Insper
IBMEC
São Paulo

A note on the use of quantile regression in beta convergence analysis

Márcio Poletti Laurini

Insper Working Paper
WPE: 102/2007

A NOTE ON THE USE OF QUANTILE REGRESSION IN BETA CONVERGENCE ANALYSIS

MÁRCIO POLETTI LAURINI

Ibmec São Paulo and IMECC-UNICAMP. Address - Rua Quatá 300, 04546-042, São Paulo-SP, Brazil. Email - marciopl@isp.edu.br

Abstract. We discuss how to interpret conflicting results obtained by the use of quantile regression methods in growth regression tests of β-convergence hypothesis and the results obtained by nonparametric methods. We show that the assumption of linearity may cause the non-rejection of the β-convergence hypothesis by quantile regression. We also show that using a nonparametric form of quantile regression, we can reject the hypothesis of β-convergence and confirm the results of divergence and formation of convergence clubs. We illustrate the discussion by using the conflicting results on convergence found in the dataset of per-capita income of Brazilian municipalities between 1970 and 1996.

In the empirical study of economic growth and income convergence, the class of growth regressions is a widely used tool. Introduced by Barro (1991), this regression assumes a general form given by:

$$\gamma_i = \beta \log(y_{i,0}) + \psi X_i + \pi Z_i + \varepsilon_i: \qquad (1)$$

where γ_i is the growth rate of the i-th economy i ; $y_{i,0}$ is the initial income of i-th economy; X_i contains explicative variables related to the growth model of Solow (1956) and Z_i is a set of variables that may affect the convergence process but are not directly related to the model of Solow (1956) .

In this regression, the estimate of a negative β parameter, controlling the effect of the variables X_i and Z_i, is indicative of a negative relationship between the initial income and the growth rate of the economies, known as β-convergence hypothesis. This property, derived from the decreasing returns from factors of production in the growth model of Solow (1956) and Swan (1956) would imply convergence in growth rates, since economies with higher initial incomes would have smaller rates of growth than those with lower initial incomes.

There are some points to criticize about conclusions on β-convergence obtained from estimating a negative parameter β in the equation (1). The first aspect is that there are models of income divergence compatible with negative β, for example Azariadis and Drazen (1990). Besides this compatibility between negative β and the possibility of divergence, there are two other points of criticism related to the use of the growth regression in the test of β-convergence hypothesis.

One basic criticism of growth regression is the possibility of Galton's fallacy, as pointed out by Friedman (1992) and Quah (1993), where a result of negative β may not indicate convergence of growth rates but rather regression toward the mean. The second criticism, as pointed out by Bernard and Durlauf (1996), is that the growth regression assumes an implicit condition of homogeneity - all the economies must have the same rate of convergence represented by the parameter β. Thus the process of formation of convergence clubs (e.g. Quah (1997)), indicative of the existence of a group of convergent economies and another group of divergent economies, cannot be captured by using this regression, given the unique β for all economies in the sample.

The methodology of quantile regression (Koenker and Basset (1978)) was pointed as a possible solution to these two problems. Koenker (2000) argues that the methodology of quantile regression allows surpassing the regression to the mean problem, corresponding to Galton's fallacy. Specifically, as the quantile regression allows heterogeneity in the coefficients of the regression, there is a vector of parameters for each conditional quantile of the dependent variable, in the case of convergence studies the growth rates of the economies. With these two properties, the methodology of quantile regression would permit capturing divergence and formation of convergence clubs. An example of divergence would be a quantile regression where economies with higher growth rates (higher quantiles) have a positive relationship with the initial income and economies with lower growth rates have a negative β, which characterizes divergence as distinct relationships with the initial income.

