increase is a key result; it is sufficient that  $k^3/T \to 0$  and that  $T^{\frac{1}{2}} \sum_{i=1}^{\infty} (|a_{k+i}|) \to 0$ , where the  $\{a_{k+i}\}, i > 0$ , are the coefficients of the truncated part of the true process. Consistency is proven for estimation by least squares.

These results were extended to cover the prediction problem for infinite-order univariate AR processes by Bhansali (1978), and by Lewis and Reinsel (1985) to multivariate cases, both for estimation and prediction. The Lewis and Reinsel results, which involve conditions like Berk's on the rate of growth of truncation order, have been important in allowing econometric applications to finite truncations of important classes of multivariate model such as the infinite-order VAR; see for example Lütkepohl and Poskitt (1996), Saikkonen and Lütkepohl (1996), Lütkepohl and Saikkonen (1997).

The other literature immediately relevant concerns the asymptotics of quantile regression estimates. The proof of the consistency of LAD (or minimum  $L_1$ -norm) estimates dates to Koenker and Bassett (1978), and has been developed and generalized by several authors including Bloomfield and Steiger (1983), Chen et al.

(1990), Knight (1989, 1991), Pollard (1991) and Davis, Knight and Liu (1992); see Pollard (1991) for a review of the related literature to that date and Zhao (2000) for a review of some later results. The Pollard, Knight, and Davis et al. approaches are related in basing results on the convexity of the LAD criterion function. Knight (1989, 1991), as well as Phillips (1995), Hercé (1996) and Koenker and Xiao (2004), treat LAD or conditional quantile estimation in non-stationary processes, and Knight (1999) in heteroskedastic processes. The methods of proof in the present study, using generalized functions to represent the LAD/quantile estimator, are most closely related to those of Phillips (1991, 1995).

Each of the contributions just listed deals with estimation in a process with a finite number of parameters. Some results are available for increasing-order models in the literature on M- estimators, which for sufficiently weak restrictions on the objective function includes quantile estimation as a special case. Portnoy (1985), for example, examines M-estimation in regression contexts where the number of parameters grows without bound, but the conditions in this and related papers include differentiability of the objective function, ruling out quantile estimation. Welsh (1988) requires weaker conditions on the objective function which allow quantile estimation, and provides an explicit application of his results to that context, but treats cases of regression models with i.i.d. errors, and so cannot accommodate many time series applications. Davis and Dunsmuir (1997) prove a limit theorem for estimation of finite-order ARMA processes and regression models with finite-order ARMA errors, but do not consider increasing-parameter models.

We therefore have well-established asymptotic results for OLS in truncations of infinite-order processes, and for quantile estimation in finite-order processes or increasing-order cases with conditions that rule out many time series applications. The next section of this paper contributes analogous results for quantile estimates of parameters of truncations of infinite-order processes, applicable in cases with dependence. As an example, the results can be applied where an AR(p) model is used to model a more general process which may be approximated arbitrarily well as  $p \to \infty$ ,  $\frac{p}{T} \to 0$ , (T being the sample size), and is therefore a case which is of substantial practical importance in time series problems; however, they are applicable much more generally.

#### 3. Asymptotic theory for conditional quantiles of an infinite-order regression

We first present the general result; the proof of the theorem, using generalized functions to represent the conditional quantile estimator (including LAD, for the conditional 50th percentile), is in the Appendix.

