Calidad de Revistas Científicas Españolas
FECYT | 2013

ISSN: 1696-2281
eISSN: 2013–8830
SORT 38 (1) January-June 2014

SORT

Statistics and Operations Research Transactions

Coediting institutions

Universitat Politècnica de Catalunya
Universitat de Barcelona
Universitat de Girona
Universitat Autònoma de Barcelona
Universitat Pompeu Fabra
Universitat de Lleida
Institut d'Estadística de Catalunya

Supporting institutions

Spanish Region of the International Biometric Society
Societat Catalana d'Estadística

Generalitat de Catalunya
Institut d'Estadística de Catalunya

SORT

Volume 38
Number 1
January-June 2014
ISSN: 1696-2281
eISSN: 2013-8830

Editor's report

Articles

Editor's Report 2014

It is a pleasure to present a new issue of SORT–Statistics and Operations Research Transactions. I am sure that readers will enjoy all the articles in this number, as much as they have enjoyed previous issues.

SORT–Statistics and Operations Research Transactions was indexed in the Web of Science on January 1st, 2007. Since then, the journal has become more popular among scientists and has gained reputation. Thomson Reuters regularly publishes impact factors of its indexed journals. In 2014, our journal will receive its best figure for impact factor[1], as it would be stated in the Journal Citation Reports, much above the average for journals of similar origin. SORT holds an excellent position compared to the past and has a very prominent rank as an academic journal in the field of Probability & Statistics.

I am enormously satisfied to see how our journal has grown up and how it succeeds every day. SORT–Statistics and Operations Research Transactions has evolved over the past from a local journal to a global publication, something that is reflected by the increasing number of submissions and by the number of authors from all over the world.

SORT is now characterized by all the ingredients for excellence.

- The FECYT seal of maximum quality for publication provided by the Spanish Ministry. This distinction is awarded every three years to top publications in recognition of their high standard.
- Participation of an increasing number of Catalan universities as co-editors of SORT who have now joined the original founders.
- An increasing number of submissions (88 in 2013).
- A reasonable review time of four months on average.
- Timely publication of two issues per year, where accepted articles are published within less than 12 months and are available online soon after acceptance.
- An electronic system for managing manuscripts, which was established in 2013 and is now fully functional.

I would like to warmly thank the editors, reviewers, authors and readers of the journal for their cooperation.

I do wish that the future will be as challenging and, at the same time, as rewarding as these past eight years have been for me as Chief Editor of SORT.

Montserrat Guillén
Department of Econometrics
Riskcenter-IREA
Universitat de Barcelona

1. Unofficial impact factor in May 2014: 0,857.

SORT 38 (1) January-June 2014, 3-12

Improving parametric Clarke and Wright algorithms by means of iterative empirically adjusted greedy heuristics*

Albert Corominas[1], Alberto García-Villoria[1,*] and Rafael Pastor[1]

Abstract

Since Clarke and Wright proposed their well-known savings algorithm for solving the Capacitated Vehicle Routing Problem, several enhancements to the original savings formula have been recently proposed, in the form of parameterisations. In this paper we first propose to use Empirically Adjusted Greedy Heuristics to run these parameterized heuristics and we also consider the addition of new parameters. This approach is shown to improve the savings algorithms proposed in the literature. Moreover, we propose a new procedure which leads to even better solutions, based on what we call Iterative Empirically Adjusted Greedy Heuristics.

MSC: 90C27 (Combinatorial Optimisation).

Keywords: EAGH-1, greedy heuristics, Clarke and Wright savings algorithm, CVRP.

1. Introduction

The Capacitated Vehicle Routing Problem (CVRP) is a well known variant of the NP-hard Vehicle Routing Problem (VRP). Heuristic methods have been proposed to solve it. Among them, the Clarke and Wright savings heuristic (*CW*) (Clarke and Wright, 1964) is one of the most popular: *CW* is simple, easy to implement, very fast and obtains quite good solutions (Altinel and Öncan, 2005).

* Supported by the Spanish MICINN project ENE2010-15509 and co-financed by ERDF

* *Corresponding author:* Alberto García-Villoria, e-mail: alberto.garcia-villoria@upc.edu

[1] Institute of Industrial and Control Engineering (IOC), Av. Diagonal, 647 (Edif. ETSEIB), 11th floor, 08028 Barcelona, Spain. E-mail: {albert.corominas/alberto.garcia-villoria/rafael.pastor}@upc.edu

Received: October 2012
Accepted: February 2014

CW is based on merging tours according to a *savings formula*, which refers to the distance or cost saved by the merged route instead of having the two original ones. Since *CW* was introduced, several enhancements to the original savings formula have been proposed (Gaskell, 1967; Yellow, 1970; Paessens, 1988; and recently Altinel and Öncan, 2005; Doyuran and Çatay, 2011). The best results achieved with a savings heuristic have been obtained by the Doyran and Çatay's *CW* version (let it be called *CW-DC*) and by the Altinel and Öncan's *CW* version (let it be called *CW-AO*).

All aforementioned enhanced savings formulae are parameterized. The performance of these parametric savings heuristics depends on the parameter values. A common practice in the literature is to set the parameter values and then run the algorithms using these values, which remain fixed during the run (Adenso-Díaz and Laguna, 2006). This approach is followed by Battarra et al. (2008) and Corominas et al. (2010) to set the parameter values of *CW-AO*. However, this approach has the drawback of needing a sufficiently representative set of training instances to calibrate the algorithm so the values of the parameters are suitable for any instance of the problem. On the other hand, to solve an instance, Altinel and Öncan (2005), and Doyran and Çatay (2011) run *CW-AO* and *CW-DC*, respectively, with a fixed set of 8,820 different combinations of parameter values and select the best combination. This other approach has the advantage that the parameters of the heuristics are specifically tuned for each instance and the drawback that more computing time is needed to solve each instance.