In a quantile regression, the objective function is directly formulated in terms of the quantile of interest τ, minimizing the objective function:

$$\min_{\beta \in R^p} \sum_{i=1}^{n} \rho_\tau (y_i - x_i \beta(\tau)) \qquad (2)$$

where ρ is a loss function conditional to the quantile τ, with τ □(0,1) . The loss function is defined as $\rho_\tau(u) = u(\tau - I(u < 0))$, where I(.) is an indicator function, and u gives us the difference between the observed value and the estimated value for each observation. The minimization of the loss function given by (2) results in estimators for $\hat{\beta}$ (τ), through the expectations of $\rho_\tau(y_i - x_i\beta(\tau))$ for each $\beta(\tau)$, allowing estimators for the conditional quantiles, including the sample median as specific case. In convergence studies, the objective function is analogous to the functional form of the regression (1) but formulated in terms of conditional quantiles:

$$\min_{\beta,\psi,\pi \in R^p} \sum_{i=1}^{n} \rho_\tau (y_i - \beta(\tau)\log(y_{i,0}) - \psi(\tau)X_i - \pi(\tau)Z_i) \qquad (3)$$

Analyses of convergence using quantile regression can be found in several articles: Mello and Novo (2002) tests convergence using the Barro-Lee and Summers and Heston datasets; Mello and Perrelli (2003) test the β-convergence hypothesis using quantile regression in Barro (1991) equations. Other articles using quantile regression to test versions of the convergence hypothesis are Barreto and Hughes (2004) and Miles (2004). In this article, we use the same database already studied in Andrade et al. (2003), Andrade et al. (2004) and Laurini et al. (2005), an example of conflicting results about the hypothesis of convergence obtained using different methodologies.

In the study of Andrade et al. (2003), quantile growth regressions were estimated using per-capita income for Brazilian municipalities1. In this study, the regression estimates for quantiles (.10,.25,.50,.75,.90) result in the same negative betas by using OLS. According to the authors, quantile regression is more robust evidence in favor of β-convergence hypothesis.

1 The sample contains the per capita income of the Brazilian municipalities in 1970 and 1996, contemplating 3781 observations. For a detailed description of this dataset see Andrade et al. (2004).

However, analyses of the same dataset using nonparametric methodologies in Andrade et al. (2004) and Laurini et al. (2005), point to evidence against the hypothesis of β-convergence and in favor of the hypothesis of formation of convergence clubs. The question is how to interpret these conflicting results, since the use of quantile regression would have to surpass the existing limitations in the OLS estimate of growth regression, allowing the formation of convergence clubs to be identified.

The point of our article is to show that the non-rejection of the β-convergence hypothesis in Andrade et al. (2003), is caused by the linear functional form assumed in the estimate of the growth regression using quantile regression. The assumption of a linear relationship between each quantile of the growth rates and the initial income may cancel all the potential advantages of using quantile regression for convergence analysis if this functional form is incorrect. This problem can also contaminate all studies using linear forms in quantile regression to test convergence.

To demonstrate this proposition, we replace the linear functional form in the quantile regression with a nonparametric form of quantile regression known as quantile smoothing spline (e.g. Bosch et al. (1995); Koenker (2005)). This methodology consists of estimating the following function:

$$\min_{g \in R} \sum_{i=1}^{n} (y_i - g(X_i))^2 + \lambda \int (g''(X_i))^2 \, dx \qquad (4)$$

where g can be any curve; X_i is the explicative variable, and λ is a smoothness parameter for the adjustment, controlling the trade-off between minimization of the residual and the roughness of the adjustment. This problem has a solution2 based on the use of cubic splines to estimate the unknown function g.

This methodology allows us to relax the linearity assumed traditionally in the convergence studies with a nonparametric estimation, where each conditional quantile can have a nonlinear relation with the initial income. It is a generalization of the methodology of smoothing splines used in Andrade et al. (2004), which allows us to join the benefits of the nonparametric estimation with

2 See Hardle (1990) for details

the robustness properties over Galton's fallacy and the heterogeneity of parameters for each conditional quantile derived from the quantile regression methodology.

We estimate the unconditional growth regression (without the inclusion of the control variables) using the described methodology of quantile smoothing spline in equation (4), for the data set of per capita incomes of the Brazilian municipalities studied in Andrade et al. (2003), Andrade et al. (2004) and Laurini et al. (2005). We estimate the nonparametric growth regression for quantiles (.01,.05.10,25,.50,.75,.95,.99) and show the estimated results in Figure(1).

FIGURE 1 HERE

We can observe that by using the methodology of quantile smoothing splines it is possible to visualize the divergence process clearly. The estimated nonparametric curves show an interval of points where the curve has a positive trend, i.e., a positive relationship with the initial log incomes per capita between 7 and 8 (approx. US$ 1100 and US$ 3000), which corresponds to the values of intermediate incomes in the sample. For the initial lower and higher incomes, the general behavior is a negative relation with the initial income.

