Mathematisches Forschungsinstitut Oberwolfach

Report No. 56/2012

DOI: 10.4171/OWR/2012/56

Mini-Workshop: Frontiers in Quantile Regression

Organised by
Victor Chernozhukov, Boston
Holger Dette, Bochum
Xuming He, Ann Arbor
Roger Koenker, Champaign

25 November – 1 December 2012

Abstract. Quantiles play an essential role in modern statistics, as emphasized by the fundamental work of Parzen (1978) and Tukey (1977). Quantile regression was introduced by Koenker and Bassett (1978) as a complement to least squares estimation (LSE) or maximum likelihood estimation (MLE) and leads to far-reaching extensions of "classical" regression analysis by estimating families of conditional quantile surfaces, which describe the relation between a one-dimensional response y and a high dimensional predictor x. Since its introduction quantile regression has found great attraction in mathematical and applied statistics because of its natural interpretability and robustness, which yields attractive applications in such important areas as medicine, economics, engineering and environmental modeling. Although classical quantile regression theory is very well developed, the implicit definition of quantile regression still yields many new mathematical challenges such as multivariate, censored and longitudinal data, which were discussed during the workshop.

Mathematics Subject Classification (2000): 62G10, 62G08, 62G30.

Introduction by the Organisers

The workshop *Frontiers of quantile regression*, organised by Victor Chernozhukov (Boston), Holger Dette (Bochum), Xuming He (Ann Arbor) and Roger Koenker (Champaign) was held 25 November – 1 December 2012. This meeting was well attended by 16 participants with broad geographic representation from all continents. During the workshop all mathematical aspects of the recent development in quantile regression analysis were discussed. A particular focus was on *Multivariate quantile regression* where several new concepts were presented by the

participants, including Bahadur representations, asymptotic normality and uniform convergence of the corresponding estimates. Other talks discussed quantile regression for longitudinal data and random effect models with applications in functional data analysis and biostatistics and the definition of new spectra of stationary time series via quantile regression methods.
Several speakers presented their results on variable selection in high-dimensional quantile regression models, especially under the framework of "large p small n paradigm (here p refers to the dimension of the parameter to be estimated and n denotes the sample size). It was shown that useful model identification is possible when sparsity of the model is expected to hold. Two other speakers discussed quantile regression methods for censored data. Specifically, the following research fields in quantile regression were discussed during the workshop.

(1) Multivariate quantile regression
(2) Quantile regression for longitudinal data and random effect models
(3) Bayesian analysis in quantile regression
(4) Variable selection in high-dimensional quantile regression models
(5) Quantile regression for censored data
(6) Quantile regression in time series

The workshop stimulated intensive discussions between all participants and new developments in various subfields of quantile regression analysis. For example, the problem of quantile regression for multivariate was discussed in three talks from different perspectives. Similarly, in the context of stationary time series a spectral theory will be developed, which avoids the existence of any moments.

References

[1] Koenker, R. and Bassett, G. Regression quantile. *Econometrica*, 46:33-50, 1978.
[2] Parzen, E. Nonparametric Statistical Data Modeling. *Journal of the American Statistical Association*, 74:105-123, 1978.
[3] Tukey, J. W. *Exploratory Data Analysis*. Addison-Wesley, 1977.

Mini-Workshop: Frontiers in Quantile Regression

Table of Contents

Abstracts

Robust Inference in High-Dimensional Sparse Quantile Regression Models

Alexandre Belloni
(joint work with Victor Chernozhukov, Kengo Kato)

Many applications of interest requires the measurement of the distributional impact of a policy (or treatment) on the relevant outcome variable. To consider the whole quantile impact instead of focusing only on the average effect allows to study the impact on the tails of the distribution which might of main interest. Examples where the tail of the distribution plays the central role of the analysis include infant birth weight, student performance, and risk management just to name a few.

Quantile treatment effects has emerged as an important paradigm to measure such distributional impact particularly when the impact might be heterogeneous. In this work we propose a method to estimate the quantile treatment effect α_0 of a poliy/treatment d of an outcome of interest y

$$\tau - \text{quantile}(y \mid z, d) = d\alpha_0 + g_\tau(z),$$

where the unknown function g_τ summarizes the confounding effects of the observed controls z. To approximate g_τ we rely on linear combinations of p-dimensional covariates $x = P(z)$. Given a sample with n independent observations, we allow for $p \gg n$ to rely on a flexible family of functions in order to achieve an accurate approximation for g_τ in finite samples. In turn, the high-dimensionality brings forth the need to perform model selection or regularization.

We are particularly interested in non-parametric settings in which it seems unrealistic to assume a separation from zero of the relevant coefficients. Therefore, in those settings, an important feature of any inferential procedure is its robustness with respect to model selection mistakes. In particular the inferential analysis presented here does not rely on the oracle property. It turns out that this feature is fundamental to achieve a inferential procedure whose properties hold uniformly over a large class of data generating processes similarly the average treatment effect case studied in [3]. This allows us to overcome the impact of model selection mistakes avoiding the criticisms in [6] which ties the oracle property with super-efficient estimators [7].

Due to inherent non-linearity of quantile estimators the proposed method proceed in three steps. Each step relies on a different identification condition which in turn requires a different estimation procedure. The first step aims to construct an accurate estimate of the control function g_τ via ℓ_1-penalized quantile regression (ℓ_1-qr) estimator [2, 5, 9]. The second step attempts to properly partial out the confounding factors z from the treatment estimating a suitable residual via heteroskedastic Lasso [8, 1]. Finally, the residual is used as the instruments for the treatment in a Instrumental quantile regression [4]. Interestingly, although we

provide a complete analysis which relies on these specific methods, several variants are of the proposed method are possible. For instance, Lasso can be substituted by Dantzig selector, SCAD, square-root Lasso, the associated post-model selection estimators or others. In our analysis we provide high-level conditions which summarizes the required properties of the estimators.

Despite of possible model selection mistakes, under suitable regularities conditions we show that the proposed estimator $\check{\alpha}$ of α_0 obeys

$$V_0^{-1/2}\sqrt{n}(\check{\alpha} - \alpha_0) \rightsquigarrow N(0,1) \tag{1}$$

where $V_0 := \tau(1-\tau)J_0^{-1}$ and $J_0 = E[v_i v_i']$. Moreover, we also propose the construction of a confidence region $\widehat{A}_{n,\xi}$ with asymptotic coverage $1-\xi$ based on an inverse statistics. Importantly, the robustness with respect to model selection mistakes allows these results to hold uniformly over a large range of dgps.

In the process of establishing the main results we also contribute to the literature of high-dimensional estimation. An intermediary step of the method required the estimation of a weighted least squares version of Lasso in which weights are estimated. Finite sample bounds of Lasso for the prediction rate are established to this new case. Finite sample bounds for the prediction norm on the estimation error of ℓ_1-penalized quantile regression in nonparametric models extending results on [2, 5, 9]. We further developed results on instrumental quantile regression problems in which we allow for the dimension to increase and estimated instruments.

References

[1] A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. forthcoming Econometrica, 2012.

[2] A. Belloni and V. Chernozhukov. ℓ_1-penalized quantile regression for high dimensional sparse models. *Annals of Statistics*, 39(1):82–130, 2011.

[3] Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection amongst high-dimensional controls. *ArXiv*, 2011.

[4] Victor Chernozhukov and Christian Hansen. Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142:379–398, 2008.

[5] K. Kato. Group lasso for high dimensional sparse quantile regression models. Preprint, ArXiv, 2011.

[6] Hannes Leeb and Benedikt M. Pötscher. Model selection and inference: facts and fiction. *Economic Theory*, 21:21–59, 2005.

[7] Hannes Leeb and Benedikt M. Pötscher. Sparse estimators and the oracle property, or the return of Hodges' estimator. *J. Econometrics*, 142(1):201–211, 2008.

[8] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58:267–288, 1996.

[9] Lie Wang. L_1 penalized lad estimator for high dimensional ilnear regression. *ArXiv*, 2012.

Random-Effects Quantile Regression

Stéphane Bonhomme
(joint work with Manuel Arellano)

Nonlinear panel data is an active area of research. Despite some recent progress, it is fair to say that we are still short of answers for panel versions of many models commonly used in applied work. In this paper we focus on one particular nonlinear model for panel data: quantile regression.

Quantile regression provides a flexible modelling of conditional distributions. This is potentially useful in panel data applications, where conditioning on covariates and initial conditions is key. Current approaches, since Koenker (2004), are based on a quantile-by-quantile fixed-effects approach. In contrast, we propose a random-effects approach, which relies on a flexible modelling of the conditional distribution of individual effects using an additional quantile regression specification.

The method. Specifically, our aim is to estimate models of the form:

$$Q_\tau\left(Y_{it} \mid X_{it}, \eta_i\right) = X_{it}'\beta\left(\tau\right) + \eta_i\gamma\left(\tau\right), \quad \tau \in (0,1),$$

where the quantile function of outcomes is linear in regressors and individual effects. With data on Y, X and η we would just run an ordinary quantile regression, i.e. minimize the check function:

$$\min_{\beta,\gamma} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_\tau\left(Y_{it} - X_{it}'\beta - \gamma\eta_i\right), \tag{1}$$

where $\rho_\tau(u) = u\left(\tau - \mathbf{1}\{u \leq 0\}\right)$.

However, since η_i is unobserved we construct instead some imputations, say M imputed values $\eta_i^{(1)}, ..., \eta_i^{(M)}$ for each individual in the panel. Then we compute:

$$\min_{\beta,\gamma} \sum_{i=1}^{N}\sum_{m=1}^{M}\sum_{t=1}^{T} \rho_\tau\left(Y_{it} - X_{it}'\beta - \gamma\eta_i^{(m)}\right).$$

For this approach to be valid, the imputed values have to be drawn from the posterior distribution of η_i conditioned on the data: $f\left(\eta \mid Y_i, X_i\right)$. The trouble is that this density depends on the distributions of $Y_i \mid X_i, \eta_i$ and $\eta_i \mid X_i$, hence on the parameters to be estimated. We solve this using an EM-type iterative procedure, which alternates between QR estimation and generation of imputed values based on an approximation to the posterior density of η_i.

To complete the model, we assume a correlated random-effects structure for the individual effects, based on a second layer of quantile regression. Formally, we specify:

$$Q_\tau\left(\eta_i \mid X_i\right) = X_i'\delta\left(\tau\right), \quad \tau \in (0,1),$$

where X_i is a function of regressors in all periods.

In each step of the iterative algorithm, the posterior density of η_i is reconstructed using the parameters $\delta(\cdot)$, $\beta(\cdot)$, and $\gamma(\cdot)$. In order to keep the dimension

of the problem manageable we use a spline approximation strategy first proposed by Wei and Carroll (2009), which takes advantage of smoothness restrictions across percentile values $\tau \in (0, 1)$.

Contribution and outline. In the approach we propose, the random-effects η_i are interpreted as missing data, assumed not to vary over time. Consistency of the estimator is then established for fixed T, as N tends to infinity. This last feature contrasts with most existing work on quantile regression for panel data, where consistency holds as both N and T tend to infinity. In addition, we establish conditions for nonparametric identification. For this purpose, we draw a connection with nonlinear measurement error models and the work by Hu and Schennach (2008).

