Open Journal of Statistics, 2016, 6, 995-1002
http://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Volatile Compounds Selection via Quantile Correlation and Composite Quantile Correlation: A Whiting Case Study

Ibrahim Sidi Zakari[1*], Assi N'guessan[2], Alexandre Dehaut[3], Guillaume Duflos[3]

[1]Department of Mathematics and Computer Science, Abdou Moumouni University, Niamey, Niger
[2]Paul Painlevé Laboratory, UMR CNRS 8524, Lille 1 University, Lille, France
[3]ANSES, Laboratoire de Sécurité des Aliments Département des Produits de la Pêche et de l'Aquaculture, Boulevard du Bassin Napoléon, Boulogne-Sur-Mer, France
Email: *isidizakari@refer.ne, assi.nguessan@polytech-lille.fr, alexandre.dehaut@anses.fr, guillaume.duflos@anses.fr

How to cite this paper: Zakari, I.S., N'guessan, A., Dehaut, A. and Duflos, G. (2016) Volatile Compounds Selection via Quantile Correlation and Composite Quantile Correlation: A Whiting Case Study. *Open Journal of Statistics*, **6**, 995-1002.
http://dx.doi.org/10.4236/ojs.2016.66079

Received: August 10, 2016
Accepted: November 8, 2016
Published: November 14, 2016

Abstract

The freshness and quality indices of whiting (*Merlangius merlangus*) influenced by a large number of chemical volatile compounds, are here analyzed in order to select the most relevant compounds as predictors for these indices. The selection process was performed by means of recent statistical variable selection methods, namely robust model-free feature screening, based on quantile correlation and composite quantile correlation. On the one hand, compounds 2-Methyl-1-butanol, 3-Methyl-1-butanol, Ethanol, Trimethylamine, 3-Methyl butanal, 2-Methyl-1-propanol, Ethylacetate, 1-Butanol and 2,3-Butanedione were identified as major predictors for the freshness index and on the other hand, compounds 3-Methyl-1-butanol, 2-Methyl-1-butanol, Ethanol, 3-Methyl butanal, 3-Hydroxy-2-butanone, 1-Butanol, 2,3-Butanedione, 3-Pentanol, 3-Pentanone and 2-Methyl-1-propanol were identified as major predictors for the quality index.

Keywords

Volatile Compounds, Freshness and Spoilage Indices, Quantile Correlation, Composite Quantile Correlation, Sure Independence Screening

1. Introduction

Fish freshness is a key attribute for the quality of fish, which is a highly perishable product. The fishing industry is an important contributor to many economies in the world. One of the senses used by consumers to determine the freshness of fish is the smell. Indeed, the volatilome of fish changes rapidly according to the product degree of freshness, and that is why sensory analysis are used by consumers and industrialists to assess fish quality. Then, the key volatile compounds that contribute to this characteristic odor can be measured and used as quality indicators [1] [2] [3]. These characteristic aromatic volatile compounds are generated by different biological pathways including the lipid autoxidation, the action of spoilage organisms and autolytic enzymes.

Recently, Duflos *et al.* [1] studied the spoilage of whiting at five stages of ice storage by comparing the analysis of volatile compounds obtained by solid phase microextraction (SPME) coupled to the combination of gaz chromatography/mass spectrometry (GC/MS) and SPME with two sensory analysis methods. Two separate steps of statistical multidimensional approaches were used to identify volatile compounds and characterize fish freshness assessed by two different indices. In the first step, control charts were used to control the daily progression of freshness and spoilage indices. The second step begins by reducing the dimension of the data set (excluding the two indices variables) to two principal components via the application of Principal Component Analysis (PCA) method.

Then, a hierarchical clustering approach and a heuristic variable selection were used for clustering the fish samples on three classes and to identify the volatile compounds that respectively characterize these classes. However, the indices (or response variables) were not directly taken into account in the later procedure.

Recently, Sidi *et al.* [4] applied stability selection and randomization techniques in L_1 norm penalized quantile regression on the same data set. These approaches highlighted volatile compounds that are more relevant for the evaluation of fish freshness throughout its storage, so, are assumed to influence more the fish freshness and quality.

Using penalized quantile regression approaches on whiting data set is motivated by the fact that consumers, generally faced different categories of fish freshness. The interest of quantile regression approach is its ability to provide a model for each level of quality. More details on quantile regression and penalized quantile regression can be found in the following references [5]-[12].