This behavior of convergent lower and higher incomes and divergent intermediary incomes is consistent with the process of formation of convergence clubs obtained by Andrade et al. (2004) and Laurini et al. (2005) for this data set. These results confirm that the problem of using quantile regression in the study of processes of divergence and formation of clubs is related to the linearity assumption, being basically a problem of incorrect specification of the functional form.

To clarify the dependence process between the growth rate and the initial income, we make nonparametric estimate of the dependence function between the empirical quantiles of growth rates and initial income. To construct Figure (2), we transform the values of the studied variables in terms of their empirical quantiles, and after we estimate the quantile smoothing spline for these transformed variables. The estimated curves directly measure the dependence function between the quantiles of growth rate and the initial income. It should be pointed out that this is a nonparametric method to estimate a Copula function. A Copula is a dependence function that links univariate margins to construct the full multivariate distributions3. Estimating the quantile smoothing spline for each possible quantile, we are able to capture nonparametrically the full dependence process, without the imposition of any functional form.

FIGURE 2 HERE

In this formulation, the nonlinear dependence is still more evident. Figure (2) clearly shows that the divergence process occurs in quantiles between .4 and .6 of the initial income, with a positive relationship with the initial income, which gives support to the hypothesis of convergence clubs.

To verify the superiority of the nonparametric quantile smoothing splines over the parametric linear quantile regression, we use the Generalized Likelihood Ratio test introduced by Fan et al. (2001). This test allows us to compare parametric and nonparametric functional forms, using a generalization of the likelihood ratio principle. The test statistic derived from Fan et al. (2001), and used in our article is in the form:

$$GLR = \frac{T}{2}\frac{SQR - SQIR}{SQR} \tag{5}$$

where SQR is the residual sum of squares of the restricted model (the linear quantile regression) and SQIR is the residual sum of squares of the unrestricted model (the quantile smoothing spline) and T is the sample size. Under regularity conditions it is possible to get the asymptotic distribution of test or then to derive the finite sample test distribution using bootstrap. We get the empirical p-values through the procedure of conditional bootstrap detailed in Fan and Yao (2003), carrying the test for quantiles (.01,.05.10,25,.50,.75,.95,.99). Table (1) show that the null hypothesis of equality between the parametric and nonparametric models is rejected by all the quantiles except the extremal quantiles. This fact indicates that the imposition of a linear

3 See Nelsen (1999) for details

parametric form is rejected in favor of the nonparametric fit, which again gives evidence in favor of the divergence and formation of convergence clubs.

TABLE 1 HERE

The analysis carried out in this article strengthens the basic point analyzed in depth in Durlauf et al. (2005), about the difficulty in analyzing the empirical models studied and their relation with the functional forms proposed by the theoretical models. Durlauf et al. (2005) points out the difference between the linear functional forms derived from the growth models of Solow (1956) and the necessity of nonlinear functional forms in endogenous growth models as in Romer (1986) or Lucas (1988). The overall result is that an incorrect functional form can cancel all the potential robustness properties of econometric methods like the quantile regression method.

References

Andrade, E., Laurini, M. P., Madalozzo, R. and Valls Pereira, P.V. (2004). Convergence clubs among Brazilian municipalities, *Economics Letters* **83(2)**, 179--184.

Andrade, E., Laurini, M. P., Madalozzo, R. and Valls Pereira, P.V (2003). Testing convergence across municipalities in Brazil using quantile regression, *In Proceedings of 18 Meeting of European Economic Association*.

Azariadis, C. and Drazen, A. (1990). Threshold externalities in economic development, *Quarterly Journal of Economics* **105(2),** 501--526.

Barreto, R. A. and Hughes, A. W. (2004). Under performers and over achievers: A quantile regression analysis of growth, *Economic Record* **80**, 17--35.

Barro, R. (1991). Economic growth in a cross-section of countries, *Quarterly Journal of Economics* **106(2),** 407--443.

Bernard, A. and Durlauf, S. (1996). Interpreting tests of the convergence hypothesis, *Journal of Econometrics* **71**, 161--173.