Importantly, we show that our approach may easily be extended in various directions. One extension is to allow for multiple random-effects, for example random slopes, by building triangular quantile regression models. A second extension is to allow for lagged outcomes as covariates, and more generally for predetermined regressors. To handle predeterminedness we rely on a third layer of quantile regression, and use another quantile regression to model the feedback process.

A practical feature of our approach is that the computational simplicity of standard quantile regression is preserved. In our iterative algorithm, each estimation step relies on a simple quantile regression, which may be solved using linear programming. To implement the algorithm, we use a modified EM algorithm with a Monte Carlo E-step. Preliminary simulations suggest good finite sample properties and numerical stability.

Lastly, we apply our method to estimate the effect of maternal smoking during pregnancy on children's birthweight. We use a sample of mothers and children from Abrevaya (2006), and allow for mother-specific correlated random-effects to address potential endogeneity of smoking. The panel data quantile regression estimates show that the negative smoking effect is reduced when allowing for mother-specific unobserved heterogeneity. Nonetheless, the average effect remains negative, and is particularly large at the bottom of the distribution.

References

[1] J. Abrevaya, *Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach,* Journal of Applied Econometrics **21** (2006), 489–519.

[2] Y. Hu and S.M. Schennach, *Instrumental Variable Treatment of Nonclassical Measurement Error Models*, Econometrica **76** (2008), 195–216.

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

[4] Y. Wei and R.J. Carroll, *Quantile Regression with Measurement Error*, J. Amer. Statist. Assoc. **104** 2009, 1129–1143.

Semiparametric Efficient Tests

Juan Carlos Escanciano

This paper proposes efficient tests for restrictions on finite-dimensional parameters in regular semiparametric models. Our theory overcomes the main limitation of the existing theory, which requires explicit computation and estimation of certain projections onto infinite-dimensional tangent spaces and a case-by-case analysis. We consider generic semiparametric models defined by an infinite number of moment conditions, including a finite-dimensional parameter of interest and possibly containing moment-specific nonparametric nuisance parameters. We investigate tests based on functionals of the sample analog of the moments, and show that the optimal functional takes the form of a Radon-Nikodym derivative or nonparametric Likelihood Ratio (LR). This functional LR test was first suggested by Grenander (1950) in an abstract setting, and since then, numerous proposals have used this principle. However, the asymptotic optimality of the nonparametric LR test remained unknown.

Our first result shows that the functional LR test is efficient in our general semiparametric setting. To prove this result we first obtain a generic asymptotic representation of the LR test statistic as a score-type process (i.e. as a sample mean of a score function). We characterize the score function in terms of the limiting covariance and mean functions of the standardized sample moments. Then, we show that the resulting score coincides with the so-called efficient score in the semiparametric model defined by the moment restrictions, which in turn establishes the semiparametric efficiency of the functional LR test.

The LR test is generally infeasible, as it assumes knowledge of the singular value decomposition of the limiting Gaussian process of the standardized sample moments. We then propose and justify feasible efficient tests based on a novel nonparametric estimator of the so-called efficient score, without requiring direct computation of projections onto tangent spaces or sample splitting techniques. The estimator of the efficient score solves an integral equation, a generalized information equality, using Tikhonov regularization and has a simple closed form expression. The test based on the estimated efficient score is a natural extension of the celebrated $C(\alpha)$ test of Neyman (1959) from parametric to semiparametric models. Thus, the proposed efficient tests are widely applicable, while being straightforward to implement.

Finally, to illustrate the benefits of the approach, we apply the new methods to a semiparametric linear quantile regression model with a continuum of quantiles. The parameter of interest is a quantile-invariant parameter. Optimal inferences in this model were not available because classical efficiency arguments are difficult to apply. In particular, the orthogonal projection into the tangent space of nuisance parameter has no simple closed form expression. In contrast, our methods deliver relatively simple efficient tests in this example.

References

[1] Grenander, U., *Stochastic Processes and Statistical Inference*, Arkiv for Matematik **1** (1950), 195–277.

[2] Neyman, J., *Optimal Asymptotic Tests of Composite Statistical Hypotheses*, In Probability and Statistics, (ed. U. Grenander), Harald Cramer Volume, Wiley, New York, 212–234.

Conditional Quantile Processes based on Series or Many Regressors

Iván Fernández-Val

(joint work with Alexandre Belloni, Victor Chernozhukov)

Quantile regression (QR) is a principal regression method for analyzing the impact of covariates on outcomes, particularly when the impact might be heterogeneous. This impact is characterized by the conditional quantile function and its functionals [2, 5, 15]. For example, we can model the log of the individual demand for some good, Y, as a function of the price of the good, the income of the individual, and other observed individual characteristics X and an unobserved preference U for consuming the good, as

$$Y = Q(X, U),$$

where the function Q is strictly increasing in the unobservable U. With the normalization that $U \sim \text{Uniform}(0, 1)$ and the assumption that U and X are independent, the function $Q(X, u)$ is the u-th conditional quantile of Y given X, i.e. $Q(X, u) = Q_{Y|X}(u|X)$. This function can be used for policy analysis. For example, we can determine how changes in taxes for the good could impact demand heterogeneously across individuals.

In this paper we develop the nonparametric QR series framework for performing inference on the entire conditional quantile function and its linear functionals. In this framework, we approximate the entire conditional quantile function $Q_{Y|X}(u|x)$ by a linear combination of series terms, $Z(x)'\beta(u)$. The vector $Z(x)$ includes transformations of x that have good approximation properties such as powers, trigonometrics, local polynomials, or B-splines. The function $u \mapsto \beta(u)$ contains quantile-specific coefficients that can be estimated from the data using the QR estimator of Koenker and Bassett [16]. As the number of series terms grows, the approximation error $Q_{Y|X}(u|x) - Z(x)'\beta(u)$ decreases, approaching zero in the limit. By controlling the growth of the number of terms, we can obtain consistent estimators and perform inference on the entire conditional quantile function and its linear functionals. The QR series framework also covers as a special case the so called many regressors model, which is motivated by many new types of data that emerge in the new information age, such as scanner and online shopping data.

We describe now the main results in more detail. Let $\hat{\beta}(\cdot)$ denote the QR estimator of $\beta(\cdot)$. The first set of results provides large-sample theory for the empirical QR coefficient process of increasing dimension $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$. We obtain uniform strong approximations to this process by a sequence of the following stochastic processes of increasing dimension:

(i) a conditionally pivotal process,
(ii) a gradient bootstrap process,
(iii) a Gaussian process, and
(iv) a weighted bootstrap process.

To the best of our knowledge, all of the above results are new. The existence of the pivotal approximation emerges from the special nature of QR, where a (sub) gradient of the sample objective function evaluated at the truth is pivotal conditional on the regressors. This allows us to perform high-quality inference without even resorting to Gaussian approximations. We also show that the gradient bootstrap, introduced by Parzen, Wei and Ying [20] in the parametric context, is effectively a means of carrying out the conditionally pivotal approximation without explicitly estimating Jacobian matrices. The conditions for validity of these two schemes require only a mild restriction on the growth of the number of series terms in relation to the sample size. We also obtain a Gaussian approximation to the entire distribution of QR process of increasing dimension by using chaining arguments and Yurinskii's coupling. Moreover, we show that the weighted bootstrap works to approximate the distribution of QR process for the same reason as the Gaussian approximation. The conditions for validity of the Gaussian and weighted bootstrap approximations, however, appear to be substantively stronger than for the pivotal and gradient bootstrap approximations.

The second set of results provides estimation and inference methods for linear functionals of the conditional quantile function, including

(i) the conditional quantile function itself, $(u, x) \mapsto Q_{Y|X}(u|x)$,
(ii) the partial derivative function, $(u, x) \mapsto \partial_{x_k} Q_{Y|X}(u|x)$,
(iii) the average partial derivative function, $u \mapsto \int \partial_{x_k} Q_{Y|X}(u|x) d\mu(x)$, and
(iv) the conditional average partial derivative, $(u, x_k) \mapsto \int \partial_{x_k} Q_{Y|X}(u|x) d\mu(x|x_k)$,

where μ is a given measure and x_k is the k-th component of x. Specifically, we derive uniform rates of convergence, large sample distributions and inference methods based on the strong pivotal and Gaussian approximations and on the gradient and weighted bootstraps. It is noteworthy that all of the above results apply to function-valued parameters, holding uniformly in both the quantile index and the covariate value, and covering pointwise normality and rate results as a special case. If the function of interest is monotone, we show how to use monotonization procedures to improve estimation and inference.

The paper contributes and builds on the existing important literature on conditional quantile estimation. First and foremost, we build on the work of He and Shao [13] that studied the many regressors model and gave pointwise limit theorems for the QR estimator in the case of a single quantile index. We go beyond the many regressors model to the series model and develop large sample estimation and inference results for the entire QR process. We also develop analogous estimation and inference results for the conditional quantile function and its linear functionals, such as derivatives, average derivatives, conditional average derivatives, and others. None of these results were available in the previous work. We also build on Lee [18] that studied QR estimation of partially linear models in

the series framework for a single quantile index, and on Horowitz and Lee [14] that studied nonparametric QR estimation of additive quantile models for a single quantile index in a series framework. Our framework covers these partially linear models and additive models as important special cases, and allows us to perform inference on a considerably richer set of functionals, uniformly across covariate values and a continuum of quantile indices. Other very important work includes Chaudhuri [7], Chaudhuri, Doksum and Samarov [8], Härdle, Ritov, and Song [12], Cattaneo, Crump, and Jansson [6], and Kong, Linton, and Xia [17], among others, but this work focused on local, non-series, methods.

Our work also relies on the series literature, at least in a motivational and conceptual sense. In particular, we rely on the work of Stone [21], Andrews [1], Newey [19], Chen and Shen [10], Chen [9] and others that rigorously motivated the series framework as an approximation scheme and gave pointwise normality results for least squares estimators, and on Chen [9] and van de Geer [22] that gave (non-uniform) consistency and rate results for general series estimators, including quantile regression for the case of a single quantile index. White [23] established non-uniform consistency of nonparametric estimators of the conditional quantile function based on a nonlinear series approximation using artificial neural networks. In contrast to the previous results, our rate results are uniform in covariate values and quantile indices, and cover both the quantile function and its functionals. Moreover, we not only provide estimation rate results, but also derive a full set of results on feasible inference based on the entire quantile regression process.

While relying on previous work for motivation, our results require to develop both new proof techniques and new approaches to inference. In particular, our proof techniques rely on new maximal inequalities for function classes with growing moments and uniform entropy. One of our inference approaches involves an approximation to the entire conditional quantile process by a conditionally pivotal process, which is not Donsker in general, but can be used for high-quality inference. The utility of this new technique is particularly apparent in our high-dimensional setting. Under stronger conditions, we also establish an asymptotically valid approximation to the quantile regression process by Gaussian processes using Yurinskii's coupling. Previously, [11] used Yurinskii's coupling to obtain a strong approximation to the least squares series estimator. The use of this technique in our context is new and much more involved, because we approximate an entire empirical QR process of an increasing dimension, instead of a vector of increasing dimension, by a Gaussian process. Finally, it is noteworthy that our uniform inference results on functionals, where uniformity is over covariate values, do not even have analogs in the least squares series literature (the extension of our results to least squares is a subject of ongoing research, [3]).

This material is based upon work supported by the National Science Foundation under Grant No. 1060889. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

References

[1] Donald W. K. Andrews. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica*, 59(2):307–345, 1991.