In this paper we follow and focus on the second approach. The objective is to improve the results obtained in the literature, by means of: i) considering to add new parameters, ii) using a more sophisticated procedure in the literature to obtain suitable parameter values according to the particular instance to be solved, which is known as EAGH-1 (Empirically Adjusted Greedy Heuristics-1) (Corominas, 2005), and iii) designing a new procedure based on EAGH-1 that we call Iterative EAGH-1 (IEAGH-1). IEAGH-1 always ensures solutions better than or equal to the solutions obtained by EAGH-1.

The remainder of the paper is organised as follows. Section 2 defines the CVRP and describes the parametric algorithms based on *CW* proposed in the literature. Section 3 explains EAGH-1 and proposes the new IEAGH-1 procedure. Section 4 presents the results of a computational experiment. Finally, the conclusions are given in Section 5.

2. Parametric Clarke and Wright algorithms to solve the CVRP

The VRP consists in finding the set of routes that minimises the total routing cost. The routes are designed for a fleet of vehicles that has to serve a set of customers with positive demand from one or several depots, subject to various constraints. The CVRP is a VRP variant in which all the vehicles have the same capacity, C, and there is only one depot. The formulation of the CVRP is as follows. Let $G = (V, E)$ be an undirected graph. V is the set of nodes $\{0, 1, \ldots, n\}$, in which node 0 represents the depot and the other n nodes represent the customers, and E is the set of edges. Each vehicle has a

capacity C and each customer i has a demand $0 < d_i \leq C$ $(i = 1,\ldots,n)$. For each edge $(i,j) \in E$ $(i = 0,\ldots,n; j = 0,\ldots,n)$, a traversing cost c_{ij} is associated. The objective is to find the routes that minimise the total cost so that each customer is visited only once and each route starts and ends at the depot.

Clarke and Wright (1964) proposed a greedy heuristic (*CW*) which solves the CVRP. Initially, this heuristic considers n routes to visit all customers, where each route includes only one customer. Next, at each iteration the two routes that can be feasibly merged with the largest *saving* are chosen to be merged. The saving $s_{ij} = c_{i0} + c_{0j} - c_{ij}$ is the cost saved by the merge of the routes $(0,\ldots,i,0)$ and $(0,j,\ldots,0)$.

Based on CW saving formula, Gaskell (1967) and Yellow (1970) proposed the parameterised saving expression $s_{ij} = c_{i0} + c_{0j} - \lambda c_{ij}$, where λ is a parameter that weights the relative importance of c_{ij}. Paessens (1988) proposed $s_{ij} = c_{i0} + c_{0j} - \lambda c_{ij} + \mu\,|c_{0i} - c_{j0}|$, where μ is another weight parameter. Later, Altinel and Öncan (2005) included the customer demands in their *CW* enhancement (*CW-AO*): $s_{ij} = c_{i0} + c_{0j} - \lambda c_{ij} + \mu\,|c_{0i} - c_{j0}| + \upsilon\frac{d_i+d_j}{\bar{d}}$, where $\bar{d}$ is the average demand and υ is a new weight parameter. To set the specific values of the three parameters when solving every instance, the authors vary the values of λ between 0.1 and 2, and the values of μ and υ between 0 and 2, using a step size equal to 0.1. So, *CW-AO* is executed 8,820 times with all combinations of the parameter values and the best solution obtained is chosen. This method improves on the original *CW* but requires much more computing time.

Lately, Doyuran and Çatay (2011) proposed the following savings formula:

$$s_{ij} = \frac{c_{i0} + c_{0j} - \lambda c_{ij}}{c^{\max}} + \mu\frac{\cos\theta_{ij}\left|c^{\max} - (c_{0i} - c_{j0})/2\right|}{c^{\max}} + \upsilon\frac{\left|\bar{d} - (d_i + d_j)/2\right|}{d^{\max}},$$

where θ_{ij} is the angular distance between the customers i and j with respect to the depot, c^{max} is the longest distance among all customer pairs and $d^{\max}$ is the maximum demand among all customers. This savings function has the same 3 parameters that the Altinel and Öncan function, and Doyuran and Çatay also propose to run *CW-DC* 8,820 times with the combinations of the parameter values. But the variation of the parameter values differs: λ between 0.1 and 2, μ between 0 and 2, and υ between -1 and 1, using a step size equal to 0.1.

3. EAGH-1 and IEAGH-1 for the CVRP

3.1. EAGH-1

The main characteristic of a greedy heuristic is that at each iteration an irreversible decision is taken according to an index value associated with each possible decision. When two or more indices are considered in a greedy heuristic, infinite mix-indices

can be generated by linear combination and, therefore, an infinite set H of heuristics can be obtained. Regarding savings algorithms, the infinite set of heuristics H can be defined by the following function h which depends on the set of attributes of the decision ($a_{ij} = \{c_{i0}, c_{0j}, c_{ij}, c_{0i}, c_{j0}, d_i, d_j, \bar{d}\}$) and a set of parameters ($\Pi = \{\lambda, \mu, \upsilon\}$) that weight the elementary indices: $h(a_{ij}, \Pi) = c_{i0} + c_{0j} - \lambda c_{ij} + \mu |c_{0i} - c_{j0}| + \upsilon \frac{d_i + d_j}{\bar{d}}$.