This paper aims at using Ma and Zhang [13] approach to select a reduced subset of volatile compounds which can be used to explain whiting spoilage during its conservation. This approach allows a robust and model-free feature screening based on quantile correlation proposed by Li *et al.* [14].

The lines below are organized as follow: The methodology is briefly presented in section 2 and section 3 is dedicated to the experimental framework. Finally, the results are discussed in section 4 followed by concluding remarks in section 5.

2. Methodology

This section is dedicated to the following methods: Quantile Correlation, Sure Independence Screening via Quantile Correlation, Composite Quantile Correlation and Sure Independence Screening via Composite Quantile Correlation.

2.1. Quantile Correlation

As advocated in Li *et al.* [14], quantile correlation is a novel measure used to examine the linear relationship between any two random variables Y and X for a given quantile $\tau \in (0,1)$. So,

$$qcor_\tau(Y,X) = \frac{qcov_\tau(Y,X)}{\sqrt{var(\psi_\tau(Y-Q_{\tau,Y}))var(X)}} = \frac{E(\psi_\tau(Y-Q_{\tau,Y}))(X-E(X))}{\sqrt{(\tau-\tau^2)var(X)}} \quad (1)$$

where

$$qcov_\tau(Y,X) = cov(I(Y-Q_{\tau,Y}>0),X) = E(\psi_\tau(Y-Q_{\tau,Y}))(X-E(X)) \quad (2)$$

$Q_{\tau,Y}$ is the τ^{th} conditional quantile of Y and

$$\psi_\tau(s) = \tau - I(s<0) \quad (3)$$

Moreover, if X is independent of Y, the $qcor_\tau(Y,X)=0$; else (X and Y are correlated), the $qcor_\tau(Y,X)\neq 0$.

2.2. Sure Independence Screening via Quantile Correlation

Consider Y as the dependent variable and $X=(X_1,\cdots,X_p)^{\mathrm{T}}$ be the p-dimensional independent variables. Let $\boldsymbol{\omega}=(\omega_1,\cdots,\omega_p)^{\mathrm{T}}$ with

$$\omega_k = E(qcor_\tau^2(Y,X_k)), k=1,\cdots,p \quad (4)$$

Sure Independent Screening via Quantile Correlation method selects the first d independent variables with largest $\hat{\omega}_k$; where

$$\hat{\omega}_k = \frac{n^3}{n(n-1)(n-2)\bar{\omega}_k} \quad (5)$$

with $\bar{\omega}_k$ the sample estimate of ω_k.

2.3. Composite Quantile Correlation

Composite Quantile Correlation(CQC) is motivated by the fact that previous quantile correlation cannot characterize the entire relationship between X and Y. So, the composite quantile correlation is defined by:

$$cqcor(Y,X) = E(qcor_\tau(Y,X)) = \int_0^1 qcor_\tau(Y,X)\,\mathrm{d}\tau \quad (6)$$

2.4. Sure Independence Screening via Composite Quantile Correlation

The CQC screening is based on the vector $\boldsymbol{\omega}=(\omega_1,\cdots,\omega_p)^{\mathrm{T}}$ with components

$$\omega_k = |cqcor(Y,X_k)|, k=1,\cdots,p. \quad (7)$$

Sub models are selected based on decreasing values of ω_k. Furthermore, as advocated by Ma and Zhang [13], when using screening techniques, the number of selected variables is often set to be $n-1$ or the integer part of $\frac{n}{\log(n)}$.

3. Materials and Methods

The sample preparation and sensory evaluation methods are briefly presented in this section. More details about the full experimental procedure can be found in [1].

3.1. Sample Preparation

As advocated by Duflos *et al.* [1], the sample considered is based on two different catches of respectively 20 and 15 fish. These catches were stored in crushed ice at 4˚C in self-draining polystyrene boxes for 7 days. Fresh crushed ice was added daily. Sensory evaluation and volatile analysis were performed on seven different fish on days 1, 2, 3, 4 and 7.

3.2. Sensory Evaluation

According to Duflos *et al.* [1], two methods were used for the sensory evaluation of fish. These methods lead to freshness and quality indices which represent two response variables for our selection process.

4. Results and Discussion

The empirical results of the analysis of freshness and quality indices influenced by a great number of volatile compounds are presented below.

The sample size and the number of predictors (volatile compounds) are respectively $n = 35$ and $p = 55$. So, the number of predictors is higher than the sample size.

In order to perform variables selection, screening methods are applied on whiting data set using QC-SIS package available for R software.