[2] O. Arias, K. F. Hallock, and Walter Sosa-Escudero. Individual heteroneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics*, 26(1):7–40, 2001.

[3] A. Belloni, X. Chen, and V. Chernozhukov. New asymptotic theory for series estimators. 2009.

[4] A. Belloni and V. Chernozhukov. ℓ_1-penalized quantile regression for high dimensional sparse models. *Ann. Statist.*, 39(1):82–130, 2011.

[5] M. Buchinsky. Changes in the u.s. wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62(2):405–458, Mar. 1994.

[6] Matias D. Cattaneo, Richard K. Crump, and Michael Jansson. Robust data-driven inference for density-weighted average derivatives. *J. Amer. Statist. Assoc.*, 105(491):1070–1083, 2010. With supplementary material available online.

[7] Probal Chaudhuri. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.*, 19(2):760–777, 1991.

[8] Probal Chaudhuri, Kjell Doksum, and Alexander Samarov. On average derivative quantile regression. *Ann. Statist.*, 25(2):715–744, 1997.

[9] Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. *In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics*, 6B(Chapter 76), 2006.

[10] Xiaohong Chen and Xiaotong Shen. Sieve extremum estimates for weakly dependent data. *Econometrica*, 66(2):289–314, 1998.

[11] V. Chernozhukov, Sokbae Lee, and Adam M. Rosen. Intersection bounds: estimation and inference. *arXiv*, (0907.3503v1), 2009.

[12] Wolfgang K. Härdle, Yaacov Ritov, and Song Song. Partial linear quantile regression and bootstrap confidence bands. SFB 649 Discussion Paper 2010-002, 2009.

[13] X. He and Q.-M. Shao. On parameters of increasing dimenions. *Journal of Multivariate Analysis*, 73:120–135, 2000.

[14] J. L. Horowitz and S. Lee. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, 100(472):1238–1249, 2005.

[15] R. Koenker. *Quantile regression*. Cambridge University Press, New York, 2005.

[16] R. Koenker and G. Basset. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[17] Efang Kong, Oliver Linton, and Yingcun Xia. Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, 26(5):1529–1564, 2010.

[18] S. Lee. Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, 19:1–31, 2003.

[19] Whitney K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168, 1997.

[20] M. I. Parzen, L. J. Wei, and Z. Ying. A resampling method based on pivotal estimating functions. *Biometrika*, 81(2):341–350, 1994.

[21] Charles J. Stone. Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053, 1982.

[22] Sara van de Geer. M-estimation using penalties or sieves. *Journal of Statistical Planning and Inference*, 108(1-2):55–69, 2002.

[23] Halbert L. White. Nonparametric estimation of conditional quantiles using neural networks. In *Proceedings of the Symposium on the Interface*, pages 190–199, 1992.

Toward Multivariate Quantile Regression

Alfred Galichon
(joint work with Guillaume Carlier and Victor Chernozhukov)

The aim of this talk is to investigate a multivariate notion of quantile regression. Let us setup some notations. Consider $(X,Y) \sim \nu$, where $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^d$, where d is some positive integer. It is assumed that $X_1 = 1$ is constant. Denote $\bar{x} = \mathbb{E}[X]$. Let $\mu = \mathcal{U}\left([0,1]^d\right)$ the uniform density on $[0,1]^d$. We shall use the notation $\dot{f}(t)$ for $df(t)/dt$.

Conditional quantiles. For now we focus on the unidimensional case, and we seek a variational principle for conditional quantiles which we will elaborate upon. When $d = 1$, μ is the uniform distribution over $[0,1]$, and $(X,Y) \sim \nu$ is a random vector where $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}$. A natural idea is to look for the $U \sim \mu$ which is maximally correlated to Y among those independent from X. Write

$$\begin{aligned} & \max \mathbb{E}[UY] \\ s.t. \quad & U \sim \mu,\ (X,Y) \sim \nu \\ & (X,U) \text{ indep.} \end{aligned} \tag{1}$$

Theorem 1. *The following holds:*

(i) An optimal coupling (U,X,Y) solution to problem (1) exists.

(ii) The value of problem (1) coincides with the value of its dual, namely

$$\begin{aligned} & \min_{\varphi,\psi} \int \varphi(x,u)\, d\mu d\nu + \int \psi(x,y)\, d\nu \\ s.t. \quad & \psi(x,y) \geq uy - \varphi(x,u). \end{aligned}$$

(iii) A solution (φ,ψ) to the dual problem exists and is such that

$$\psi(x,y) = \sup_u \{uy - \varphi(x,u)\} \quad \text{and} \quad \varphi(x,u) = \sup_y \{uy - \psi(x,y)\},$$

in particular $\varphi(x,u)$ is convex with respect to u for all x.

(iv) The primal solution (U,X,Y) and the dual solution are related by

$$Y = \frac{\partial \varphi}{\partial u}(X,U)$$

and $u \to \frac{\partial \varphi}{\partial u}(x,u)$ exists almost everywhere and is nondecreasing.

(v) As a consequence,

$$\frac{\partial \varphi}{\partial u}(x,u) = Q_{Y|X=x}(u)$$

is the conditional quantile of $Y|X = x$.

Quantile regression. We see that problem (1) leads to the conditional quantile representation $u \to Q_{Y|X=x}(u)$ without shape restriction regarding the dependence in x. If we want to impose shape restrictions (like linearity w.r.t. x) on

the dual variable φ, this means loosening the constraints in the primal problem. As one looks for partial linearity of φ w.r.t. x, it is natural to relax the constraint of independence of U and X into mean independence of X from U, namely $\mathbb{E}[X|U] = \bar{x}$. Accordingly, write

$$\begin{aligned} & \max \mathbb{E}[UY] \\ s.t. \quad & U \sim \mu,\ (X,Y) \sim \nu \\ & \mathbb{E}[X|U] = \bar{x} \end{aligned} \tag{2}$$

The next result relates this problem to the classical Quantile Regression procedure. Call $\beta^{QR}(t)$ the classical Quantile Regression estimator, obtained from

$$\beta^{QR}(t) = \text{argmin}_{\beta \in \mathbb{R}^n} \left\{ \mathbb{E}[(Y - x.\beta)^+] + (1-t)\bar{x}.\beta \right\}.$$

Theorem 2. *The following holds:*

(i) An optimal coupling (U, X, Y) solution to Problem (2) exists.

(ii) The value of Problem (2) coincides with the value of its dual, namely

$$\begin{aligned} & \inf_{b \in L^1(\mu), \psi \in L^1(\nu)} \bar{x} \int b d\mu + \int \psi d\nu \\ s.t. \quad & \psi(x,y) \geq uy - x.b(u) \end{aligned} \tag{3}$$

(iii) A solution (b, ψ) to the dual problem exists in L^1, an is such that

$$\psi(x,y) = \sup_{u \in [0,1]} \{uy - x.b(u)\} \tag{4}$$

(iv) Whenever the quantile regression estimator $\beta^{QR}(t)$ has no crossing, then the solution to the primal problem (U, X, Y) and to the dual problem (ψ, b) are related by the first order conditions in (4), that is

$$Y = X.\dot{b}(U)$$

(where $\dot{b}$ is the derivative of b), and

$$\dot{b}(u) = \beta^{QR}(u) \quad \textit{and} \quad \psi(x,y) = \sup_u \left\{ \int_0^u y - x.\beta^{QR}(t)\, dt \right\}.$$

Multivariate quantiles. We now would like to extend Quantile Regression to the multivariate case. We do this by appealing to the notion of multivariate quantiles introduced in [4], which we now recall. Let μ be the uniform distribution over the unit hypercube $[0,1]^d$. Let $Y \sim Q$ be a random vector of $\mathbb{R}^d$. By the Monge-Kantorovich duality

$$\sup_{\substack{U \sim \mu \\ Y \sim Q}} \mathbb{E}_\pi[Y.U] = \inf_{\varphi(u) + \psi(y) \geq \langle u, y \rangle} \int \varphi(u)\, d\mu(u) + \int \psi(y)\, dQ(y)$$

where $u.y$ denotes the standard scalar product in $\mathbb{R}^d$, and the equality of the primal and the dual, as well as the existence of solutions to both problems is a basic result in Optimal transport theory (see [7] for an excellent introduction). Further, the solutions of the dual (φ, ψ) can be taken convex, and are hence almost everywhere

differentiable. Finally, the primal solutions (U, Y) and dual solutions (φ, ψ) are related by the complementary slackness relation

$$Y = \nabla\varphi\,(U)$$

where $\nabla\varphi = (\partial\varphi/\partial u_1, ..., \partial\varphi/\partial u_d)$ is the gradient of φ. Following [4], [5] and [6], we take the map $u \to \nabla V(u)$ as the definition of our multivariate quantile, and call $\varphi(u)$ the *Kantorovich potential* of Q. Note that in dimension 1, this definition is consistent with the usual one: the gradient of a convex function is a nondecreasing function–which then coincides with the quantile function. The link with the notion of statistical depth and fundamental symmetries of the problem are investigated in a work in progress [6]. The link with the Knothe-Rosenblatt quantile is made precise in [1].

Multivariate empirical quantile process. Let Q be a probability measure on $\mathbb{R}^d$ with a density $f(y)$ and Kantorovich potential $\varphi_Q(u)$. Let Q_N be the empirical distribution of an i.i.d. sample $\{y_1, ..., y_N\}$ from Q, that is $Q_N = \frac{1}{N}\sum_{k=1}^{N}\delta_{y_k}$, and let $\varphi_{Q_N}(u)$ be the Kantorovich potential of Q_N. As explained in [4], φ_{Q_N} is piecewise affine and can be expressed $\varphi_{Q_N}(u) = \max_{k=1...N}\{u.y_k - \alpha_k\}$ where the α_k's are determined by

$$\min_{\alpha\in\mathbb{R}^N}\left\{\mathbb{E}\left[\max_{k=1...N}\{U.y_k - \alpha_k\}\right] + \frac{1}{N}\sum_{k=1}^{N}\alpha_k\right\}.$$

In ongoing work [2], the *multivariate empirical quantile process* is defined by

$$\nabla V_N(u) = \sqrt{N}\left(\nabla\varphi_{Q_N}(u) - \nabla\varphi_Q(u)\right)$$

where $\nabla\varphi_{Q_N}$ and $\nabla\varphi_Q$ are the multivariate quantiles of Q_N and Q respectively. Letting $\mathcal{A}_Q$ be the second order elliptic differential operator defined by

$$\left(\mathcal{A}_Q V\right)(u) = f\left(\nabla\varphi_Q(u)\right) Tr\left(D^2V(u).D^2\varphi_Q(u)^{-1}\right) + \nabla V(u).\nabla f\left(\nabla\varphi_Q(u)\right),$$

it is shown that at the first order, V_N is the solution to the following generalized Laplace equation with Neumann-type boundary conditions

$$\mathcal{A}_Q V_N + \sqrt{N}\left(Q_N - Q\right) = 0.$$

Multivariate quantile regression. The ideas exposed in the last two paragraphs combine very naturally with the first two, to lead to a convenient notion of Multivariate Quantile Regression. Theorem 2 essentially holds without modification. In ongoing work [3], we propose to use the multivariate version of program (2) to compute the optimal coupling $(U, X, Y) \sim \pi$, and to look for β such that distance between Y and $X.\beta(U)$ is minimized, which thus provides a multivariate analog of quantile regression.