The EAGH-1 procedure (Corominas, 2005) seeks the best heuristic function $h \in H$ for a given instance according to the objective function $f(X_\Pi)$ to be minimised, where X_Π is the solution obtained by $h(a_{ij}, \Pi)$. In the case of the CVRP, f is the total route cost. Note that to find the best heuristic function $h(a_{ij}, \Pi)$ is equivalent to find the Π parameter values that minimise f. Because the function f is not expected to have any special or recognisable property, only a direct optimisation algorithm (i.e. an optimisation algorithm that only uses the values of the function) may be used. Corominas (2005) proposes to apply the Nelder and Mead algorithm (N&M), also known as the flexible polyhedron algorithm (Corominas et al., 2010), as the optimisation algorithm for minimising f.

N&M is based on $q+1$ points ($q = |\Pi|$) that are the vertices of a q-dimensional simplex: $V_0, V_1, \ldots, V_i, \ldots, V_q$. The coordinates of the vertices represent the parameter values (Π), and each vertex is evaluated using f, i.e., the value of the solution (according to f) obtained when the instance is solved using $h(a_{ij}, \Pi)$ and Π is defined by the coordinates of the vertex. At each iteration of N&M, the vertices of the simplex are moved over the q-dimensional space according to their evaluations. N&M starts from an initial regular simplex. To build it, we have to provide N&M with an initial vertex (V_0) and the length of the edges of the initial simplex (δ). Thus, the N&M parameters are f, V_0 and δ. A more detailed description of N&M is provided in the Appendix.

We propose to solve the CVRP using four procedures based on EAGH-1, and the *CW-AO* and *CW-DC* saving formulae. First define the following two sets of decision attributes: $a_{ij}^{AO} = \{c_{i0}, c_{0j}, c_{ij}, c_{0i}, c_{j0}, d_i, d_j, \bar{d}\}$ and $a_{ij}^{DC} = \{c_{i0}, c_{0j}, c_{ij}, c_{0i}, c_{j0}, d_i, d_j, \bar{d}, \theta_{ij}, c^{\max}, d^{\max}\}$, and define the following three sets of parameters: $\Pi_3 = \{\lambda, \mu, \upsilon\}$, $\Pi_6 = \{\lambda, \mu, \upsilon, \beta_1, \beta_2, \beta_3\}$ and $\Pi_8 = \{\lambda, \mu, \upsilon, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$. Finally, define the following four sets of heuristics:

$$h_3^{AO}(a_{ij}^{AO}, \Pi_3) = c_{i0} + c_{0j} - \lambda c_{ij} + \mu |c_{0i} - c_{j0}| + \upsilon \frac{d_i + d_j}{\bar{d}}$$

$$h_8^{AO}(a_{ij}^{AO}, \Pi_8) = c_{i0}^{\beta_1} + c_{0j}^{\beta_2} - \lambda c_{ij}^{\beta_3} + \mu |c_{0i} - c_{j0}|^{\beta_4} + \upsilon \left(\frac{d_i + d_j}{\bar{d}}\right)^{\beta_5}$$

$$h_3^{DC}(a_{ij}^{DC}, \Pi_3) = \frac{c_{i0} + c_{0j} - \lambda c_{ij}}{c^{\max}} + \mu \frac{\cos\theta_{ij} \left| c^{\max} - \frac{(c_{0i} - c_{j0})}{2} \right|}{c^{\max}} + \upsilon \frac{\left| \bar{d} - \frac{(d_i + d_j)}{2} \right|}{d^{\max}}$$

$$h_6^{DC}(a_{ij}^{DC}, \Pi_6) = \left(\frac{c_{i0} + c_{0j} - \lambda c_{ij}}{c^{\max}}\right)^{\beta_1} + \mu \left(\frac{\cos\theta_{ij} \left| c^{\max} - \frac{(c_{0i} - c_{j0})}{2} \right|}{c^{\max}} \right)^{\beta_2} + \upsilon \left(\frac{\left| \bar{d} - \frac{(d_i + d_j)}{2} \right|}{d^{\max}} \right)^{\beta_3}$$

The four procedures based on EAGH-1 that we propose, $EAGH-1_3^{AO}$, $EAGH-1_3^{AO}$, $EAGH-1_3^{DC}$ and $EAGH-1_6^{DC}$, consist in applying EAGH-1 to the set of heuristics defined by h_3^{AO}, h_8^{AO}, h_3^{DC} and h_6^{DC}, respectively. Regarding the initial vertex V_0, we use a point that corresponds to the original Clarke and Wright algorithm: $\lambda=1$, $\mu=0$, $\upsilon=0$, $\beta_1=\beta_2=\ldots=1$. Regarding the initial length of the edges δ, since different solutions may be obtained according to its value, 20 δ values between 0.25 and 5 with a step size equal to 0.25 have been considered. Thus, $EAGH-1_3^{AO}$, $EAGH-1_3^{AO}$, $EAGH-1_3^{DC}$ and $EAGH-1_6^{DC}$ are run using the 20 δ values and the best found solution is returned.

Since the sets of heuristics defined by h_3^{AO} and h_3^{DC} are subsets of the sets defined by h_8^{AO} and h_6^{DC}, respectively, we expect that better solutions are found by $EAGH-1_3^{AO}$ and $EAGH-1_6^{DC}$, although we also expected that the computing time will be larger since the best heuristic will be sought in a larger search space.