Bosch, R. J. Y, Ye, Y. and Woodworth, G. G. (1995). A Convergent algorithm for the quantile regression with smoothing splines, *Computational Statistics & Data Analysis* **19**, 613--630.

Durlauf, S., Johnson, P. and Temple, J. (2005). Handbook of Economic Growth, chap. Growth econometrics, pp. 5–41. North-Holland.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon, *The Annals of Statistics* **29**, 153--193.

Friedman, M. (1992). Do old fallacies ever die? *Journal of Economic Literature* **30,** 2129--2132.

Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.

Koenker, R. and G. Basset (1978). Regression quantiles, *Econometrica* **46**, 33--50.

Koenker, R. (2000). Galton, Edgeworth, Frisch, and prospects for quantile regression in econometrics, *Journal of Econometrics* **95**, 347--374.

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

Laurini, M. P., Andrade, E. and Valls Pereira, P. V. (2005). Income convergence clubs for Brazilian Municipalities: a non-parametric analysis, *Applied Economics* **37(18),** 2099--2118.

Lucas, R. (1988). On the mechanics of economic development, *Journal of Monetary Economics* **22(1),** 3--42.

Mello, M. and Novo, A. (2002). The new empirics of economic growth: Quantile regression estimation of growth equations. *Unpublished Working Paper*.

Mello, M. and Perrelli, R. (2003). Growth equations: a quantile regression exploration, *The Quarterly Review of Economics and Finance* **43(4)**, 643--667.

Miles, W. (2004). Human capital and economic growth: A quantile regression approach, *Applied Econometrics and International Development* **2**.

Nelsen, R. (1999). *An Introduction to Copulas*. Lectures Notes in Statistics, Springer Verlag.

Quah, D. (1993). Galton's fallacy and tests of convergence hypothesis, *Scandinavian Journal of Economics* **95(4),** 427--443.

Quah, D. (1997). Empirics for growth and distribution: Stratification, polarization and convergence clubs, *Journal of Economic Growth* **2,** 27--59.

Romer, P. (1986). Increasing returns and long run growth, *Journal of Political Economy* **94(5),** 1002--1037.

Solow, R. (1956). A contribution to the theory of economic growth, *Quarterly Journal of Economics* **70**, 65--94.

Swan, T. (1956). Economic growth and capital accumulation, *Economic Record* **32**, 334--361.

Footnotes:

1 The sample contains the per capita income of the Brazilian municipalities in 1970 and 1996, contemplating 3781 observations. For a detailed description of this dataset see Andrade et al. (2004).