References

[1] Carlier, G., A. Galichon and L. Santambrogio (2010). "From Knothe's transport to Brenier's map and a continuation method for optimal transport," *SIAM Journal of Mathematical Analysis* 41, pp. 2554-2576.

[2] Chernozhukov, V., and Galichon, A. (2013). "The multivariate empirical quantile process". In preparation.

[3] Carlier, G., Chernozhukov, V., and Galichon, A. (2013). "Multivariate quantile regression via maximal correlation". In preparation.

[4] Ekeland, I., A. Galichon and M. Henry (2012). "Comonotone measures of multivariate risks," *Mathematical Finance* 22 (1), pp. 109-132.

[5] Galichon, A. and M. Henry, "Dual theory of choice with multivariate risks," *Journal of Economic Theory* 147 (4), pp. 1501–1516.

[6] Galichon, A., Hallin, M., and Henry, M. (2013). "Monge-Kantorovich Quantiles and Multivariate Statistical Depth". In preparation.

[7] Villani, C., *Topics in Optimal Transportation*, Providence: American Mathematical Society, 2003.

Of Distribution Functions, Halfspace Depth, Multivariate Quantiles, and Multiple-Output Quantile Regression

Marc Hallin

(joint work with Zudi Lu, Davy Paindaveine and Miroslav Šiman)

1. Quantiles and Depth

1.1. **Quantiles and quantile contours.** The classical concept of quantile is intrinsically univariate, and strongly rooted into the well-ordered nature of the real line; based on "analytical" characterizations, its empirical counterparts are obtained through numerically "efficient" L_1 algorithms and, via Bahadur representation results, yield standard central-limit asymptotics.

Let $X \sim P$ be a real-valued, absolutely continuous random variable with strictly increasing distribution function $F_X = F$. The quantile function of X is the inverse distribution function $q \in (0,1) \mapsto F^{-1}(q)$. It is characterized by a a collection of nested regions of the form $(-\infty, F^{-1}(q)]$, indexed by their probability contents $q \in (0,1)$. Those regions are not equivariant under affine transformations: for instance, $x \mapsto -x$ yields $-F_X^{-1}(q) = F_{-X}^{-1}(1-q) \neq F_{-X}^{-1}(q)$. The same quantile function also can be characterized by the collection of *nested* regions $[F_+^{-1}(\tau), F_-^{-1}(\tau)]$ indexed by their probability contents $1-2\tau$, $\tau \in (0, 1/2]$, where $F_+(x) := P[X \leq x] = F(x)$ and $F_+(x) := P[X \geq x]$. These regions, contrary to the previous ones, are equivariant under affine (location-scale-symmetry) transformations (actually, equivariance holds under the much larger group of continuous monotone transformations).

Note that F_+^{-1} and F_-^{-1} can be interpreted as the quantile functions associated with the directions $u = \pm 1 \in \mathcal{S}_0$ (the unit sphere in dimension one), respectively, yielding *directional quantiles*, thus, but also *projected quantiles*, computed from the projections $\pm X$ of X on the directions $u = \pm 1$, respectively. The intervals $[F_+^{-1}(\tau), F_-^{-1}(\tau)]$ can be considered as *quantile regions* (of order τ, with

probability contents $1-2\tau$), and their boundaries $\{F_+^{-1}(\tau), F_-^{-1}(\tau)\}$ as *quantile contours* (of order τ, with probability contents $1-2\tau$). Contrary to the quantiles themselves, quantile regions and contours are coordinate-free concepts, invariant under affine (location-scale-symmetry) transformations.

Traditional quantiles also can be defined as the minimizers of weighted L_1 criteria. Let $\rho_\tau(z) := \tau|z|I[z \geq 0] + (1-\tau)|z|I[z \leq 0]$ denote the so-called *check function*. More precisely,

$$F_\pm^{-1}(\tau) = \mathrm{argmin}_{a\in\mathbb{R}} \mathrm{E}\big[\rho_\tau\big(u(X-a)\big)\big], \qquad u \in \mathcal{S}_0 = \{-1,1\},\ \tau \in (0,1/2]$$

They can be represented, asymptotically, as sums of independent summands; such representations are called *Bahadur representations*, and yield "classical" central-limit behavior. Finally, they define a *probability-integral transformation* which, in terms of quantile contours, can be described as follows. Each $x \in \mathbb{R}$ belongs to one and only one τ-contour: denote by $\tau(x)$ (which possibly takes value 0) the order of that contour, by $p(\tau)$ its probability contents (here, $p(\tau) = 1-2\tau$). Then, $p(\tau(X))$ is uniformly distributed over $[0,1]$, that is, for all $u \in [0,1]$,

$$\mathrm{P}[p(\tau(X))] \leq u = \mathrm{P}[X \in \text{quantile region with probability contents } u] = u.$$

1.2. **Depth and depth contours.** Contrary to quantile functions, depth functions have a multivariate (or multiple output) origin, and were introduced as an attempt to overcome the absence of natural order in $\mathbb{R}^k$. Depth concepts typically are based on "geometric" characterizations, their empirical versions require computationally intensive combinatorial algorithms, and their asymptotics are non-standard.

Let P be a probability measure on $\mathbb{R}^k$. It will be convenient, throughout, to assume that P has a nonvanishing Lebesgue-density. The *halfspace depth* $HD(\mathbf{y}, \mathrm{P})$ of $\mathbf{y} \in \mathbb{R}^k$ with respect to P is $\inf_{\mathbf{u}\in\mathcal{S}_{k-1}} \mathrm{P}[H_{\mathbf{y},\mathbf{u}}]$, where $\mathcal{S}_{k-1} := \{\mathbf{u} : \|\mathbf{u}\| = 1\}$ is the unit sphere in $\mathbb{R}^k$ and $H_{\mathbf{y},\mathbf{u}}$ denotes the halfspace lying above the hyperplane orthogonal to $\mathbf{u}$ and running through $\mathbf{y}$: the depth of $\mathbf{y}$ is thus the minimal probability lying above such hyperplanes—clearly, a measure of *outlyingness*. The empirical version of that concept is the depth of $\mathbf{y}$ with respect to the empirical measure P_n associated with an observed n-tuple $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. The collections of points with given depth τ are called *depth contours*. Population depth contours are continuous, convex, nested, and entirely characterize P.

The *halfspace depth region* $D(\tau)$ of order $\tau \in [0,1]$ associated with P collects all points of $\mathbb{R}^k$ with depth at least τ, that is,

$$D(\tau) = D_{\mathrm{P}}(\tau) := \{\mathbf{y} \in \mathbb{R}^k : HD(\mathbf{y}, \mathrm{P}) \geq \tau\}.$$

Clearly, $D(0) = \mathbb{R}^k$. Also, it can be shown that, for any $\tau > 0$,

$$D(\tau) = \bigcap \{H : H \text{ is a closed halfspace with } \mathrm{P}[\mathbf{Y} \in H] > 1-\tau\}.$$

The empirical version $D^{(n)}(\tau)$ of $D(\tau)$, as usual, is obtained by replacing P with the empirical measure associated with an observed n-tuple $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$.

1.3. **(Directional) quantile regression.** Apparently, quantiles and depth belong to distinct statistical planets. In the univariate case, however, the depth of a point $y \in \mathbb{R}$ with respect to a probability distribution P with strictly monotone distribution function F is either $F(y)$ (y below the median) or $(1 - F(y))$ (y above the median), that is (the unique hyperplane running through y is $\{y\}$ itself), $\min(F(y),\ 1 - F(y)) = \min(F_+(y),\ F_-(y))$. The only points with depth τ (the τ-*depth contour*), $0 < d < 1/2$, are $F^{-1}(\tau)$ and $F^{-1}(1-\tau)$, that is, $F_{\pm}^{-1}(\tau)$, the *directional quantiles* of order τ. This establishes, in the one-dimensional case, a relation between depth contours and quantile contours (one quantile associated with each direction $u = \pm 1$ of $\mathcal{S}_0$). It also suggests extending to higher dimensions this relation between depth and quantiles as a relation between depth contours and the contours of a directional version of quantiles.

Recall the argmin form of the definition of quantiles: the τ-quantile ($\tau \in (0,1)$) of a univariate random variable Y is any element of the collection $\operatorname{argmin}_{a\in\mathbb{R}} \Psi_\tau(a)$, with $\Psi_\tau(a) := \mathrm{E}[\rho_\tau(Y-a)]$, where $x \mapsto \rho_\tau(x) =: \tau|x|I[x > 0] + (1-\tau)|x|I[x \le 0]$. This actually defines a hyperplane dividing the total probability mass into two parts, τ and $1-\tau$, while minimizing a weighted *directional* (here, direction is $+1$) L_1 criterion. This is exactly what the Koenker and Bassett (1978) quantile regression hyperplanes also do, in higher dimension, for a direction which is the "vertical" direction (in a coordinate system where the covariates are "horizontal").

In the case of a single, k-variate random variable $\mathbf{Y}$ (no covariates), there is no reason for privileging any "vertical" direction. The following concept is developed in Hallin et al. (2010). Fix $\mathbf{u} \in \mathcal{S}_{k-1}$, and use it as the "vertical direction" for a (fully k-variate) L_1 regression quantile hyperplane construction in the Koenker and Bassett style. That is, denoting by $\mathbf{\Gamma_u}$ an arbitrary $k \times (k-1)$ matrix of unit vectors such that $(\mathbf{u}, \mathbf{\Gamma_u})$ constitutes an orthonormal basis of $\mathbb{R}^k$, decompose $\mathbf{Y}$ into $\mathbf{Y_u} := (\mathbf{u}'\mathbf{Y})\mathbf{u}$ and $\mathbf{Y_u^\perp} := \mathbf{\Gamma}_\mathbf{u}'\mathbf{Y}$. Define the $\boldsymbol{\tau}$-*quantile hyperplane* $\Pi_{\mathrm{HP\check{S}},\tau\mathbf{u}}$ (with $\boldsymbol{\tau} := \tau\mathbf{u}$) of $\mathbf{Y}$ as any element of the collection of $(k-1)$-dimensional hyperplanes

$$\Pi_{\boldsymbol{\tau}} := \{\mathbf{y} \in \mathbb{R}^k : \mathbf{u}'\mathbf{y} = \mathbf{b}_{\boldsymbol{\tau}}'\mathbf{\Gamma}_\mathbf{u}'\mathbf{y} + a_{\boldsymbol{\tau}}\}$$

such that $(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}')' \in \operatorname{argmin}_{(a,\mathbf{b}')'\in\mathbb{R}^k} \mathrm{E}[\rho_\tau(\mathbf{Y_u} - \mathbf{b}'\mathbf{Y_u^\perp} - a)]$. Those quantile hyperplanes characterize *lower* and *upper quantile halfspaces*,

$$H_{\boldsymbol{\tau}}^- = H_{\boldsymbol{\tau}}^-(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) := \left\{\mathbf{y} \in \mathbb{R}^k : \mathbf{u}'\mathbf{y} < \mathbf{b}_{\boldsymbol{\tau}}'\mathbf{\Gamma}_\mathbf{u}'\mathbf{y} + a_{\boldsymbol{\tau}}\right\}$$

and

$$H_{\boldsymbol{\tau}}^+ = H_{\boldsymbol{\tau}}^+(a_{\boldsymbol{\tau}}, \mathbf{b}_{\boldsymbol{\tau}}) := \left\{\mathbf{y} \in \mathbb{R}^k : \mathbf{u}'\mathbf{y} \ge \mathbf{b}_{\boldsymbol{\tau}}'\mathbf{\Gamma}_\mathbf{u}'\mathbf{y} + a_{\boldsymbol{\tau}}\right\}$$

and, varying $\mathbf{u}$ for fixed τ, *quantile contours* as the boundaries $\partial R(\tau)$ of the *quantile regions* $R(\tau) := \bigcap_{\mathbf{u}\in\mathcal{S}^{k-1}} \bigcap \{H_{\tau\mathbf{u}}^+\}$.