3.2. IEAGH-1

We propose a new procedure that we call Iterative EAGH-1 (IEAGH-1). It takes advantage of the two following properties of the optimisation algorithm that is used by EAGH-1 (the N&M algorithm): i) it is recommended to use a good starting point V_0, and ii) N&M ensures that the best set of parameters found Π^* is always better than or equal to V_0 (i.e, $f(X_{\Pi^*}) \leq f(X_{V_0})$).

IEAGH-1 is based on applying iteratively EAGH-1. It first apply EAGH-1. Let $\prod_0^*$ be the set of parameters found. Then, at each iteration *it* ($it=1,2,\ldots$), it applies EAGH-1 using the initial vertex $\prod_{it-1}^*$, where $\prod_{it}^*$ is the set of parameters found by EAHG-1 at the iteration *it*. The stop condition of IEAGH-1 is that no improvement is achieved, that is, $f(X_{\Pi_{it-1}^*}) = f(X_{\Pi_{it}^*})$. Note that IEAGH-1 always ensures solutions better than or equal to EAGH-1 but the computing time will be, of course, larger.

Analogously to the EAGH-1 procedures proposed for the CVRP, we propose four IEAGH-1 procedures for solving this problem: $IEAGH-1_3^{AO}$, $IEAGH-1_8^{AO}$, $IEAGH-1_3^{DC}$ and $IEAGH-1_6^{DC}$. First, $EAGH-1_3^{AO}$, $EAGH-1_3^{AO}$, $EAGH-1_3^{DC}$ and $EAGH-1_6^{DC}$ are run, respectively, using the 20 aforementioned δ values. Then, at each, iteration, EAGH-1 is run using only 5 δ values around the δ value that returned the best solution at previous iteration. That is, if at iteration $it-1$ the best solution was found using a δ value equal to δ^*, then the δ values $\delta^*-0.50$, $\delta^*-0.25$, δ^*, $\delta^*+0.25$, $\delta^*+0.50$ are used at the iteration *it*.

Table 1: Average percentage solution improvement over CW (average computing times, in seconds).

	Global	P	A	B	CE	CMT
Altinel and Öncan	2.97 (6.96)	4.47 (3.96)	2.44 (4.69)	2.10 (4.79)	3.26 (7.11)	2.88 (32.15)
Doyuran and Çatay	3.16 (7.43)	5.20 (4.85)	2.47 (5.14)	2.10 (5.24)	3.22 (7.21)	2.86 (31.80)
$EAGH - 1_3^{AO}$	3.22 (5.73)	4.98 (3.74)	2.62 (4.08)	2.21 (4.34)	3.34 (5.65)	3.18 (23.00)
$EAGH - 1_3^{AO}$	3.40 (37.73)	5.39 (24.19)	2.69 (26.90)	2.35 (26.91)	3.43 (37.11)	3.31 (158.31)
$EAGH - 1_3^{DC}$	3.17 (6.53)	5.04 (4.26)	2.54 (4.72)	2.20 (5.10)	3.12 (6.44)	2.97 (25.45)
$EAGH - 1_6^{DC}$	3.46 (24.25)	5.47 (15.42)	2.85 (17.42)	2.33 (18.96)	3.53 (23.58)	3.18 (96.47)
$IEAGH - 1_3^{AO}$	3.26 (12.60)	5.03 (8.30)	2.64 (9.16)	2.27 (9.84)	3.34 (12.16)	3.20 (48.99)
$IEAGH - 1_8^{AO}$	3.51 (88.85)	5.40 (51.15)	2.92 (64.49)	2.39 (64.03)	3.55 (83.34)	3.46 (389.21)
$IEAGH - 1_3^{DC}$	3.29 (22.71)	5.15 (14.04)	2.63 (16.77)	2.35 (18.22)	3.29 (21.56)	3.13 (88.90)
$IEAGH - 1_6^{DC}$	3.51 (30.83)	5.47 (18.71)	2.89 (21.63)	2.42 (24.36)	3.53 (30.18)	3.27 (126.46)

4. Computational experiment

The test instances used in the computational experiment are the same instances used in Altinel and Öncan (2005), which were also used in Doyuran and Çatay (2011). Namely, 72 instances of Augerat et al. (1995) grouped in three sets (22 in set P, 27 in set A and 23 in set B), 8 instances of Christofides and Eilon (1969) (set CE) and 7 instances of Christofides et al. (1979) (set CMT).

The algorithms were coded in Java and the computational experiment was carried out using a 1.17 GHz Intel Core i7 with 3.0 GB of RAM. Because the Altinel and Öncan (2005) and Doyuran and Çatay (2011) experiments were carried out on computers different from ours, we coded and ran again their experiments on our computer so the comparison of the computational times is fair. We found a slight variability in the numerical results when we rerun their experiments. This phenomenon in savings algorithms have been reported, for instance, in Laporte et al. (2000) and in Doyuran and Çatay (2011). The reason that may cause this difference is the computer code (for instance, in savings

algorithms it is not specified how to break the ties between pair of costumers with equal savings). For the sake of consistency, all results reported here are the ones found with our code.

As it is done in the literature, we consider the following two criteria when evaluating the procedures: the average percentage solution improvement over the original *CW* and the average computing time. Table 1 shows the obtained results, in which the grey rows indicate dominated procedures (worse solutions, on average, are obtained in equal or higher computing time or equal solutions, on average, are obtained in higher computing time).

