Young Researchers' Day

1 February, 2013

9^{00}	**Nathan Uyttendaele**	Nonparametric reconstruction of the tree structure of a nested archimedean copula
9^{30}	**Benjamin Colling**	Goodnest of fit tests in semiparametric transformation models
10^{00}	**Aurélie Bertrand**	The Simex method for correcting the bias in a survival cure model with mismeasured covariates

Coffee Break

11^{00}	**Vincent Bremhorst**	Estimation of the latent distribution in cure survival models using a flexible Cox Model
11^{30}	**George Babajan**	Mathematical finance applied to gas and power derivatives
12^{00}	**Aleksandar Sujica**	The copula-graphic estimator in censored nonparametric location-scale regression models

The seminar is followed by the annual lunch of the ISBA.

The Young Researchers' Day is held in room c.115 of the Institut de statistique, biostatistique et sciences actuarielles, Voie du Roman Pays 20, Louvain-la-Neuve.

Nonparametric reconstruction of the tree structure of a nested archimedean copula

Nathan Uyttendaele (nathan.uyttendaele@uclouvain.be)

One of the features inherent in nested Archimedean copulas, also called hierarchical Archimedean copulas, is their rooted tree structure. In this paper, a completely nonparametric method to estimate this structure is developed. Our approach consist in representing the rooted tree structure as a set of trivariate structures that can be individually estimated. Indeed, for any triple of variables there are only four possible rooted tree structures and, based on a sample, a choice can be made by performing comparisons between the three marginal empirical bivariate distributions of the triple. The set of estimated trivariate structures can then be used to build an estimate of the mother rooted tree structure. This approach leads to an estimator that has reasonable properties, and a simulation study strongly suggest it can be made a consistent estimator for any nested Archimedean copula rooted tree structure.

References

[1] Barbe, P., C. Genest, K. Ghoudi, and B. Rémillard (1996). On Kendall's process. *Journal of Multivariate Analysis 58*(2), 197–229.

[2] Genest, C., J. Nešlehová, and J. Ziegel (2011). Inference in multivariate Archimedean copula models. *Test 20*(2), 223–256.

[3] Genest, C. and L. Rivest (2001). On the multivariate probability integral transformation. *Statistics & probability letters 53*(4), 391–399.

[4] Genest, C. and L.-P. Rivest (1993). Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association 88*(423), pp. 1034–1043.

[5] Hering, C., M. Hofert, J. Mai, and M. Scherer (2010). Constructing hierarchical Archimedean copulas with Lévy subordinators. *Journal of Multivariate Analysis 101*(6), 1428–1433.

[6] Hofert, J. (2010). *Sampling Nested Archimedean Copulas: With Applications to CDO Pricing*. Südwestdeutscher Verlag.

[7] Hofert, M. and M. Maechler (2011, 3). Nested Archimedean Copulas Meet R: The nacopula Package. *Journal of Statistical Software 39*(9), 1–20. Please note the package nacopula has been merged with the package copula.

[8] Hofert, M. and D. Pham (2012). Densities of nested Archimedean copulas. *arXiv preprint arXiv:1204.2410*.

[9] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.

[10] McNeil, A. J. (2008). Sampling nested Archimedean copulas. *Journal of Statistical Computation and Simulation 78*(6), 567–581.

[11] McNeil, A. J. and J. Nešlehová (2009, August). Multivariate Archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *ArXiv e-prints*.

[12] Nelsen, R., J. Quesada-Molina, J. Rodríguez-Lallena, and M. Úbeda-Flores (2003). Kendall distribution functions. *Statistics & probability letters 65*(3), 263–268.

[13] Okhrin, O., Y. Okhrin, and W. Schmid (2009). Properties of hierarchical Archimedean copulas. SFB 649 discussion paper 2009,014, Berlin.

[14] Okhrin, O., Y. Okhrin, and W. Schmid (2011). Determining the structure and estimation of hierarchical Archimedean copulas. second time revised and resubmitted to Journal of Econometrics on 26.08.2012.

[15] Okhrin, O. and A. Ristig (2012). Hierarchical Archimedean Copulae: The HAC Package. Technical report, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

[16] Puzanova, N. (2011). A hierarchical archimedean copula for portfolio credit risk modelling. *Natalia Puzanova. A hierarchical Archimedean copula for portfolio credit risk modelling. Deutsche Bundesbank Discussion Paper, Series 2.*

Goodnest of fit tests in semiparametric transformation models

Benjamin Colling (benjamin.colling@uclouvain.be)

Taking transformations of the data has been an important part of statistical practice for many years. A major contribution to this methodology was made by Box and Cox (1964), who proposed a parametric power family of transformations that includes the logarithm and the identity. They suggested that the power transformation, when applied to the dependent variable in a linear regression model, might induce normality and homoscedasticity. The transformation methodology has been quite successful and a large literature exists on this topic for parametric models. See Carroll and Ruppert (1988) and Sakia (1992) and the references therein.