Consider a pair of discrete stochastic processes  $\{y_t, X_t\}$ , and an increasing

sequence of  $\sigma$ -fields  $\{\Im_t\}$ , where the vector (of possibly infinite dimension)  $X_t$  is measurable w.r.t.  $\Im_t$ . The process  $\{y_t\}$  is related to  $\{X_t\}$  by an approximation of possibly infinite order. Denote by  $\chi_q(y_t) \equiv \chi_q(y_t|\Im_t)$  the  $q^{th}$  quantile of the conditional distribution of  $y_t$ . Define the check function:

$$f_q(x) \equiv \left[ \left( q - \frac{1}{2} \right) + \frac{1}{2} \operatorname{sgn}(x) \right] x; \qquad (2.1)$$

this function, which is the basis of the criterion function for estimation of the parameters of the quantile regression model, reduces to the function |x| where  $q = \frac{1}{2}$ . Like the absolute value function, it is non-differentiable at x = 0. However, following Phillips (1991, 1995), we can treat the function  $f_q(x)$  as a weak limit of a sequence of smooth functions  $f_q^m(x)$  (defined in the Appendix, at A.1), which have the property of being continuously differentiable. For the purposes of the present paper, we need only consider functions which are three times continuously differentiable. We can also define the generalized derivatives for the generalized function  $f_q(x)$ , allowing us to speak of  $f'_q(x)$  and  $f''_q(x)$  as weak limits of  $f_q^{m'}(x)$  and  $f_q^{m''}(x)$  (see the Appendix, A.7). These operations are treated in detail in the Appendix. For a general reference on generalized functions, see Gel'fand and Shilov (1964).

Next define the row vector  $X_t(k) \equiv (X_{0,t}, X_{1,t}, \dots, X_{k,t})$  with  $X_{0,t} = 1$ ; we allow for infinite k and corresponding  $X_t(\infty)$ . For any  $k < \infty$ ,  $X_t(\infty)$  can be partitioned as

 $X_t(\infty) = (X_t(k), X_t(k+1, \infty));$  analogously, partition the column vector  $\gamma(\infty)$  as  $(\gamma(k)', \gamma(k+1, \infty)')'.$ 

The assumption following uses a scaling matrix  $V_T(k)$ , which may be thought of as being such that  $V_T(k)'V_T(k) = \Sigma$ , the covariance matrix of  $X_t(k)$ , where the latter exists.

Assumption 1. For a sequence of (possibly random) non-singular matrices  $\{V_T(k)\}$ ,

- (a) $X_t(k)V_T(k)^{-1}$  is  $\mathfrak{T}_t$ -measurable for all T, k
- (b)  $\chi_q(y_t|\mathfrak{S}_t) = X_t(\infty) \gamma_q(\infty)$

- (c)  $e_t - \varpi_q = y_t - X_t(\infty)\gamma_q(\infty)$ , where  $\varpi_q$  is a constant, is such that

- . (i)  $\{e_t, X_t\}$  is a stationary ergodic sequence
- . (ii) the p.d.f. of e,  $p_e(x)$ , exists and is continuous at  $x = \overline{\omega}_q$
- . (iii)  $\{f'_q(e_t \varpi_q), \mathfrak{F}_t\}$  is a m.d. sequence

- (d)<sup>4</sup>. 
$$\sup_{1 \le t \le T} \max |X_t(k)V_T^{-1}(k)| = o_p(1)$$
  
- (e)  
 $\max \left| \sum_{t=1}^T V_T(k)^{-1}X_t(k)'X_t(k)V_T(k)^{-1} - I_{k+1} \right| = o_p(1)$ 

- (f) There exists a monotonically increasing function  $\omega(x)$  such that  $k = \omega(T) \to \infty$  as  $T \to \infty$  and

$$\sup_{1 \le t \le T} |X_t [k+1, \infty) \gamma_q [k+1, \infty)| = o_p(T^{-\frac{1}{2}}).$$