The tuning parameter d used to select covariates with significant effect on each response variable can be set to $n - 1 = 34$ or the integer part of $\frac{n}{\log(n)} = 9$.

The results for $d = 9$ are presented in **Table 1** and **Table 2**.

Furthermore, **Figure 1** and **Figure 2** display the Pearson correlation matrix through bivariate scatter plots for each index with corresponding selected compounds. These figures have been made using *Performance Analytics* package available for R software.

Table 1. Selected volatile compounds ranked by decreasing weights for freshness index

Selected	Weights	
Volatile Compounds	QC SIS	CQC SIS
2-Methyl-1-butanol	0.314	0.515
3-Methyl-1-butanol	0.292	0.492
Ethanol	0.264	0.478
Trimethylamine	0.140	0.327
3-Methyl butanal	0.132	0.322
2-Methyl-1-propanol	0.119	0.305
Ethylacetate	0.106	0.302
1-Butanol	0.102	0.291
2,3-Butanedione	0.099	0.278

Table 2. Selected volatile compounds ranked by decreasing weights for quality index.

Compounds	QC SIS Weights	Compounds	CQC SIS Weights
3-Methyl-1-butanol	0.302	3-Methyl-1-butanol	0.501
2-Methyl-1-butanol	0.288	2-Methyl-1-butanol	0.499
Ethanol	0.224	Ethanol	0.450
3-Methyl butanal	0.158	3-Methyl butanal	0.363
3-Hydroxy-2-butanone	0.122	1-Butanol	0.327
1-Butanol	0.119	3-Hydroxy-2-butanone	0.286
2,3-Butanedione	0.097	3-Pentanol	0.272
3-Pentanol	0.094	2,3-Butanedione	0.271
3 Pentanone	0.090	2-Methyl-1-propanol	0.245