We can see in Table 1 that the results of Altinel and Öncan (2005) and Doyuran and Çatay (2011) are dominated by $EAGH-1_3^{AO}$ and $EAGH-1_3^{DC}$. Moreover, $EAGH-1_3^{DC}$ is dominated by $EAGH-1_3^{AO}$. Therefore, we compare $EAGH-1_3^{AO}$ with them. Specifically, the average improvement obtained with $EAGH-1_3^{AO}$ over *CW* is 7.76%, 1.86% and 1.55% better than the improvement obtained in Altinel and Öncan (2005), Doyuran and Çatay (2011) and by $EAGH-1_3^{DC}$, respectively, and the average computing time is 1.21, 1.30 and 1.14 times smaller, respectively. Other existing dominated procedures are $EAGH-1_3^{AO}$ (by $EAGH-1_6^{DC}$) and $IEAGH-1_8^{AO}$ (by $IEAGH-1_6^{DC}$). On average, the best solutions obtained with EAGH are those that correspond to $EAGH-1_6^{DC}$, whose improvement over *CW* is 14.16%, 8.67% and 6.94% better than the improvement achieved by Altinel and Öncan, Doyuran and Çatay and $EAGH-1_3^{AO}$, respectively.

As we expected, all IEAGH-1 procedures obtain better solutions than their respective EAGH-1 procedures, although the computing times are larger. The average improvements over *CW* obtained with $IEAGH-1_3^{AO}$, $IEAGH-1_8^{AO}$, $IEAGH-1_3^{DC}$ and $IEAGH-1_6^{DC}$ are 1.23%, 3.13%, 3.65% and 1.42% better than the improvements obtained with $EAGH-1_3^{AO}$, $EAGH-1_3^{AO}$, $EAGH-1_3^{DC}$ and $EAGH-1_6^{DC}$, respectively, and the average computing times are 2.20, 2.35, 3.48 and 1.27 times larger, respectively. The best solutions are obtained with $IEAGH-1_6^{DC}$, which dominates $IEAGH-1_8^{AO}$ (they obtain the same average improvement but the computing time of $IEAGH-1_8^{AO}$ is larger). Specifically, the $IEAGH-1_6^{DC}$ improvement over *CW* is 15.38% and 9.97% better than the improvement achieved by Altinel and Öncan, and Doyuran and Çatay, respectively.

5. Conclusions

In this paper we have improved the resolution of the CRVP with parametric Clarke and Wright savings heuristics. To achieve this objective, we have considered the addition of new parameters in the parameterized savings formula.

We first propose to use EAGH-1 (Corominas, 2005) and the computational experiment shows that better solutions, on average, can be found with less computing time. Specifically, the $EAGH-1_3^{AO}$ improvement over *CW* is, on average, 1.86% better than

Doyuran and Çatay's improvement whereas their average computing times are 5.73 s and 7.43 s, respectively. With some more computing time (24.25 s), $EAGH-1_6^{DC}$ is able to obtain an improvement 8.67% better than the Doyuran and Çatay's procedure improvement.

Moreover, we propose a new procedure based on EAGH-1 that we call Iterative EAGH-1 (IEAGH-1). It is shown that the solutions are improved with respect to the ones obtained by EAGH-1 at the expense of a larger computing time. The best results are obtained by $IEAGH-1_6^{DC}$, which slightly improves $EAGH-1_6^{DC}$ but the average computing time is gone up by 6.58 s.

Although we have proposed IEAGH-1 to solve the CRVP, this procedure can be also applied to solve other combinatorial optimisation problems.

APPENDIX

The N&M algorithm is a direct search method for minimising $f(x)$ where $f:\mathbb{R}^q\rightarrow\mathbb{R}$ is the objective function and q the dimension. It is based on $q+1$ points that are the vertices of a simplex in the q-dimensional space: $x_1,x_2,\ldots,x_{q+1}$. N&M starts from an initial simplex (usually regular) and iteratively moves the vertices over the q-dimensional space according to their objective function values until the differences between the values of the vertices are small enough and the simplex is small enough.

At each iteration of N&M, the vertices of the simplex are labelled and ordered such that $f(x_1)\leq f(x_2)\leq\ldots\leq f(x_{q+1})$. In the case of ties, the oldest vertex has priority. Let $x_r=\bar{x}+\alpha(\bar{x}-x_{q+1})$ be the reflection of x_{q+1}, where $\bar{x}$ is the centroid of the q best vertices (i.e., $\bar{x}=\sum_{i=1}^{q}x_i/q$) and $\alpha>0$ is a parameter. Four cases are considered according to the $f(x_r)$ value:

1. *Expansion*. If $f(x_r)<f(x_1)$ then calculate $x_e=\bar{x}+\gamma(x_r-\bar{x})$, where $\gamma>1$ is a parameter. If $f(x_e)<f(x_1)$, replace x_{q+1} with x_e; otherwise, replace x_{q+1} with x_r.
2. *Reflection*. If $f(x_1)\leq f(x_r)<f(x_q)$ then replace x_{q+1} with x_r.
3. *Outside contraction*. If $f(x_q)\leq f(x_r)<f(x_{q+1})$ then calculate $x_{oc}=\bar{x}+\eta(x_r-\bar{x})$, where $0<\eta<1$ is a parameter. If $f(x_{oc})<f(x_r)$, replace x_{q+1} with x_{oc}; otherwise, replace x_{q+1} with x_r and shrink all vertices except x_1: $x_i=x_1+\delta(x_i-x_1)$ $i=2,\ldots,q+1$, where $0<\delta<1$ is a parameter.
4. *Inside contraction*. If $f(x_{q+1})\leq f(x_r)$ then calculate $x_{ic}=\bar{x}+\eta(x_{q+1}-\bar{x})$. If $f(x_{ic})<f(x_{q+1})$, replace x_{q+1} with x_{ic}; otherwise, shrink all vertices except x_1 as in 3).