Parts (a) through (e) of this assumption have antecedents in the previous literature. Parts (a), (d) and (e) are similar to parts (ii), (iii), (iv) respectively of Theorem 2 (LAD with random regressors) of Pollard (1991). Part (b) states linearity of the conditional quantile function, also assumed in typical treatments of LAD asymptotics. Part (c) is analogous to the error assumption of Pollard (1991), and embodies the common requirement that the density of the error exist and be continuous at a particular point (typically 0). Part (f) is particular to the case we treat here: it states that the approximation error induced by a truncation to order k of the infinite linear process can be made suitably small as  $k, T \to \infty$ . Note that condition (f) is satisfied if components of  $X_t$  are bounded in probability and  $\sum_{i=1}^{\infty} |\gamma_{qi}| < \infty$  (as for example for an AR( $\infty$ ) representation of an ARMA(p, q)), or if  $\gamma(k)$  and  $X_{k,t}$  satisfy complementary conditions such as that  $\gamma(k)$  declines exponentially in k while  $X_{k,t}$  grows at most at a polynomial rate as  $k \to \infty$ , for  $k = T^{\alpha}, \ \alpha \in (0, \frac{1}{2})$ .

We denote by  $\hat{\gamma}_q(k)$  the quantile estimator of  $\gamma_q(k)$ :

$$\hat{\gamma}_q(k) = \arg\min_{\gamma} \sum_{t=1}^T f_q \left( y_t - X_t(k)\gamma \right).$$
(2.2)

For any fixed<sup>5</sup> k' < k, define  $\Omega_{k'} \equiv \begin{bmatrix} I_{k'} & 0 \\ 0 & 0 \end{bmatrix}$ , where  $I_{k'}$  is the identity matrix of order k'. We are ready now to formulate the result about the asymptotic distribution of the quantile estimator.

<sup>&</sup>lt;sup>4</sup>For any matrix X, max |X| denotes in this paper the absolute value of the largest component of the matrix.

<sup>&</sup>lt;sup>5</sup>While k' could be defined to grow with k, we concentrate here on a fixed k' to simplify the derivations below.

**Theorem 1.** Under Assumption 1, as  $T \to \infty$ ,  $k = \omega(T)$ ,

$$\Omega_{k'}V_T(k)\left(\hat{\gamma}_q(k) - \gamma_q(k)\right) \Rightarrow N\left(0, \frac{q(1-q)}{p_e(\varpi_q)}\Omega_{k'}\right).$$

**Proof:** See the Appendix.

With additional conditions (sufficient conditions would comprise the existence of the first two moments of X, and a rate condition such as  $k = o(T^{-\frac{1}{2}})$ —see for example Berk 1974) consistent estimates of  $V_T(k)^{-1}$  can be obtained as  $\hat{\Sigma}^{-\frac{1}{2}}$ , where  $\hat{\Sigma}$  is a consistent estimate of the  $k \times k$  covariance matrix of  $X_t(k)$ . The theorem then fully characterizes the asymptotic distribution of the estimated quantiles.

#### 4. Simulation example

In this section we illustrate the application of these results using the example of the approximation of a moving average process by an autoregression, estimated by quantile regression. Consider the  $MA(\ell)$  process with  $\ell = 2$ ,

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}, \tag{3.1}$$

where  $\{\epsilon_t\}$  is a white-noise process and the roots of the polynomial  $z^2 + \theta_1 z + \theta_2 = 0$  are all inside the unit circle, implying an invertible MA. This process (or its analogue for other values of  $\ell$ ) can be represented as an infinite-order autoregression  $y_t = \sum_{i=1}^{\infty} \alpha_i y_{t-i} + \epsilon_t$ , where the coefficients are absolutely summable, so that the approximation can be made arbitrarily good for a finite number of autoregressive terms p, as long as  $p \to \infty$  and  $\frac{p}{T} \to 0$  at an appropriate rate. Properties of this autoregression are well understood in the LS regression case; see Galbraith and Zinde-Walsh (1994) on estimation of an MA via this approximation. The results in section 3 of the present paper establish that we can also estimate the autoregression consistently via quantile regression.