Consider a semiparametric transformation model of the form

$$\Lambda_{\theta_0}(Y) = m(X) + \epsilon \quad ,$$

where Y is a univariate dependent variable, X is a 1-dimensional covariate, and ϵ is independent of X and has mean zero. We assume that $\{\Lambda_\theta : \theta \in \Theta\}$ is a parametric family of strictly increasing functions, while m is the unknown regression function. This model has been extensively studied by Linton, Sperlich and Van Keilegom (2008). They proposed two estimation methods for the unknown true parameter vector θ_0 : a profile likelihood method and a mean squared distance from independence-method. Here, we will use the profile likelihood estimator denoted by $\widehat{\theta}$, since it has been shown that it outperforms the other estimator.

The goal is to develop a test for the parametric form of the regression function m. We like to test the hypothesis

$$H_0 : m \in \mathcal{M} \qquad\qquad H_1 : m \notin \mathcal{M} \quad ,$$

where $\mathcal{M} = \{m_\beta : \beta \in \mathcal{B}\}$ is some parametric class of regression functions and $\mathcal{B} \subset \mathbb{R}^p$. The two test statistics that we'll use are the Kolmogorov-Smirnov type statistic and the Cramer-von Mises type statistic, where the basic idea is to compare the distribution function of ϵ estimated in a nonparametric way to the distribution function of ϵ estimated under the null hypothesis :

$$T_{KS} = n^{1/2} \sup_{y \in \mathbb{R}} |\widehat{F}_\epsilon(y) - \widehat{F}_{\epsilon_0}(y)|$$

and

$$T_{CM} = n \int (\widehat{F}_\epsilon(y) - \widehat{F}_{\epsilon_0}(y))^2 \, d\widehat{F}_\epsilon(y) \quad ,$$

where

$$\widehat{F}_\epsilon(y) = n^{-1} \sum_{i=1}^{n} I(\widehat{\epsilon}_i \leq y) \quad ,$$

$\widehat{\epsilon}_i = \Lambda_{\widehat{\theta}}(Y_i) - \widehat{m}(X_i, \widehat{\theta})$ are the nonparametric residuals, and $\widehat{m}$ is the Nadaraya-Watson estimator of the function m. Moreover,

$$\widehat{F}_{\epsilon_0}(y) = n^{-1} \sum_{i=1}^{n} I(\widehat{\epsilon}_{i0} \leq y) \quad ,$$

where $\widehat{\epsilon}_{i0} = \Lambda_{\widehat{\theta}}(Y_i) - m_{\widehat{\beta}}(X_i, \widehat{\theta})$ are the residuals estimated under H_0 and $\widehat{\beta}$ is a minimizer over $\beta \in \mathcal{B}$ of the expression $n^{-1} \sum_{i=1}^{n} (\Lambda_{\widehat{\theta}}(Y_i) - m_\beta(X_i))^2$.

We prove that the empirical process $n^{1/2}(\widehat{F}_\epsilon(y) - \widehat{F}_{\epsilon_0}(y))$, $y \in \mathbb{R}$, converge to $f_\epsilon(y)W$ where W has a normal distribution with zero mean and finite variance. Hence we can deduce the asymptotic distributions of the two test statistics under H_0 and under a local alternative. We see that these limiting distributions depend on the normal random variable W.

References

- Akritas, M.G., Van Keilegom, I. (2001), *Non-parametric Estimation of the Residual Distribution*, Scandinavian Journal of Statistics, Volume 28, Number 3, 549–568.

- Bickel, P.J. and Doksum, K. (1981), *An analysis of transformations revisited*, Journal of the American Statistical Association, Volume 76, 296–311.

- Box, G.E.P. and Cox, D.R. (1964), *An analysis of transformations*, Journal of the Royal Statistical Society - Series B, Volume 26, Number 2, 211–252.

- Carroll, R.J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman and Hall, New-York.

- Chen, X., Linton, O., Van Keilegom, I. (2003), *Estimation of Semiparametric Models when the Criterion Function is not Smooth*, Econometrica, Volume 71, Number 5, 1591–1608.

- Escanciano, J.C., Pardo-Fernandez, J.C., Van Keilegom, I. (2012), *A nonparametric test for risk-return relationships.*

- Hardle, W. and Mammen, E. (1993), *Comparing nonparametric versus parametric regression fits*, The Annals of Statistics, Volume 21, Number 4, 1926–1947.

- Linton, O., Sperlich, S., Van Keilegom, I. (2008), *Estimation of a Semiparametric Transformation Model*, Annals of Statistics, Volume 36, Number 2, 686–718.