Empirical versions $\Pi_{\mathrm{HP\check{S}},\tau\mathbf{u}}^{(n)}$ and $R^{(n)}(\tau)$, as usual, are obtained by replacing P with the empirical measure associated with an observed n-tuple $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, yielding polygonal (polyhedral) contours with a finite number of edges; those contours can be computed via linear programming.

1.4. **(Directional) projection quantiles.** Another directional approach to multivariate has been proposed by Kong and Mizera (2008). Still for $\mathbf{u} \in \mathcal{S}_{k-1}$, instead of constructing regression hyperplanes in the Koenker-Bassett style, they propose projecting (orthogonally) $\mathbf{Y}$ on the directed straight line with unit vector $\mathbf{u}$, then compute the ordinary τ-quantile of the projection. Varying $\mathbf{u}$ for fixed τ, they also obtain contours, the *envelopes* of which, as we shall see, enjoy interesting properties. This is equivalent to replacing the quantile regression hyperplanes with hyperplanes $\Pi^{(n)}_{\mathrm{KM};\tau\mathbf{u}}$ orthogonal to $\mathbf{u}$ and running through the projection quantile, then considering the envelope, as $\mathbf{u}$ ranges over $\mathcal{S}_{k-1}$, of the (infinite number of) hyperplanes thus obtained.

2. From Directional Quantile Regression and Projection Quantiles to Halfspace Depth

2.1. **Two fundamental results.** Hallin et al. (2010) establish the following fundamental result: *the directional quantile regression contours (regions) coincide with the halfspace depth contours (regions).* More precisely,

Theorem *(i) (population case) For all $\tau \in [0,1)$, $R(\tau) = D(\tau)$, and (ii) (empirical case) assuming that the $n(\geq k+1)$ data points are in general position, for any $\ell \in \{1,2,\ldots,n-k\}$ such that $D^{(n)}(\frac{\ell}{n})$ has a non-empty interior, $R^{(n)}(\tau) = D^{(n)}(\frac{\ell}{n})$ for all positive τ in $[\frac{\ell-1}{n}, \frac{\ell}{n})$.*

Kong and Mizera (2008) similarly prove the equally fundamental result that *the envelopes of their projection quantile contours coincide with the halfspace depth contours.*

2.2. **The best of two worlds?** Directional quantile regression and directional projection quantiles thus both establish a very direct and constructive bridge between quantile and depth contours. Projection quantile contours are conceptually simpler; but it takes an infinite number of $\Pi^{(n)}_{\mathrm{KM};\tau\mathbf{u}}$ hyperplanes, even in the empirical case, to construct a contour, while only a finite number of $\Pi_{\mathrm{HP\check{S}},\tau\mathbf{u}}$ is required.

From their relation to halfspace depth, quantile contours inherit the nice geometric features of depth: convexity, connectedness, nestedness, affine-equivariance, ... essentially, all the properties that have been listed for traditional univariate quantile contours. Via the same relation, quantile contours are bringing to depth the nice analytic features of quantiles: tractable asymptotics (Bahadur representation and central-limit asymptotics), L_1 characterizations and optimality, implementable linear programming algorithms, along with the simplex byproducts (duality and Lagrange multipliers) and, when contours are indexed by their probability contents, a multivariate form of the classical probability integral transformation.

3. Multiple Output Quantile Regression

Similar ideas apply in a regression context, where an m-dimensional response $\mathbf{Y}$ is observed along with a p-dimensional vector of covariates $\mathbf{X} := (1, \mathbf{W})$. The

relevant quantile hyperplanes, depth regions and contours then are those of the m-dimensional distributions of $\mathbf{Y}$ conditional on $\mathbf{W}$ or, more precisely, the collection, for $\mathbf{w}_0$ ranging over $\mathbb{R}^{p-1}$, of the HPŠ hyperplanes, regions and contours associated with the distributions $P^{\mathbf{Y}|\mathbf{W}=\mathbf{w}_0}$ of $\mathbf{Y}$ conditional on $\mathbf{W} = \mathbf{w}_0$. These contours, as functions of $\mathbf{w}_0$, constitute conditional quantile or depth *tubes* that completely characterize the impact of the covariates on the conditional distributions of the response. Consistent *local bilinear* estimators of those contours are provided in Hallin et al. (2012).

References

[1] M. Hallin, Z. Lu, D. Paindaveine, and M. Šiman, *Local bilinear multiple-output quantile regression*, ECARES Working Papers 2012-003, ULB – Université libre de Bruxelles (2012).

[2] M. Hallin, D. Paindaveine, and M. Šiman, *Multivariate quantiles and multiple-output regression quantiles : from L1 optimization to halfspace depth*, Annals of Statististics **38** (2010), 635–669.

[3] R. Koenker and G.J. Bassett, *Regression quantiles*, Econometrica **46** (1978), 33–50.

[4] L. Kong and I. Mizera, *Quantile tomography: using quantiles with multivariate data*, Preprint, arXiv0805.0056v1 (2008).

Composite Quantile Regression with high dimensional Single-Index Models

Wolfgang Karl Härdle

(joint work with Weining Wang, Lixing Zhu, Yan Fan, Lining Yu)

Regression between Y and covariates X is a standard element of statistical data analysis. We consider an efficient model setup that combines a solution to both dimension reduction problems and variable selection problems. Specifically, a composite regression with general weighted loss for the single index model with possibly ultra high dimensional variables. Our setup is very general, and typically include the interesting case of quantile regression and expectile regression. We have derived the theoretical property of our estimation and demonstrate our method with applications to firm risk analysis.

Quantile regression is one of the major statistical tools and is "gradually developing into a comprehensive strategy for completing the regression prediction" Koenker and Hallock (2001). In many fields of applications like quantitative finance, econometrics, marketing and also in medical and biological sciences, quantile regression (QR) is a fundamental element for modeling and inference. An application in financial time series analysis is the estimation of conditional Value-at-Risk (VaR). Engle and Manganelli (2004) proposed the CaViaR framework to model VaR dynamically. Koenker and Bassett (1982) used their QR techniques to test heteroscedasticity in the field of labor market discrimination. The set of conditional quantile curves are key elements in various statistical problems and are therefore of great interest in practice. As a set of curves, they describe the conditional behavior of a response variable given the explanatory variable, and

display changes in all parts of the distribution, other than just the mean. In addition, quantile regression is robust w.r.t. outliers, a feature that is welcome in many practical applications.

The QR estimation implicitly assumes an asymmetric ALD likelihood, and may not be efficient in the QMLE case. Therefore, different type of flexible loss functions are considered in literature to improve the estimation efficiency, such as, composite quantile regression, Zou and Yuan (2008), Kai and Zou (2010) and Kai and Zou (2011). Moreover, Fan and Wang (2011) proposed a more general loss function framework for linear model, with a weighted sum of different kind of loss function, and the weight is selected to be data driven. Another special type of loss considered in Newey and Powell (1987) corresponds to expectile regression (ER) that is in spirit similar to QR but contains mean regression as its special case, while nonparametric expectile smoothing work could be found in Schnabel and Eilers (2009). The ER curves are alternatives to the QR curves and give us a full picture of regression of Y on X.

The difficulty of characterizing an entire distribution partly arises from the high dimensionality of covariates, which asks for striking a balance between model flexibility and statistical precision. To crack this tough nut, dimension reduction techniques of semiparametric type such as single index models came into the focus of statistical modeling. Tracy Z. Wu (2010) considers a quantile regression problem in single index model. However, to our knowledge there are no further literature on a more generalized regression framework for single-index model.

In addition to the dimension reduction, there is however the problem of choosing right variables for projection. This motivates our second goal of this research, variable selection via Lasso. Kong and Xia (1994), Wang and Yin (2008) and Zeng and Zhu (2012) focused on variable selection in mean regression for single index model. Considering the uncertainty on the multi-index model structure, we restrict ourselves to the single-index model at the moment. An application of our research is presented in the relevant financial risk area: to investigate how the revenue distribution of companies depend on financial ratios describing risk factors for possible failure. Such kind of research has important consequences for rating and credit scoring.

When the dimension of X (the explanatory variables) is high, severe nonlinear dependencies between X and the mean (quantile) curves are expected. This triggers the nonparametric approach, but in its full gear, it runs into the "curse of dimensionality" trap, meaning that the convergence rate of the smoothing techniques is so slow that it is actually impractical to use in such situations. A balanced dimension reduction space for quantile regression is therefore needed. The MAVE technique, Xia, Tong, Li and Zhu (2002) provides us 1) with a dimension reduction and 2) good numerical properties for semiparametric function estimation. The set of ideas presented there, however, have never been applied to composite quantile framework or a even more general composite quasi-likelihood framework. The semiparametric multi-index approach that we consider here will provide practitioners with a tool that combines flexibility in modeling with applicability for

even very high dimensional data. Consequently the curse of dimensionality is circumvented. The Lasso idea in combination with the minimum average contrast estimate (MACE) technique will provide a set of relevant practical techniques for a wide range of disciplines.

References

[1] B. Kai, R. Li and H. Zou, *Local CQR smoothing: An efficient and safe alternative to local polynomial regression*, J. R. Statist. Soc. B **72** (2010), 49–69.

[2] B. Kai, R. Li and H. Zou, *New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models*, Annals of Statistics **39** (2011), 305–332.

[3] C. J. Hull, *Risk Management and Financial Institutions – 2nd ed.*, Prentice Hall (2010).

[4] C. Leng, Y. Xia and J. Xu, *An adaptive estimation method for semiparametric models and dimension reduction*, In preparation (2010).

[5] D. Ruppert, S. J. Sheather and M. P. Wand, *An effective bandwidth selector for local least squares regression*, J. Amer. Statist. Assoc. **90** (1995), 1257–1270.

[6] E. Kong and Y. Xia, *Variable selection for the single-index model*, Biometrika **94** (1994), 217–229.

[7] H. Zou and M. Yuan, *Composite quantile regression and the oracle model selection theory*, Annals of Statistics **36** (2008), 1108–1126.

[8] J. Bradic, J. Fan and W. Wang, *Penalized composite quasi-likelihood for ultrahigh dimensional variable selection*, J. R. Statist. Soc. B **73**(3) (2011), 325–349.

[9] J. Fan and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Statist. Assoc. **96** (2001), 1348–1360.

[10] J. Franke, W. Härdle and C. Hafner, *Statistics of Financial Markets*, Springer (2004).

[11] J. Y. Campbell, A. W. Lo and A. C. MacKinlay, *The Econometrics of Financial Markets*, Princeton University Press (1996).

[12] K.-C. Li, *Sliced Inverse Regression for Dimension Reduction (with discussion)*, J. Amer. Statist. Assoc. **86** (1991), 316–342.

[13] K. Yu and M. C. Jones, *Local linear quantile regression*, J. Amer. Statist. Assoc. **98** (1998), 228–237.