To illustrate the results of doing so, we generate examples of (3.1) by simulation, using relatively heavy tailed  $t_5$  – distributed errors to emulate a case in which LAD (or other quantile) regression might be chosen for its robustness to large errors. (In this case the estimated quantiles are essentially the same except for the values of the intercept, although in other contexts one might wish to allow different models for different quantiles of the distribution.) We use values of the parameter vector  $(\theta_1, \theta_2)$  of (0.8, 0.15), (0.6, 0.3), (0.6, -0.2) and (-0.6, 0.3) respectively, and 10 000 simulated samples. The Figures 1(a–d) and 2(a–d) present results for LAD estimation (q = 0.5) of the first AR coefficient; similar results are obtained for other quantiles (the empirical conditional quantiles differ primarily in the value of the intercept) and the other AR coefficients. For a first sample size of T = 200, the infinite-order AR representation of the process is truncated to an AR(8), and this truncation is estimated by quantile regression. The empirical distribution of the t-type statistic for this first coefficient,  $\hat{\gamma}_q^{(1)}(k) - \gamma_q^{(1)}(k)$  scaled by its standard error,<sup>6</sup> is presented in Figure 1 for each case, together with the normal distribution scaled to the same empirical variance. We see reasonably good conformity with the normal at this sample size and AR order, although some considerable finite-sample truncation bias is observable in Figure 1c, a process for which the coefficients in the AR expansion decay relatively slowly. Next, in Figure 2, we present analogous results, but for T = 1000 and an AR order of 16, so that the truncation bias is smaller. We now observe very good conformity with the asymptotic normal distribution even in the less-well-behaved case c.

### 5. Concluding remarks

Theorem 1 states that consistent and asymptotically normal estimates result from application of the quantile estimator to a finite truncation of an infiniteorder model that represents a true process, where the order of truncation increases with sample size. In time series contexts, the results will be useful for quantile estimation where a process can be represented by, e.g., an infinite-order AR, MA or ARCH process, but where a representation with fixed, finite order is not valid; the simulations reported in Section 4 for one such case suggest that the asymptotics provide a reasonable guide to the finite-sample distributions for moderately large samples and appropriate truncation order. As well, the results can be applied to series expansion for a conditional quantile.

Note that we assume only that a specific quantile is represented as in Assumption 1(b), and that the approximation error satisfies the other conditions formulated in Assumption 1. With respect to time series applications, note also that while we can use this method to estimate parametric models such as the ARMA, the infinite-order representation addressed in this theorem is more general than representations such as the ARMA. For ARMA estimation we would take  $y_t$ in Assumption 1 to be an observable process and  $\{X_t\}$  to be a sequence of lags

<sup>&</sup>lt;sup>6</sup>The coefficients  $\gamma$  of the infinite-order representation can be obtained from the standard recursive expression; see for example Fuller (1976).

of this process and innovations. However, Theorem 1 says nothing specific about the nature of the set of series used for approximation of the original series: it can consist of other non-linear functions of the innovations, or the set can include auxiliary variables. Such variables in the linear representation need only meet the requirements of Assumption 1.

This result therefore complements those listed earlier which have established asymptotic properties of  $L_1$ -norm estimates in finite stationary linear processes, unit root processes, infinite-variance processes, and heteroskedastic processes, and validates  $L_1$ -norm estimation in a class of cases for which the estimated model is used non-parametrically as an approximation to some underlying process whose precise form may be unknown or non-finite.



Figure 1b: Estimated density of t-type statistic v. Normal



## References

Berk, K.N. (1974) Consistent Autoregressive Spectral Estimates. Annals of Statistics 2, 489–502.

Bhansali, R, (1978) Linear prediction by autoregressive model fitting in the time domain. Annals of Statistics 6, 224-231.

Bierens, H.J. (1994) Topics in Advanced Econometrics. Cambridge University Press, Cambridge.

Bloomfield, P. and W.L. Steiger (1983) Least Absolute Deviations: Theory, Applications, and Algorithms. Birkhauser, Boston.