- Sakia, R.M. (1992), *The Box-Cox transformation technique : a review*, The Statistician, Volume 41, 169–178.

- Samb, R., Heuchenne, C., Van Keilegom, I. (2012), *Estimating the Residual Distribution in Semiparametric Transformation Models.*

- Van Keilegom, I., Gonzalez Manteiga, W., Sanchez Sellero, C. (2008), *Goodness of fit tests in parametric regression based on the estimation of the error distribution*, TEST, Volume 17, 401–415.

- Zellner, A. and Revankar, N.S. (1969), *Generalized production functions*, The Review of Economic Studies, Volume 36, Number 2, 241–250.

The Simex method for correcting the bias in a survival cure model with mismeasured covariates

AURÉLIE BERTRAND (aurélie.bertrand@uclouvain.be)

In traditional survival analysis, all subjects in the population are assumed to be susceptible to the event of interest, that is, every subject has either already experienced the event or will experience it in the future. In many situations, however, it may happen that a fraction of individuals (long-term survivors) will never experience the event, that is, they are considered to be event free. The promotion time cure model is one of the survival models taking this feature into account.

We consider the case where the explanatory variables in the model are supposed to be subject to measurement error. This occurs e.g. when the instrument used to measure the variable (blood pressure, cholesterol level, etc.) has some calibration error. This measurement error should be taken into account in the estimation of the model, to avoid biased estimators of the model. Several approaches exist in the literature that correct for the presence of measurement error: they have been applied in survival models without a cure fraction, in the promotion time cure model (a corrected score approach has been proposed by Ma and Yin, 2008), or in other fields (the SIMEX algorithm).

We extend the SIMEX approach to the promotion time cure model. We show via simulations that the suggested method performs well in practice by comparing it with the method proposed by Ma and Yin (2008), which is, as far as we know, the only paper that has studied this problem before in the literature.

References

[1] Ma, Y. and Yin, G. (2008), Cure rate model with mismeasured covariates under transformation. *Journal of the American Statistical Association*, 103, 743–756.

Estimation of the latent distribution in cure survival models using a flexible Cox Model

VINCENT BREMHORST (vincent.bremhorst@uclouvain.be)

A common hypothesis in the analysis of survival data is that any observed unit will experience the monitored event if it is observed for a sufficient long time. Alternatively, one can explicitly acknowledge that an unknown and unidentified proportion of the patient population under study is cured and will never experience the event of interest. The promotion time model, which is motivated using biological mechanisms in the development of cancer, is one of the survival models taking this feature into account. The promotion time model assumes that each subject is exposed to N carcogenic cells. Given this number of carcogenic cells, we define latent event times $(Y_1, ..., Y_N)$, which are independent with a common distribution $F(t) = 1 - S(t)$ and can be seen as incubation time. Since we assume that 1 out of N latent factors need to be activated, the observed failure time is defined as the minimum of the latent event times.
In this work, we estimate the latent distribution $F(t)$ using a flexible Cox proportional hazard model where the logarithm of the baseline hazard function is specified using Bayesian P-splines. The identification issues of the related model are also investigated.

References

[1] Chen, M.-H., Ibrahim, J.G. and Sinha, D. (1999), "A New Bayesian Model for Survival Data with a Surviving Fraction." *Journal of the American Statistical Association.* 94, 909-919.

[2] Cooner, F., Banerjee, S., Carlin, B.P. and Sinha, D. (2007), "Flexible cure rate modelling under latent activation schemes" *Journal of the American Statistical Association,* 102, 360-72.

[3] Eilers, P. H. C. and Marx, B. D. (1996) "Flexible smoothing with B-splines and penalties (with discussion)" *Statistical Sciences,* 11, 89-121.

[4] Lang, S., Brezger, A., 2004. "Bayesian P-splines." *J. Comput. Graphical Statist.*, 13, 183-212.

The copula-graphic estimator in censored nonparametric location-scale regression models

Aleksandar Sujica (aleksandar.sujica@uclouvain.be)

A common assumption when working with randomly right censored data, is the independence between the variable of interest Y (the "survival time") and the censoring variable C. This assumption, which is not testable, is however unrealistic in certain situations. In this paper we assume that for a given X, the dependence between the variables Y and C is described via a known copula. Additionally we assume that Y is the response variable of a heteroscedastic regression model $Y = m(X) + \sigma(X)\varepsilon$, where the explanatory variable X is independent of the error term ε, and the functions m and σ are 'smooth'. We propose an estimator of the conditional distribution of Y given X under this model, and show the asymptotic normality of this estimator. When estimating from right censored data there is always an area in the right tail where the survival function can not be estimated consistently. In our approach we are forced to estimate that area for pre-estimation of the conditional survival function. We also study the small sample performance of the estimator, and discuss the advantages/drawbacks of this estimator with respect to competing estimators.