[14] L. Hao and Q. D. Naiman, *Quantile Regression*, Sage Publications (2007).

[15] M. Hristache, A. Juditsky, J. Polzehl and V. Spokoiny, *Structure Adaptive Approach for Dimension Reduction*, Annals of Statistics 2**9** (2001), 1537–1566.

[16] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, J. R. Statist. Soc. B **68**(1) (2006), 49–67.

[17] N. Hautsch, J. Schaumburg and M. Schienle, *Financial Network System Risk Contributions*, Discussion Paper 2011–072, CRC 649, Humboldt-Universität zu Berlin (2011).

[18] P. J. Bickel, Y. Ritov and A. B. Tsybakov, *Hierarchical selection of variables in sparse high-dimensional regression*, Borrowing Strength: Theory Powering Applications - A Festschrift for Lawrence D. Brown. IMS Collections **6** (2010), 56–69.

[19] P. J. Bickel, Y. Ritov and A. B. Tsybakov, *Simultaneous analysis of lasso and dantzig selector*, Ann. Statist. **37**(4) (2009), 1705–1732.

[20] P. J. Huber, *Robust estimation of a location parameter*, The Annals of Mathematical Statistics **35**(1) (1964), 73–101.

[21] P. Zeng, T. He and Y. Zhu, *A lasso-type approach for estimation and variable selection in single-index models*, Journal of Computational and Graphical Statistics **21** (2012), 92–109.

[22] P. Zhao and B. Yu, *On model selection consistency of Lasso*, J. Machine Learning Res., **7** (2006), 2541–2567.

[23] Q. Wang and X. Yin, *A nonlinear multi-dimensional variable selection methods for high-dimensional data: sparse MAVE*, Computational Statistics and Data Analysis **52** (2008), 4512–4520.

[24] R. D. Cook and S. Weisberg, *Sliced Inverse Regression for Dimension Reduction: Comment*, J. Amer. Statist. Assoc. **86** (1991), 328–332.
[25] R. F. Engle and S. Manganelli, *CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles*, Journal of Business and Economic Statistics **22** (2004), 367–381.
[26] R. J. Serfling, *Approximation theorems of mathematical statistics*, Wiley, New York (1980).
[27] R. Koenker, *Quantile Regression*, Cambridge Univ. Press (2000).
[28] R. Koenker and G. W. Bassett, *Robust tests for heteroscedasticity based on regression quantiles*, Econometrica **50** (1982), 43–61.
[29] R. Koenker and K. F. Hallock, *Quantile regression*, Journal of Econometric Perspectives **15**(4) (2001), 143–156.
[30] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R.Statist. Soc. B **58** (1) (1996), 267–288.
[31] S. K. Chao, W. K. Härdle and W. Wang, *Quantile regression in Risk Calibration*, Discussion Paper 2012–006, CRC 649, Humboldt-Universität zu Berlin (2012).
[32] S. Schnabel and P. Eilers, *Optimal expectile smoothing*, Computational Statistics and Data Analysis **53**(12) (2009), 4168–4177.
[33] T. Adrian and M. K. Brunnermeier, *CoVaR*, Staff Reports 348, Federal Reserve Bank of New York (2011).
[34] T. Z. Wu, K. Yu and Y. Yu, *Single-index quantile regression*, Journal of Multivariate Analysis **101** (2010), 1607–1621.
[35] W. Hädle and T. M.Stoker, *Investigating Smooth Multiple Regression by the Method of Average Derivatives*, J. Amer. Statist. Assoc. **84** (1989), 986–995.
[36] W. Newey and J. Powell, *Asymmetric least squares estimation and testing*, Econometrica **55** (1987), 819–847.
[37] Y. Kim, H. Choi and H. Oh,*Smoothly clipped absolute deviation on high dimensions*, J. Amer. Statist. Assoc. **103** (2008), 1656–1673.
[38] Y. Li and J. Zhu, *L1- norm quantile regression*, Journal of Computational and Graphical Statistics **17** (2008), 163–185.
[39] Y. Xia, *Asymptotic Distributions for Two Estimators of the Single-Index Model*, Econometric Theory **22** (2006), 1112–1137.
[40] Y. Xia and H. Tong, *On the Efficiency of Estimation for a Single-Index Model in Frontiers in Statistics*, Imperial College Press, London (2006), 63–85.
[41] Y. Xia, H. Tong, W. Li and L. Zhu, *An adaptive estimation of dimension reduction space*, J. R. Statist. Soc. B **64** (2002), 363–410.
[42] Zou, H. *The adaptive Lasso and its oracle properties*, J. Amer. Statist. Assoc. **101**(476) (2006), 1418–1429.

Some Open Questions about Quantile Regression

Roger Koenker

I like to think of quantile regression as one of those fancy salumi slicers that you see in good italian restaurants. Kernel regression taught us how to do local regression by slicing the design space into pieces, and quantile regression sought to do local regression in the y-domain. It has allowed us to see local slices of the conditional distribution of $y|x$. In effect we get a deconstruction of the classical global models. Robust statistics taught us to be suspicious about global likelihoods, especially Gaussian ones, so it semed useful to suspend belief in global likelihoods and slice them up into more digestible pieces. Some open questions about expanding the scope of this approach were proposed in the talk:

- Model Selection and Post Selection Inference

- Survival Analysis and Censored Data
- Multivariate Conditional Quantiles
- Time Series Methods
- Binary Response Models
- Likelihood Interpretations for QR Models

Most of the attention focused on the final topic that raised questions about semi-parametric forms of the likelihood based on the quantile regression paradigm.

Quantile regression in longitudinal analysis

Chenlei Leng
(joint work with Cheng Yong Tang and Weiping Zhang)

Longitudinal data are characterized by dependence among the observations from the same subject. However, how to account for this dependence is difficult in quantile regression, due to the lack of a likelihood.

We study two approaches that deal with this difficulty. For the first approach (Tang and Leng, 2011), we borrow strength from the mean regression to enhance the estimation efficiency, by formulating multiple sets of smooth working estimating equations. The empirical likelihood method is utilized to produce subject-specific weights that are fed into the quantile regression. We show that efficiency is improved as long as the correlation between the estimating equations between the quantile regression and those of the mean regression is nonzero. For the second approach (Leng and Zhang, 2012), we aggregate multiple sets of unbiased non-smooth estimating equations. To overcome the discreteness of these equations, we develop an induced smoothing algorithm for parameter estimation and statistical inference. Methodological and computational aspects of both approaches will be discussed.

References

[1] C. Y. Tang and C. Leng, *Empirical likelihood and quantile regression in longitudinal data analysis*, Biometrika **98** (2011), 1001–1006.
[2] C. Leng and W. Zhang, *Smoothing aggregated estimating equations in quantile regression for longitudinal data*, Statistics and Computing (2012), to appear.

The Jackknife's Edge: Inference on Censored Regression Quantiles

Stephen Portnoy

For (right) censored data, it is very common for the right tail of the survival function to be non-identifiable because of the abundance of censored observations in the right tail. This is especially prominent in censored regression quantile analysis, and introduces a serious problem with inference, especially near the point of non-identifiability. The lack of readily estimable formulas for asymptotic variances requires the use of resampling methods. Unfortunately, the bootstrap (in any of

its versions) generates samples for which the point of non-identifiability has sufficient variability over the resamples so that (in practice) an appreciable number of the resamples can no longer estimate a quantile that the original data could estimate. This leads to very poor coverage probabilities. Thus, resampling methods that provide less variability over the resamples may be very useful in reducing the number of such samples.

The jackknife is one such method, though even for single sample quantiles it is necessary to use a "delete-d" jackknife with d of order strictly larger than $\sqrt{n}$. Another alternative is to use randomly reweighted "bootstrap" samples with weights of the form $(1+v)$, with v larger than $1/\sqrt{n}$. It appears that these approaches can be justified under appropriate asymptotic regimes. The argument for the usual regression quantiles is relatively straightforward, and the basic outline is presented in this talk. Some brief comments will suggest that it should be possible to extend these results to the censored regression quantiles of Portnoy (2003) using the inductive argument in Portnoy and Lin (2010).

After giving some introductory material on censored regression quantiles and outlining some of the arguments mentioned above, I will present a small scale simulation experiment, showing that randomly sampling a relatively modest number of delete-d jackknifed samples with $d = 2\sqrt{n}$ provides quite excellent coverage probabilities for all τ-values for which the quantile is estimable. A canonical real-data example will be presented suggesting that the delete-d jackknife method works at least as well as previous methods. Some comments about future directions will be offered.

References

[1] S. Portnoy, *Censored Regression Quantiles*, J. Amer. Statist. Assoc **98** (2003), 1001–1012.
[2] S. Portnoy and G. Lin, *Asymptotics for Censored Regression Quantiles*, J. Nonparametric Statistics **22**, (2010). 115–130.

Quantile based spectral analysis

Stanislav Volgushev
(joint work with Holger Dette, Marc Hallin, Tobias Kley)

A classical quantity that is used in both time series analysis and signal detection is the so-called periodogram. For real-valued observations $X_1, X_2, ..., X_n$ it is defined by

$$I_n(\omega) := \frac{1}{n}\Big|\sum_{t=1}^{n} X_t e^{-\mathrm{i}t g_n(\omega)}\Big|$$

where $g_n(\omega)$ denotes the Fourier frequency $\frac{2\pi j}{n} \in (-\pi, \pi]$ closest to ω. An alternative representation of the periodogram can be obtained by considering L^2-projections of the data $X_1, X_2, ..., X_n$ onto the harmonic basis. More precisely, for

$\omega_j = \frac{2\pi j}{n}$

$$I_n(\omega_j) = \frac{n}{4}\hat{b}_n(\omega_j)' \begin{pmatrix} 1 & -\mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} \hat{b}_n(\omega_j)$$

where $\hat{b}_n(\omega_j) = (\hat{b}_{1n}(\omega_j), \hat{b}_{2n}(\omega_j))'$ with

$$(\hat{a}_n(\omega_j), \hat{b}_{1n}(\omega_j), \hat{b}_{2n}(\omega_j)) = \arg\min_{b\in\mathbb{R}^3} \sum_{t=1}^{n} \left(X_t - c_t(\omega_j)'b\right)^2$$

and $c_t(\omega_j) = (1, \cos(t\omega_j), \sin(t\omega_j))'$. Because of its intrinsic connection to L^2-methods, the classical periodogram inherits all the drawbacks of this approach, including inherent non-robustness and a very narrow view of the world that is essentially based on means and covariances. In a non-Gaussian setting, such an approach clearly has many disadvantages. In the regression setting, quantile regression (see [4]) is known to provide an alternative to mean regression. It has nice robustness properties and at the same time provides a broader view on conditional distributions. This leads to the natural question of whether this approach can be successfully applied for the analysis of dynamic time series features. In this talk, we proposed to consider a collection of so-called *Laplace-periodograms* that are based on weighted L^1 rather than L^2 projections. More precisely, define

$$\begin{aligned} (\tilde{a}^\tau_n(\omega), \tilde{b}^\tau_{1n}(\omega), \tilde{b}^\tau_{2n}(\omega)) &:= \arg\min_{b\in\mathbb{R}^3} \sum_{t=1}^{n} \rho_\tau\left(X_t - c_t(\omega)'b\right), \\ (\tilde{a}^\tau_{n,R}(\omega), \tilde{b}^\tau_{1n,R}(\omega), \tilde{b}^\tau_{2n,R}(\omega)) &:= \arg\min_{b\in\mathbb{R}^3} \sum_{t=1}^{n} \rho_\tau\left(\hat{F}_n(X_t) - c_t(\omega)'b\right), \end{aligned}$$

where $\rho_\tau(u) := u(\tau - I_{(-\infty,0]}(u))$ is the *check function* and $\hat{F}_n$ is the ecdf of $X_1, ..., X_n$. We then proposed to consider the *Laplace-periodogram kernel* [see also [2, 3] for the special case $\tau_1 = \tau_2$]

$$L_n^{\tau_1,\tau_2}(\omega) := \frac{n}{4}\tilde{b}_n^{\tau_1}(\omega)' \begin{pmatrix} 1 & -\mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} \tilde{b}_n^{\tau_2}(\omega), \quad \tau_1, \tau_2 \in (0,1),$$

and the *rank-based Laplace-periodogram kernel*

$$L_{n,R}^{\tau_1,\tau_2}(\omega) := \frac{n}{4}\tilde{b}_{n,R}^{\tau_1}(\omega)' \begin{pmatrix} 1 & -\mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} \tilde{b}_{n,R}^{\tau_2}(\omega), \quad \tau_1, \tau_2 \in (0,1).$$