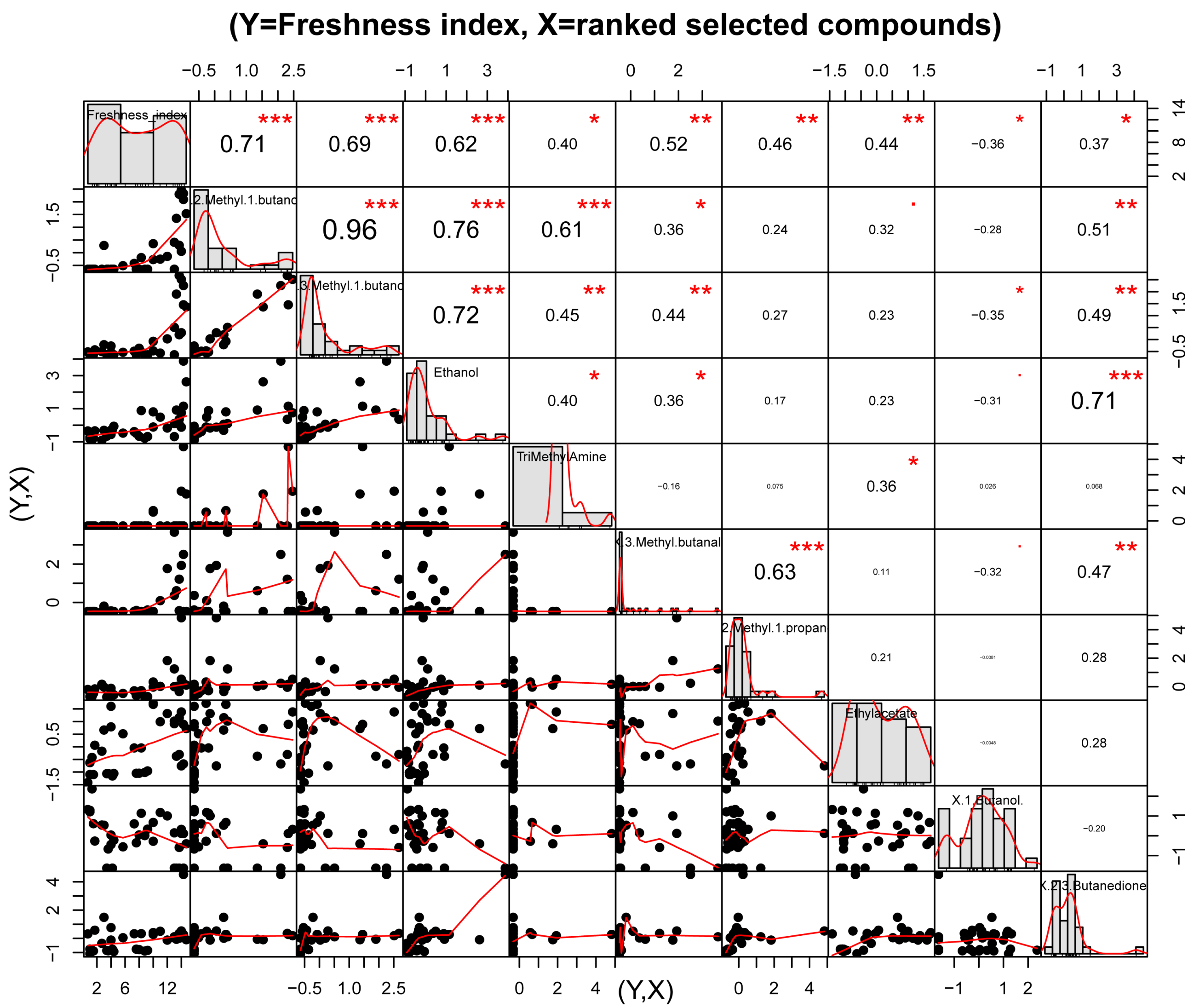

Figure 1. Correlation matrix chart for freshness index and nine(9) related compounds: On top; the (absolute) value of the correlation with significance levels. The distribution of each variable is represented on the diagonal and at bottom, the bivariate scatterplots, with a fitted line. Components of the vector are respectively tagged with symbols corresponding to the associated p-values: "***" (p-value ≤ 0.001), "**" (p-value ≤ 0.01), "*" (p-value ≤ 0.05), "." (p-value ≤ 0.1), " " (p-value ≤ 1).