Braun, P.A. and S. Mittnik (1993) Misspecifications in Vector Autoregressions and their Effects on Impulse Responses and Variance Decompositions. *Journal of Econometrics* 59, 319-341.

Chen, X.R., Z.D. Bai, L.C. Zhao and Y.H. Wu (1990) Asymptotic Normality of Minimum  $L_1$  – Norm Estimates in Linear Models. Science in China A, 1311-1328.

Davis, R.A. and W.T.M. Dunsmuir (1997) Least Absolute Deviations Estimation for Regression with ARMA Errors. *Journal of Theoretical Probability* 10, 481-497.

Davis, R.A., K. Knight and J. Liu (1992) M-estimation for autoregressions with infinite variance. *Stochastic Processes and Applications* 40, 145-180.

Fuller, W. A. (1976) Introduction to Statistical Time Series. Wiley, new York.

Galbraith, J.W. and V. Zinde-Walsh (1994) A Simple, Non-iterative Estimator for Moving Average Models. *Biometrika* 81, 143–155.

Galbraith, J.W. and V. Zinde-Walsh (1997) On Some Simple, Autoregressionbased Estimation and Identification Techniques for ARMA Models. *Biometrika* 84, 685–696.

Gel'fand, I.M. and G.E. Shilov (1964) Generalized Functions, Vol.1: Properties and Operations. Academic Press, San Diego.

Gonçalves, S. and L. Kilian (2005) Asymptotic and bootstrap inference for  $AR(\infty)$  processes with conditional heteroskedasticity. Working paper.

Hercé, M. (1996) Asymptotic Theory of LAD Estimation in a Process with Finite Variance Errors. *Econometric Theory* 12, 129-153.

Koenker, R. and G. Bassett (1978) Regression Quantiles. *Econometrica* 56, 33-50.

Koenker, R. and Z. Xiao (2004) Unit Root Quantile Autoregression Inference. Journal of the American Statistical Association 99, 775-787. Knight, K. (1989) Limit Theory for Autoregressive Parameter Estimates in an Infinite-Variance Random Walk. *Canadian Journal of Statistics* 17, 261-278.

Knight, K. (1991) Limit Theory for M Estimators in an Integrated Infinite-Variance Process. *Econometric Theory* 7, 200-212.

Knight, K. (1999) Asymptotics for  $L_1$  Estimators of Regression Parameters under Heteroskedasticity. Canadian Journal of Statistics 27, 497-507.

Lewis, R. and G. Reinsel (1985) Prediction of Multivariate Time Series by Autoregressive Model Fitting. *Journal of Multivariate Analysis* 16, 393-411.

Lütkepohl, H. and D.S. Poskitt (1996) Testing for Causation Using Infinite Order Vector Autoregressive Processes. *Econometric Theory* 12, 61-87.

Lütkepohl, H. and P. Saikkonen (1997) Impulse Response Analysis in Infinite-Order Cointegrated Vector Autoregressive Processes. *Journal of Econometrics* 81, 127-157.

Phillips, P.C.B. (1991) A Shortcut to LAD Estimator Asymptotics. *Econometric* Theory 7, 450-463.

Phillips, P.C.B. (1995) Robust Nonstationary Regression. *Econometric Theory* 11, 912-951.

Pollard, D. (1991) Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory* 7, 186-199.

Portnoy, S. (1985) Asymptotic Behavior of M Estimators of p Regression Parameters When  $p^2/n$  is Large: II. Normal Approximation. Annals of Statistics 13, 1403-1417.

Saikkonen, P. and H. Lütkepohl (1996) Infinite-order Cointegrated Vector Autoregressive Processes. *Econometric Theory* 12, 814-844.

Welsh, A.H. (1989) On M-Processes and M-Estimation. Annals of Statistics 17, 337-361.

Zhao, L. (2000) Some Contributions to M-estimation in Linear Models. *Journal of Statistical Planning and Inference* 88, 189-203.