Next, the asymptotic properties of the kernels defined above were derived under the assumption that the data $X_1, ..., X_n$ come from a strictly stationary time series. More precisely, let $\Omega := \{\omega_1, \ldots, \omega_\nu\} \subset (0, \pi)$ denote distinct frequencies and $T := \{\tau_1, \ldots, \tau_p\} \subset (0,1)$ distinct quantile orders. Then, under certain technical assumptions we have (see [1] for details)

$$(L_n^{\tau_1,\tau_2}(\omega_1), \ldots, L_n^{\tau_1,\tau_2}(\omega_\nu)) \rightsquigarrow (L^{\tau_1,\tau_2}(\omega_1), \ldots, L^{\tau_1,\tau_2}(\omega_\nu))$$

where $L^{\tau_1,\tau_2}(\omega_j)$ are independent random variables such that

$$L^{\tau_1,\tau_2}(\omega_j) \sim \frac{1}{2} f^{\tau_1,\tau_2}(\omega_j)\chi^2_2 \quad \text{if } \tau_1 = \tau_2,$$

and

$$f^{\tau_1,\tau_2}(\omega) := \frac{1}{f(q_{\tau_1})f(q_{\tau_2})} \sum_{k=-\infty}^{\infty} \mathrm{Cov}(I\{F(X_1) \le \tau_1\}, I\{F(X_k) \le \tau_2\})e^{-\mathrm{i}k\omega}$$

with F denoting the distribution function of X_1 as well as

$$L^{\tau_1,\tau_2}(\omega_j) \stackrel{d}{=} \frac{1}{4} \begin{pmatrix} Z_{11} \\ Z_{12} \end{pmatrix}' \begin{pmatrix} 1 & -\mathrm{i} \\ \mathrm{i} & 1 \end{pmatrix} \begin{pmatrix} Z_{21} \\ Z_{22} \end{pmatrix} \quad \text{if } \tau_1 \neq \tau_2$$

where $(Z_{11}, Z_{12}, Z_{21}, Z_{22}) \sim \mathcal{N}_4(0, \Sigma_4(\omega_j))$, with

$$\Sigma_4(\omega_j) = \frac{1}{2} \begin{pmatrix} f^{\tau_1,\tau_1}(\omega_j) & 0 & \Re f^{\tau_1,\tau_2}(\omega_j) & -\Im f^{\tau_1,\tau_2}(\omega_j) \\ 0 & f^{\tau_1,\tau_1}(\omega_j) & \Im f^{\tau_1,\tau_2}(\omega_j) & \Re f^{\tau_1,\tau_2}(\omega_j) \\ \Re f^{\tau_1,\tau_2}(\omega_j) & \Im f^{\tau_1,\tau_2}(\omega_j) & f^{\tau_2,\tau_2}(\omega_j) & 0 \\ -\Im f^{\tau_1,\tau_2}(\omega_j) & \Re f^{\tau_1,\tau_2}(\omega_j) & 0 & f^{\tau_2,\tau_2}(\omega_j) \end{pmatrix}.$$

For the quantity $L_{n,R}^{\tau_1,\tau_2}(\omega)$ a similar result with all instances of $f^{\tau_1,\tau_2}(\omega)$ replaced by

$$f_R^{\tau_1,\tau_2}(\omega) := \sum_{k=-\infty}^{\infty} \mathrm{Cov}(I\{F(X_1) \le \tau_1\}, I\{F(X_k) \le \tau_2\})e^{-\mathrm{i}k\omega}$$

can be derived. In particular, the absence of weighting in the definition of $f_R^{\tau_1,\tau_2}(\omega)$ shows that the rank-based version is preferable. Moreover, the results above demonstrate that by considering rank based Laplace periodogram kernels we are able to obtain a rich view of time series dynamics and at the same time separate marginal features of a time series from its dynamic structure.
In the final part of the talk, we demonstrated that smoothing can be used to obtain consistent estimators of the quantities $f_R^{\tau_1,\tau_2}$ and f^{τ_1,τ_2}. Interesting open questions that we are currently working on include the asymptotic distribution of smoothed Laplace peridograms, their application in the setting of signal detection as well as extensions of the above methods to the setting of locally stationary time series.

References

[1] H. Dette, M. Hallin, T.Kley and S. Volgushev *Of Copulas, Quantiles, Ranks and Spectra: an L_1-approach to spectral analysis*, **arXiv:1111.7205** (2011).

[2] T.-H. Li. *Laplace periodogram for time series analysis.* Journal of the American Statistical Association, **103** (2008) 757–768.

[3] T.-H. Li. *Quantile periodograms.* Journal of the American Statistical Association, **107** (2012) 765–776.

[4] R. Koenker and G. Bassett. *Regression quantiles.* Econometrica, **46** (1978) 33–50.

Quantile based spectral analysis for nonstationary time series

Stefan Skowronek

(joint work with Holger Dette, Marc Hallin, Tobias Kley, Stanislav Volgushev)

In the field of time series analysis a lot of powerful tools have been developed under the assumption of (strict) stationarity. However in typical applications such a restriction can hardly be justified and non-stationary techniques seem to be appropriate. Recently the concept of locally stationary processes could be used to transfer important tools from stationary to non-stationary time series. The idea of modeling local stationarity is based on the assumption, that a given process behaves approximately stationary over a short period of time. This idea also underlines the definitions from [4], [2] and [3]. Although the technical definitions have to be adapted to the specific problems, all three approaches have in common, that a process $(X_{t,T})_{1\leq t\leq T}$ is called locally stationary, if for each rescaled point in time $\vartheta = \frac{t}{T}$ there exists a process $X_t(\vartheta)$ which satisfies

- $X_t(\vartheta)$ is stationary
- $X_t(\vartheta)$ approximates $X_{t,T}$ locally in time in a suitable fashion.

To obtain an meaningful asymptotic theorie consider rescaled time $\frac{t}{T}$ instead of the original time t. We want to apply the concept of local stationarity to quantile based spectral analysis, where the objects of interest are cross covariant kernels, that can be written in terms of distribution functions. Therefore it seems intuitively natural to use differences of distribution functions to measure the distance between the non-stationary process and its stationary approximation.

Definition 1. *A sequence of stochastic processes $(X_{t,T})_{1\leq t\leq T}$ is called locally strictly stationary (of order 2) if for each time point ϑ there exists a strictly stationary processes $X_j(\vartheta)$ such that*

$$||F_{u,v}(x,y;T) - G_{u-v}(x,y;\vartheta)||_\infty \leq L\max(|u/T-\vartheta|, |v/T-\vartheta|), \tag{1}$$

where $F_{u,v}(x,y;T)$ and $G_{u-v}(x,y;\vartheta)$ denote the joint distribution functions of $(X_{u,T}, X_{v,T})$ and $(X_j(\vartheta), X_{j+h}(\vartheta))$ respectively.

If we let y tend to infinity, we get an analogous condition for the marginal distributions $F_{t,T}(x)$ and $G(x;\vartheta)$. From a more heuristic point of view this means, that the distribution function $F_{t,T}(x)$ and the dependence structure of the time series $(X_{t,T})_{1\leq t\leq T}$ are allowed to change continuously over time t. One advantage of this definition is its nonparametric character, as it doesn't depend on the data generating mechanism. A similar definition in [4] considers processes that have a tvMA(∞) representation with time varying coefficients. In fact under conditions used by Dahlhaus these processes are locally strict stationary in the sense of the definition above. The approach in [3] exhibits the same nonparametric structure but is based on a distance in probability rather than our distance in distributions.

With our concept of local strict stationarity we can define a nonstationary version of the time varying covariance kernel of lag h of $(X_{t,T})_{t\in\mathbb{Z}}$ at time ϑ by

$$\gamma_h(x_1,x_2;\vartheta) = \mathrm{Cov}(1\{X_j(\vartheta)\leq x_1\}, 1\{X_{j+h}(\vartheta)\leq x_2\}). \tag{2}$$

It is important, that unlike in traditional spectral analysis we do not need assumptions on the moments of the process itself, as we consider covariances of indicators. These covariances always exist and provide a canonical description of the joint distributions. If we assume, that the covariance kernels are absolutely summable we can define a time varying spectral density by

$$\mathfrak{f}_{x_1,x_2}(\omega,\vartheta) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\gamma_h(x_1,x_2;\vartheta)e^{-ih\omega}. \tag{3}$$

That this is indeed a meaningful definition can be seen if we consider an indicator version of the Wigner-Ville spectrum (we set $X_{t,T}=0$ for $t\notin[1,T]$)

$$f^T_{x_1,x_2}(\omega,\vartheta) := \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\mathrm{Cov}(1_{\{X_{\lfloor\vartheta T-s/2\rfloor,T}\le x_1\}},1_{\{X_{\lfloor\vartheta T+s/2\rfloor,T}\le x_2\}})e^{-i\omega s}. \tag{4}$$

Under the assumption of local strict stationarity and if $\mathfrak{f}_{x_1,x_2}(\omega,\vartheta)$ and $\gamma_h(x_1,x_2;\vartheta)$ are finite, we get

$$\sup_{\omega,\vartheta}|\mathfrak{f}_{x_1,x_2}(\omega,\vartheta)-\gamma_h(x_1,x_2;\vartheta)| = o(1).$$

That means that our tv spectral density can be interpreted as the limit of the indicator version of the Wigner-Ville spectrum. Analogue to the strictly stationarity case Dette et al. [1] have investigated, $\mathfrak{f}_{x_1,x_2}(\omega,\vartheta)$ can be estimated by a local version of the Laplace-periodogram kernel

$$L_T^{\tau_1,\tau_2}(\omega,\vartheta) := \frac{T}{4}\tilde{b}_n^{\tau_1}(\omega,\vartheta)'\begin{pmatrix}1 & -\mathrm{i}\\ \mathrm{i} & 1\end{pmatrix}\tilde{b}_T^{\tau_2}(\omega,\vartheta),\quad \tau_1,\tau_2\in(0,1),$$

where for $N=o(\sqrt{T})$, $c_u(\omega)=(1,\cos(u\omega),\sin(u\omega))'$ and $\rho_\tau(x):=x(\tau-1_{(-\infty,0]}(x))$

$$(\tilde{a}_T^\tau(\omega,\vartheta),\tilde{b}_{1T}^\tau(\omega,\vartheta),\tilde{b}_{2T}^\tau(\omega,\vartheta)) := \arg\min_{b\in\mathbb{R}^3}\sum_{u=1}^{N}\rho_\tau\left(X_{\lfloor\vartheta T-N/2\rfloor}-c_u(\omega)'b\right).$$

Within this setting it is straight forward to proof that the results from Dette et al. still hold in a nonstationary context.