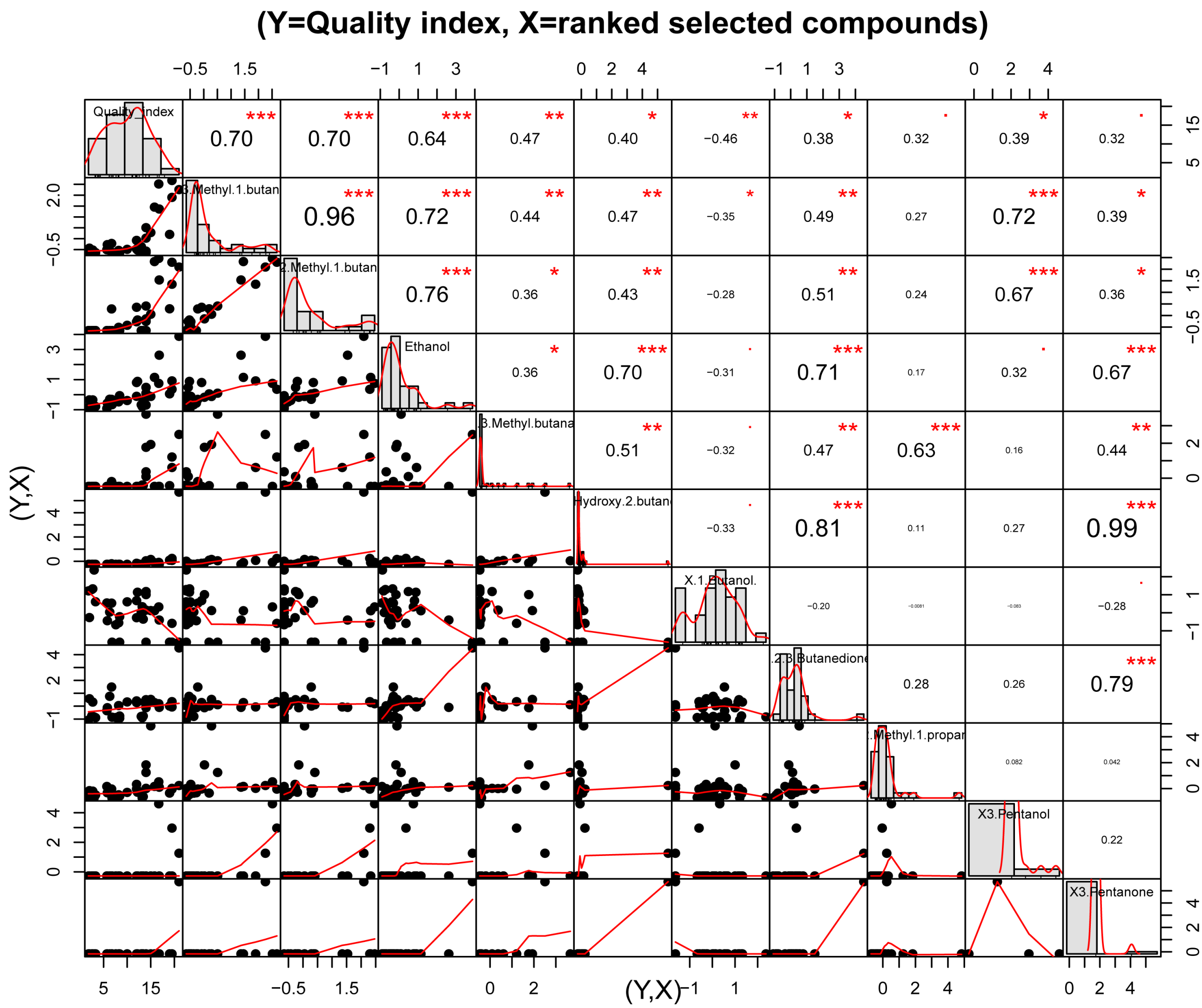

Figure 2. Correlation matrix chart for quality index and ten(10) related compounds: On top; the (absolute) value of the correlation with significance levels. The distribution of each variable is represented on the diagonal and at bottom, the bivariate scatterplots, with a fitted line. Components of the vector are respectively tagged with symbols corresponding to the associated p-values: "***" (p-value ≤ 0.001), "**" (p-value ≤ 0.01), "*" (p-value ≤ 0.05), "." (p-value ≤ 0.1), " " (p-value ≤ 1).

4.1. Results for Freshness Index

According to Table 1, Quantile Correlation and Composite Quantile Correlation Sure Independent Screening methods select the same subset of volatile compounds for freshness index.

These compounds have been previously identified as spoilage markers.

For example, the compounds Ethanol, 3-Methyl-1-butanol, 2,3-Butanedione, 2-Methyl-1-butanol, 3-Methyl butanal, 2-Methyl-1-propanol and Ethylacetate are identified as correlated to the second principal component axis in Duflos *et al.* [1].

The previous seven compounds were included in the eight compounds (with limonene) that characterized category 2/3 (intermediate category between freshness and spoilage) in Duflos *et al.* [1].

Moreover, compounds Trimethylamine and 1-Butanol were not identified as correlated to the first principal component axis in Duflos *et al*. [1].

Finally, considering Freshness index, only Trimethylamine, 1-Butanol and 2-Methyl-1-butanol were not selected by L_1 QR and other randomization approaches highlighted in Sidi *et al*. [4].

4.2. Results for Quality Index

According to **Table 2**, Quantile Correlation and Composite Quantile Correlation Sure Independent Screening methods do not select the same subset of volatile compounds for quality index.

The compounds Ethanol, 3-Methyl-1-butanol, 2,3-Butanedione, 2-Methyl-1-butanol, 3-Methyl butanal and 2-Methyl-1-propanol are identified as correlated to the second principal component axis in Duflos *et al*. [1].

The previous six compounds were included in the eight compounds (with limonene) that characterized category 2/3 (intermediate category between freshness and spoilage) in Duflos *et al*. [1].

The compounds 1-Butanol, 3-Pentanol, 3-Pentanone and 3-Hydroxy-2-butanone are not identified as correlated to the first principal component axis in Duflos *et al*. [1].

For the quality index, only Ethanol, 3-Methyl butanal, 3-Methyl-1-butanol and 2-Methyl-1-butanol were selected in Sidi *et al*. [4].

Choosing $d = n - 1 = 34$ includes freshness markers like Propanal, Hexanal, 1-Penten-3-ol, Pentanal, 2,3-Pentanedione, 1-Penten-3-one, Heptanal, (E)-2-Pentenal, (Z)-2-Penten-1-ol, 1-Pentanol, Butanal, Octanal, 1-hexanol and 4,4-Dimethyl-1,3-dioxane.

5. Concluding Remarks

Sure Independence Screening via Quantile Correlation and Composite Quantile Correlation methods highlighted relevant volatile compounds influencing freshness and quality indices during whiting conservation.

The selected compounds include Ethanol, 3-Methyl-1-butanol, 2,3-Butanedione, 2-Methyl-1-butanol, 3-Methyl butanal, 2-Methyl-1-propanol and Ethylacetate, previously identified as spoilage markers.