The values of parameters α, γ, η and δ that we have adopted are 1, 2, 0.5 and 0.5, respectively, which have been almost always used in the literature (Lagarias et al., 1998).

For more details of N&M, see Nelder and Mead (1965) or Corominas et al. (2010).

References

Adenso-Díaz, B. and Laguna, M. (2006). Fine-tuning of algorithms using fractional experimental designs and local search. *Operations Research*, 54, 99–114.

Altinel, I. K. and Öncan, T. (2005). A new enhancement of the Clarke and Wright savings heuristic for the capacitated vehicle routing problem. *Journal of the Operational Research Society*, 56, 954–61.

Augerat, P., Belenguer, J. M., Benavent, E., Corberán, A., Naddef, D. and Rinaldi, G. (1995). Computational results with a Branch and Cut Code for the Capacitated Vehicle Routing Problem. *Technical report RR 949-M*, U. Joseph Fourier, France.

Battarra, M., Golden, B. and Vigo, D. (2008). Tuning a parametric Clarke-Wright heuristic via a genetic algorithm. *Journal of the Operational Research Society*, 59, 1568–72.

Christofides, N. and Eilon, S. (1969). An algorithm for the vehicle routing dispatching problem. *Operations Research Quarterly*, 20, 309–18.

Christofides, N., Mingozzi, A. and Toth, P. (1979). The Vehicle Routing Problem. *Combinatorial Optimization*, Christofides, N., Mingozzi, A., Toth, P. and Sandi, C. (Eds.), Wiley, Chichester, 318–38.

Clarke, G. and Wright, J. (1964). Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 12, 568–81.

Corominas, A. (2005). Empirically Adjusted Greedy Algoriths (EAGH): A new approach to solving combinatorial optimisation problems. *Working paper IOC-DT-P-2005-22*, Universitat Politècnica de Catalunya, Spain.

Corominas, A., García-Villoria, A. and Pastor, R. (2010). Fine-tuning a parametric Clarke and Wright heuristic by means of EAGH (empirically adjusted greedy heuristics). *Journal of the Operational Research Society*, 61, 1309–14.

Doyuran, T. and Çatay, B. (2011). A robust enhancement to the Clarke-Wright savings algorithm. *Journal of the Operational Research Society*, 62, 223–31.

Gaskell, T. J. (1967). Bases for vehicle fleet scheduling. *Operations Research Quarterly*, 18, 281–95.

Lagarias, J. C., Reeds, J. A., Wright, M. H. and Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9, 112–47.

Laporte, G., Gendreau, M., Potvin, J.and Semet, F. (2000). Classical and modern heuristics for the vehicle routing problem. *International Transactions in Operational Research*, 7, 285–300.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7, 308–13.

Paessens, H. (1988). The savings algorithm for the vehicle routing problem. *European Journal of Operational Research*, 34, 336–44.

Yellow, P. (1970). A computational modification to the savings method of vehicle scheduling. *Operations Research Quarterly*, 21, 281–83.

SORT 38 (1) January-June 2014, 13-42

Time-varying market beta: does the estimation methodology matter?

Belén Nieto[1], Susan Orbe[2] and Ainhoa Zarraga[3]

Abstract

This paper compares the performance of nine time-varying beta estimates taken from three different methodologies never previously compared: least-square estimators including nonparametric weights, GARCH-based estimators and Kalman filter estimators. The analysis is applied to the Mexican stock market (2003-2009) because of the high dispersion in betas. The comparison between estimators relies on their financial applications: asset pricing and portfolio management. Results show that Kalman filter estimators with random coefficients outperform the others in capturing both the time series of market risk and their cross-sectional relation with mean returns, while more volatile estimators are better for diversification purposes.

MSC: 62M10, 62J15, 62G08, 91G70.

Keywords: Time-varying beta, nonparametric estimator, GARCH-based beta estimator, Kalman filter.

1. Introduction

Precise estimates for market betas are crucial in many financial applications, including asset pricing, corporate finance and risk management. From a pricing perspective, the empirical failure of the unconditional Capital Asset Pricing Model (CAPM) has led to two possible ways of relaxing restrictive assumptions under the model being considered: the first is the use of an intertemporal framework, as in Merton (1973), that implies multiple sources of systematic risk. The ad-hoc three-factor model of

[1] Department of Finance. University of Alicante. Ap. 99. E-03080 Alicante. Spain. belen.nieto@ua.es

[2] Department of Applied Economics III, BETS. University of the Basque Country (UPV/EHU). Avda. Lehendakari Aguirre, 83. 48015 Bilbao. Spain. susan.orbe@ehu.es

[3] Department of Applied Economics III, BETS. University of the Basque Country (UPV/EHU). Avda. Lehendakari Aguirre, 83. 48015 Bilbao. Spain. ainhoa.zarraga@ehu.es

Received: November 2012
Accepted: September 2013

Fama and French (1993) and the four-factor model of Carhart (1997) are successful examples of multifactor models. The second is to eliminate the static context in the relationship between expected return and risk by allowing time variation in both factors and loadings. In that sense, Jagannathan and Wang (1996), Lettau and Ludvigson (2001) and Petkova and Zhang (2005) find that betas of assets with different characteristics move differently over the business cycle and Campbell and Vuolteenaho (2004), Fama and French (1997) and Ferson and Harvey (1999) show that time-variation in betas helps to explain anomalies such as value, industry and size. However, this conditional time-varying framework does not seem to be enough to improve the weak fit of the CAPM, as shown by Lewellen and Nagel (2006). The main problem in beta dynamics literature is that the investor's set of conditioning information is unobservable and consequently some assumptions have to be made. There are two main alternatives: making assumptions about the dynamics of the betas and making assumptions about the conditional covariance matrix of the returns.