References

[1] H. Dette, M. Hallin, T.Kley and S. Volgushev *Of Copulas, Quantiles, Ranks and Spectra: an L_1-approach to spectral analysis*, **arXiv:1111.7205** (2011).

[2] Z. Zhou and W. B. Wu, *Local linear quantile estimation for nonstationary time series*, Annals of Statistics **37** (2009), 2696–2729.

[3] M. Vogt, *Nonparametric Regression for Locally Stationary Time Series*, Forthcoming in Annals of Statistics.

[4] R. Dahlhaus, *A likelihood approximation for locally stationary processes*, Annals of Statistics **28** (2000), 1762–1794.

Analysis on Censored Quantile Residual Life Model via Spline Smoothing

Ying Wei
(joint work with Yanyuan Ma)

Residual life is defined as the remaining time to event given the fact that the survival time T of a patient is at least t, i.e., $T-t \mid T \geq t$. In many clinical studies, especially when the associated diseases are chronic or/and incurable, knowing residual life is the major concern to patients. We propose a general class of quantile residual life models, where a specific quantile of the residual life time, conditional on an individual has survived up to time t, is a function of certain covariates with their coefficients varying over time. Specifically, the model we consider can be written as following.

$$Q_\tau(T_i - t \mid X_i, T_i \geq t) = m(X_i, \beta(t)), \quad t > 0,$$

where $Q_\tau(T|A)$ denotes the τth conditional quantile function of a random variable T conditional on an event A, τ is a quantile level ranging between 0 and 1, and t is the time at which the residual life is considered. Here, $m(\cdot)$ is a parametric function of covariate X, while the parameter $\beta(t) = \{\beta_1(t), \beta_2(t), \ldots, \beta_p(t)\}^T$ consists of p unknown smooth functions of t. A special case of the model is the familiar linear varying-coefficient model

$$Q_\tau(T_i - t \mid X_i, T_i \geq t) = X_i^T \beta(t), \quad t > 0.$$

By taking into consideration that $\beta(t)$ is a smooth function of t, we can obtain a unified presentation of the residual life over a period of time, which is of interest in many applications. Moreover, compared to estimating the residual life at given times separately, we can achieve a more efficient estimator by estimating $\beta(t)$ globally.

We propose to estimate the coefficient functions using spline approximation. Specifically, we take $b(t) = [\pi_1(t), \ldots, \pi_{k_n}(t)]^T$ as k_n B-spline basis functions given a set of internal knots and the order of spline, and then approximate $\beta(t)$ by $\beta(t) \approx \alpha b(t)$, where α is a $p \times k_n$ matrix of unspecified parameters. With spline representation of $\beta(t)$, we proposed two estimation approaches. In the first approach, we construct a cumulative estimation equations, and solve for $\beta(t)$ with one step optimization. Specifically, at a given time t_j, we construct the following estimation equations

$$\begin{aligned} & s(\alpha, t_j) \\ &= \sum_{i=1}^{n} \frac{\partial m\{X_i, \boldsymbol{\alpha B}(t_j)\}}{\partial \alpha} \left(\frac{I\left[Y_i \geq t_0 + m\{X_i, \boldsymbol{\alpha B}(t_j)\}\right]}{G\left[t_j + m\{X_i, \boldsymbol{\alpha B}(t_j)\}\right]} - (1-\tau) \frac{I(Y_i \geq t_j)}{G(t_j)} \right) = 0, \end{aligned}$$

and then combining $t = t_1, \ldots, t_J$. The estimator of α can be obtained from

$$\min_{\boldsymbol{\alpha}} S(\boldsymbol{\alpha}) = \sum_{j=1}^{J} s(\alpha, t_j)^{\otimes 2}$$

One can choose an arbitrary set of times $t_1, \ldots, t_J$ as long as at least k_n of the J corresponding equations are linearly independent, for then the estimator is uniquely defined. Note that this requires that $J \geq k_n$, so the number of distinctive event/censor times is larger than the number of B-spline basis functions. Our subsequent theoretical development further requires that there exist $\epsilon > 0$ so that $J = o(n^{1/2-\epsilon})$.

The one-step optimization is computationally intensive. Another way to alleviate the computational burden is to estimate $\beta(t)$ separately at $t_1, \ldots, t_J$. Denoting the resulting estimator $\check{\boldsymbol{\beta}}(t_j)$'s, we then obtain the estimate of α by

$$\min_{\boldsymbol{\alpha}} \sum_{j=1}^{J} \left\{ \boldsymbol{\alpha} \boldsymbol{B}(t_j) - \check{\boldsymbol{\beta}}(t_j) \right\}^{\otimes 2}$$

to obtain

$$\widetilde{\boldsymbol{\alpha}} = \left\{ \sum_{j=1}^{J} \check{\boldsymbol{\beta}}(t_j) \boldsymbol{B}^{\mathrm{T}}(t_j) \right\} \left\{ \sum_{j=1}^{J} \boldsymbol{B}(t_j) \boldsymbol{B}^{\mathrm{T}}(t_j) \right\}^{-1}.$$

The two step approach reducing a one step $p \times k_n$ dimensional optimization into J separate p dimensional problem, andn hence it is more computationally efficient.

We developed large sample theorems for the estimators from both one-step and two-step optimizations. We show that both estimators are consistent and asymptotically normally distributed. Based on the derived limiting variance-covariance matrix, inference tools to test the significance of the covariate effect on residual life are available following the classical Wald test theories. We compare the two methods in terms of their asymptotic efficiency and computational complexity. They are asymptotically equivalent when $J = k_n$, and when incorporating the variance-covariance matrix into the estimation.

We finally investigated the finite sample performance of the estimation and testing procedures through numerical experiments. We also apply the methods to a real data set from a neurological study.

References

[1] Y. Ma and Y. Wei *Analysis on Censored Quantile Residual Life Model via Spline Smoothing*, Statistica Sinica **22** (2012), 47–68.

Mardenized Multivariate Quantiles?

Xuming He

Quantile is a useful notion for univariate data. Because it is closely related to ranking and ordering, its generalization to multivariate data is not so straightforward. It is not even clear whether there is a universally accepted notion of multivariate ranking and quantile. Two recent papers, Hallin, Paindaveine, and Siman (2010) and Kong and Mizera (2012), provided complementary perspectives on the problem of multivariate quantiles. In this talk, I introduce an earlier discovery of John Marden that could lead to a very attractive notion of multivariate quantiles. I

first learned about this in an informal seminar given by Marden at the University of Illinois in 1999.

Given a set of points $\mathbf{X} = \{x_1, \cdots, x_n\}$ in R^p, we define their positions through the following transformation

$$P(x; \mathbf{X}) = \frac{c_p}{n} \sum_{i=1}^{n} \frac{x - x_i}{||x - x_i||}$$

for some constant c_p. If we call the set of points after the transformation as $\mathbf{X^1} = \{P(x_1; \mathbf{X}), \cdots, P(x_n; \mathbf{X})\}$, we can then iterate this transformation as $\mathbf{X} \to \mathbf{X^1} \to \mathbf{X^2}$.... This iteration has a stationary point. Under mild conditions, this iteration seems to converge to a stationary distribution that is independent of the distribution of $\mathbf{X}$. For bivariate data $p = 2$ with $c_p = \pi/4$, the stationary distribution is well-known as the least informative distribution in a unit circle. It is also interesting to note that the inverse operation of this transformation can be characterized by a convex optimization, so we can use the positions in the circle to represent the bivariate quantile of the original data. I shall demonstrate numerically that the convergence of this iterative transformation is usually achieved with a small number of steps. Limited experience with this approach suggests that this approach leads to (nearly?) affine equivariant solutions, even though each step of the iteration does not have this property.

My talk is based on preliminary investigations of what I might call "Mardenized quantiles", and I hope that my presentation will stimulate interest in this promising direction.

References

[1] M. Hallin, D. Paindaveine, and M. Siman, *Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth*, Annals of Statistics **38** (2010), 635 –669.

[2] L. Kong and I. Mizera, *Quantile tomography: using quantiles with multivariate data*, Statistica Sinica **22** (2012), 1589-1610.

Reporter: Stefan Skowronek

Participants

Prof. Dr. Alex Belloni
Duke University
The Fuqua School of Business
100 Fuqua Drive
Durham, NC 27708
UNITED STATES

Prof. Dr. Stephane Bonhomme
CEMFI
Centro de Estudios Monetarios y Financ.
Calle Casado del Alisal 5
28014 Madrid
SPAIN

Prof. Dr. Victor Chernozhukov
Department of Economics
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge, MA 02142-1347
UNITED STATES

Prof. Dr. Holger Dette
Fakultät für Mathematik
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum
GERMANY

Dr. Juan Carlos Escanciano
Department of Economics
Indiana University
100 S. Woodlawn
Bloomington, IN 47401
UNITED STATES

Prof. Dr. Ivan Fernandez-Val
Boston University
Department of Economics
270 Bay State Road
Boston, MA 02215-1403
UNITED STATES

Prof. Dr. Alfred Galichon
Departement de Economique
Sciences Po
28 Rue des Saint-Peres
75007 Paris
FRANCE

Prof. Dr. Marc Hallin
Université Libre de Bruxelles
CP 114
50 avenue F. D. Roosevelt
1050 Bruxelles
BELGIUM

Prof. Dr. Wolfgang Karl Härdle
Wirtschaftswissenschaftliche Fakultät
Ladislaus v. Bortkiewicz Chair of Statistics CASE
Humboldt-Universität zu Berlin
Unter den Linden 6
10117 Berlin
GERMANY

Prof. Dr. Xuming He
Department of Statistics
University of Michigan
462 West Hall
1085 South University Ave.
Ann Arbor, MI 48109-1107
UNITED STATES

Prof. Dr. Roger W. Koenker
Department of Economics
University of Illinois
330 Commerce Bldg. (West)
Champaign, IL 61820
UNITED STATES

Prof. Dr. Chenlei Leng
Dept. of Statistics & Applied Probability
National University of Singapore
2 Science Drive 6
Singapore 117 546
SINGAPORE

Prof. Dr. Ivan Mizera
Dept. of Mathematics & Statistics
University of Alberta
632 Central Academic Bldg.
Edmonton AB T6G 2G1
CANADA

Prof. Dr. Stephen Portnoy
University of Illinois at
Urbana-Champaign
116 E. Illini Hall
725 S.Wright St.
Champaign, IL 61820
UNITED STATES

Stefan Skowronek
Fakultät für Mathematik
Ruhr-Universität Bochum
Universitätsstr. 150
44801 Bochum
GERMANY

Prof. Dr. Stanislav Volgushev
Fakultät für Mathematik
Ruhr Universität Bochum
Universitätsstr. 150
44780 Bochum
GERMANY

Prof. Dr. Ying Wei
Columbia University
Department of Biostatistics
722 West 168th Street
New York, NY 10032
UNITED STATES