For future investigation on whiting data, it will be very interesting to explore the following issues:

1) Simultaneous model selection in multiple quantile regression [11]

2) Selection of groups of highly correlated compounds [8]

3) Quantile regression models and inference processes based on [15] and [16].

Acknowledgements

The authors wish to thank Pr Abdallah Mkhadri (Cadi Ayyad University of Morroco), François Leduc (ANSES), The Nord-Pas-de-Calais Regional Council (France) and ANSES (France). This support is greatly appreciated.

References

[1] Duflos, G., Leduc, F., N'Guessan, A., Krzewinski, F., Ossarath, K. and Malle, P. (2010) Freshness Characterisation of Whiting (*Merlangius merlangus*) Using an SPME/GC/MS Method and a Statistical Multivariate Approach. *Journal of the Science of Food and Agriculture*, **90**, 2568-2575. http://dx.doi.org/10.1002/jsfa.4122

[2] Duflos, G., Cornu, M., Coin, V.M., Antinelli, J.F. and Malle, P. (2006) Determination of Volatile Compounds to Characterize Fish Spoilage Using HS/MS and SPME-GC/MS Analysis. *Journal of Agricultural and Food Chemistry*, **86**, 600-611. http://dx.doi.org/10.1002/jsfa.2386

[3] Edirisinghe, R.K.B., Graffham, A.J. and Taylor, S.J. (2007) Characterisation of the Volatiles of Yellowfin Tuna (*Thunnus albacares*) during Storage by Solid Phase Microextraction and G-MS and Their Relationship to Fish Quality Parameters. *International Journal of Food Science & Technology*, **42**, 1139-1147. http://dx.doi.org/10.1111/j.1365-2621.2006.01224.x

[4] Sidi Zakari, I., Mkhadri, A. and N'Guessan, A. (2014) Stability Selection and Randomization in L1 Quantile Regression. In: Skiadas, C.H., Ed., *Theoretical and Applied Issues in Statistics and Demography*, International Society for the Advancement of Science and Technology (ISAST), Barcelona, 8 p.

[5] Koenker, R. and Basset, G. (1978) Regression Quantiles. *Econometrica*, **46**, 33-50. http://dx.doi.org/10.2307/1913643

[6] Koenker, R. (2005) Quantile Regression. Cambridge University Press, Cambridge.

[7] Sidi Zakari, I. (2013) Sélection de variables et régression sur les quantiles. Ph. D. Dissertation, Cadi Ayyad and Lille 1 Universities.

[8] Slawski, M. (2012) The Structured Elastic Net for Quantile Regression and Support Vector Classification. *Statistics and Computing*, **22**, 153-168. http://dx.doi.org/10.1007/s11222-010-9214-z

[9] Belloni, A. and Chernozhukov, V. (2011) L_1-Penalized Quantile Regression in High-Dimensional Sparse Models. *Annals of Statistics*, **39**, 82-130. http://dx.doi.org/10.1214/10-AOS827

[10] Wu, Y. and Liu, Y. (2009) Variable Selection in Quantile Regression. *Statistica Sinica*, **19**, 801-817.

[11] Zou, H. and Yuan, M. (2008) Regularized Simultaneous Model Selection in Multiple Quantiles Regression. *Computational Statistics and Data Analysis*, **52**, 5296-5304. http://dx.doi.org/10.1016/j.csda.2008.05.013

[12] Li, Y. and Zhu, J. (2008) L_1-Norm Quantile Regression. *Journal of Computational and Graphical Statistics*, **17**, 163-185. http://dx.doi.org/10.1198/106186008X289155

[13] Ma, X. and Zhang, J. (2015) Robust Model-Free Feature Screening via Quantile Correlation. *Journal of Multivariate Analysis*, **143**, 472-480. http://dx.doi.org/10.1016/j.jmva.2015.10.010

[14] Li, G., Li, Y. and Tsai, C.-L. (2015) Quantile Correlations and Quantile Autoregressive Modeling. *Journal of the American Statistical Association*, **110**, 246-261. http://dx.doi.org/10.1080/01621459.2014.892007

[15] Koenker, R. (2011) Additive Models for Quantile Regression: Model Selection and Confidence Bandaids. *Brazilian Journal of Probability and Statistics*, **25**, 239-262. http://dx.doi.org/10.1214/10-BJPS131

[16] Koenker, R. and Machado, A.F. (1999) Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association*, **94**, 1296-1310. http://dx.doi.org/10.1080/01621459.1999.10473882