2 See Hardle (1990) for details.

3 See Nelsen (1999) for details.

Figures

Figure 1: Non-Parametric Quantile Growth Regression

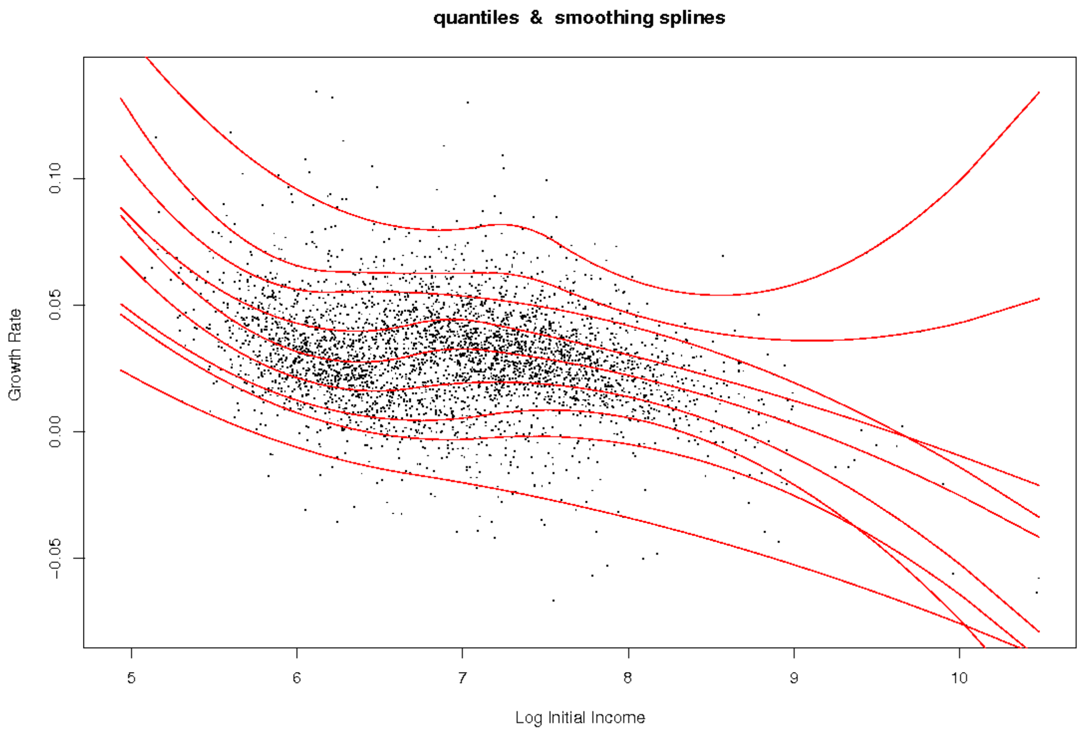
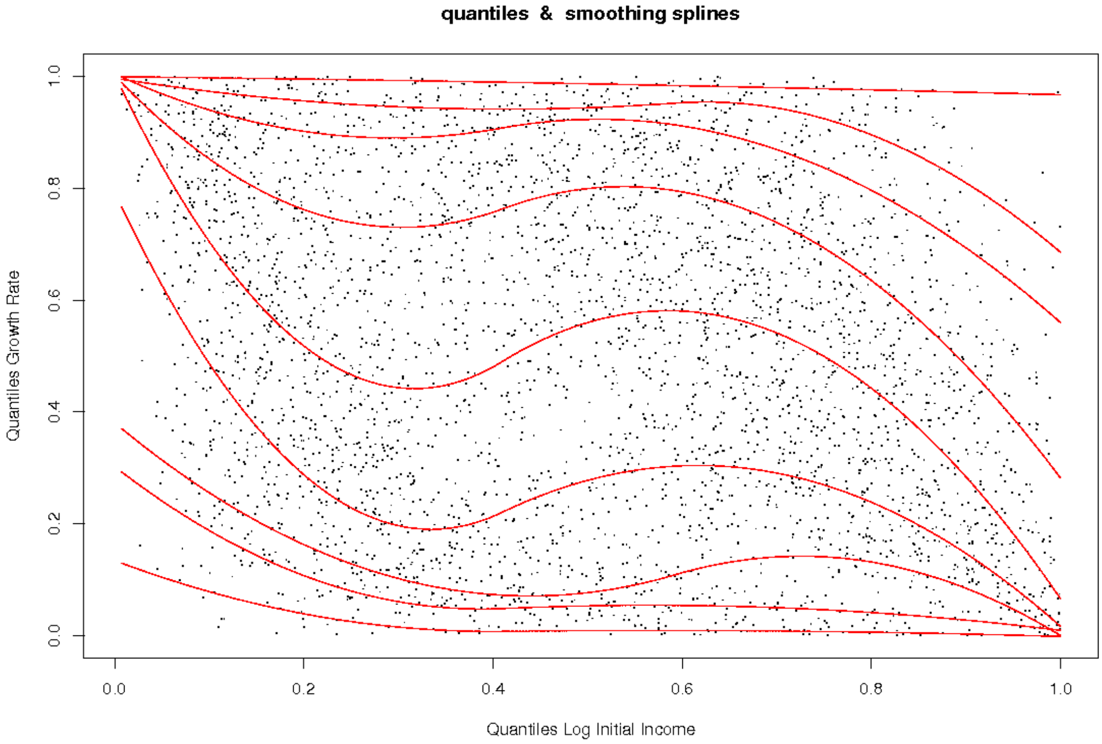


Figure 2: Non-Parametric Quantile Smoothing Spline - Quantile-Quantile Estimation



Tables

Table 1: Generalized Likelihood Ratio Test

Quantile	.01	.05	.10	.25	.50	.75	.90	.95	.99
P-Value	0.178	0.004	0.000	0.000	0.000	0.000	0.000	0.003	0.103