For the first alternative, many different structures have been considered. There are studies that estimate the dynamics of betas by Kalman filter assuming standard stochastic processes such as random walk, autoregressive, mean reverting and switching models driving those dynamics. Some examples can be found in Wells (1994), Moonis and Shah (2003) and Mergner and Bulla (2008). Other studies use parametric approaches based on Shanken (1990), in which betas are modelled as a function of state variables or firm characteristics as in Jagannathan and Wang (1996) and in Lettau and Ludvigson (2001). A nonparametric version of this approach can be found in Ferreira, Gil and Orbe (2011). Betas have also been assumed as a function of time, with both linear and parabolic functional forms, as in Lin, Chen and Boot (1992) and Lin and Lin (2000). Nonetheless neither empirical estimation nor simulation results can produce a clear conclusion about the best way to model betas. If no parametric functions are specified and no additional conditions are assumed except that betas vary smoothly over time, then the seminal work of Fama and MacBeth (1973) suggests the use of a rolling window ordinary least squares (OLS) estimation of the market model. This data-driven approach has the advantage of no parameterization but requires prior selection of the window length. More recently, other estimators from the family of rolling least squares have been considered. In this sense, based on the nonparametric time-varying estimator proposed by Robinson (1989), time-varying conditional betas have been nonparametrically estimated by Esteban and Orbe (2010), Li and Yang (2011) and Ang and Kristensen (2012) assuming that betas vary smoothly over time and possibly nonlinearly. The flexibility of this nonparametric setting avoids the problem of misspecification derived from selecting a functional form but it also requires that window length be selected.

The second alternative, consisting of making assumptions about the conditional covariance matrix of the returns, relies on the parametric approach of ARCH-class models. In this context the assumptions under multivariate GARCH (MGARCH) models make it possible to estimate time-varying betas. In fact, the transmission of volatility between assets is captured by a time-varying conditional covariance matrix whose

elements are used to calculate the beta as a ratio of covariance to variance. As the conditional covariance matrix is time dependent, the beta obtained will also be time dependent. There has been a great proliferation of multivariate models with GARCH structures in the last few decades, see Bauwens, Laurent and Rombouts (2006) or Silvennoinen and Teräsvirta (2009) for a survey. Some examples of the use of MGARCH models to estimate time-varying betas can be found in Bollerslev, Engle and Wooldridge (1988), Ng (1991), De Santis and Gérard (1998) and more recently in Choudhry (2005) and Choudhry and Wu (2008), among others.

Given the wide variety of time-varying beta estimates, some papers compare different approaches. The most common comparison is between GARCH-based estimators and Kalman filter approaches. In general, results indicate that the latter class of estimators performs better in terms of forecasting ability (Faff, Hillier and Hillier (2000) and Choudhry and Wu (2008)). However, there is no agreement about the best process assumption for beta dynamics. Moreover, when the Kalman filter is compared with estimators in the class of least squares, as in Ebner and Neumann (2005), the latter outperform the former.

In this paper three different methodologies for estimating time varying betas are compared: least-squares-based estimators, including the well-known rolling window OLS and the nonparametric time-varying estimator proposed in Esteban and Orbe (2010), beta estimators based on GARCH processes for the conditional covariance matrix of returns, including also asymmetric versions, and dynamic beta estimators obtained by the Kalman filter method. The main theoretical difference between the OLS and nonparametric estimators is that the latter have guaranteed consistency if the bandwidth is optimally chosen. In practice, there is an advantage in using the nonparametric estimator since there are many data-driven window selection criteria while the OLS estimator uses the rule of the thumb. The GARCH-based beta estimator does not rely on a smoothness assumption but has the advantage of taking into account the potential conditional heteroscedasticity of the returns. Finally, the Kalman filter method, unlike the other estimators, imposes assumptions about the specific functional form of beta dynamics. To the best of our knowledge, this is the first paper to compare these different methodologies simultaneously for the specific estimation of market risk.

Specifically, the OLS, the nonparametric estimator with both a uniform and a Gaussian kernel, the bivariate BEKK (after Baba, Engle, Kraft and Kroner) and the bivariate dynamic conditional correlation (DCC) structures together with their corresponding asymmetric versions, and random walk and random coefficient structures for the dynamic of the betas under the Kalman filter estimation are considered. The analysis is applied to daily returns for the Mexican stock market between 2003 and 2009. This market is selected because of the high cross-sectional dispersion in the sensitivity of individual returns to market returns in terms of both level and variability. Thus, grouping stocks into portfolios on the basis of trading volume provides high dispersion in time series and cross-sectionally which allows the performance of the beta estimates covering very different patterns to be analysed. The sample period also contributes to the aim of

the paper because it includes the recent financial and economic crisis, ensuring enough time variation in betas potentially related, in this case, to the business cycle. Finally, the data frequency selection seeks to exploit the benefits of using high-frequency data in measuring systematic risk while avoiding problems of errors in variables that stem from nonsynchronous trading effects.

A second distinctive feature of the paper is the way in which the different estimates are compared. Instead of using only standard statistical measures based on the standard errors of the estimates or on the fit of the simple market model, the accuracy of the estimators is also determined by financial criteria. Specifically, the estimators are compared in terms of their usefulness for asset pricing or portfolio management purposes. On the one hand, the CAPM fit in both time series (pricing errors) and cross-sectional (risk premia) frameworks is analysed. On the other hand, the power in achieving the next period out of the sample minimum variance portfolio based on the use of each estimate is also compared.

Interesting results are found. The time-series analysis reveals that the wide time fluctuation combined with the moderate dispersion of the Kalman filter estimate assuming a random coefficient makes this the best beta estimator for reducing the adjustment errors in both the market model and the CAPM when daily frequency returns are used. At the same time, this estimator also produces a positive and significant risk premium in the cross-sectional estimation of the CAPM with monthly frequency data. This good fit between betas and mean returns is also obtained when the two nonparametric beta estimators are used. On the other hand, for the purpose of risk diversification, beta estimators with high volatility are more appropriate. The Kalman filter with random walk estimator and the GARCH-based beta estimators do a good job of estimating the composition of the portfolio with the minimum risk.

The rest of the paper is structured as follows. Section 2 presents the estimation methodologies. Section 3 describes the data. Section 4 compares beta estimates descriptively. Section 5 provides the empirical results for the comparison of the beta estimators in two frameworks: asset pricing and mean-variance portfolio analysis. Section 6 concludes and the Appendix contains the data information.

2. Methodology

The Capital Asset Pricing Model due to Sharpe (1964) and Lintner (1965) relates the expected return on an asset to its systematic market risk or beta. This beta is the sensitivity of the asset return to changes in the return on the market portfolio. That is, the beta is the slope of the market model:

$$R_{it} = \alpha_i + \beta_i R_{mt} + u_{it}, \qquad i = 1,\ldots,N, \quad t = 1,\ldots,T, \tag{1}$$

where R_{it} and R_{mt} are the return on asset or portfolio i and on the market portfolio at time t, respectively. Commonly, the unknown coefficients in (1) are estimated by OLS applied to the linear regression for each portfolio.

If it is assumed that these coefficients vary with time, model (1) must be rewritten as:

$$R_{it} = \alpha_{it} + \beta_{it} R_{mt} + u_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T \tag{2}$$

2.1. Least-squares-based time-varying beta estimators

As proposed by Fama and MacBeth (1973), one simple way to obtain time series estimates of betas is by a rolling OLS estimation of the market model. This consists of minimising a local sum of squared residuals for each portfolio i:

$$\min_{(\alpha_{it}, \beta_{it})} \sum_{j=t-1}^{t-r} (R_{ij} - \alpha_{it} - \beta_{it} R_{mj})^2, \tag{3}$$

where r indicates the amount of past observations to be considered at each estimation point. From the first order conditions of the optimisation problem (3), the rolling OLS estimator is obtained as:

$$(\hat{\alpha}_{it} \;\; \hat{\beta}_{it})'_{ROLL} = \left(\sum_{j=t-1}^{t-r} \mathbf{X}_j \mathbf{X}'_j \right)^{-1} \sum_{j=t-1}^{t-r} \mathbf{X}_j R_{ij}, \qquad i = 1, \ldots, N,$$

where $\mathbf{X}_j = (1 \;\; R_{mj})'$ is the jth observation of the data matrix, the subscript *ROLL* denotes the OLS rolling estimator and $'$ denotes matrix and vector transpose.

In the empirical application of this estimator, a window of 120 observations for data with daily frequency is used. The sampling frequency is selected based on the findings of Bollerslev and Zhang (2003) and Ghysels and Jacquier (2006), who show that high-frequency data result in a more effective measure of betas than the commonly used monthly returns. Since, in general, stocks in the Mexican market are not continuously traded, intraday data are discarded in order to avoid nonsynchronicity effects on beta estimates. Regarding the window length, an alternative number of observations was also considered but it did not alter the main conclusions of the paper.[1]

The nonparametric time-varying beta estimator can be considered within the family of rolling least-squares estimators. It relies on the assumption that the unknown time-varying coefficients, α_{it} and β_{it}, are smooth functions (linear or nonlinear) of the time

1. Specifically, windows of 90 and 400 days were analysed. Results are available upon request.

index. It is derived from minimising a smoothed sum of squared residuals for a given portfolio i and for a pre-selected smoothness degree h_i:

$$\min_{(\alpha_{it},\,\beta_{it})} \sum_{j=t-1}^{t-Th_i} K_{h_i,tj}(R_{ij} - \alpha_{it} - \beta_{it}R_{mj})^2,$$

where $K_{h_i,tj} = h_i^{-1}K((t/T - j/T)/h_i)$ is a weight function and $K(\cdot)$ is a symmetric second order kernel. The shape of this kernel determines how past observations are to be weighted. If a uniform kernel is used all past observations selected are equally weighted but if the Epanechnikov or Gaussian kernels are used, higher weights are given to those observations closer to the estimation time point and lower weights to those farther away in time. The parameter h_i is the bandwidth that controls the amount of smoothness imposed on the coefficients associated with the ith portfolio. Solving the first-order conditions, the estimator has the following expression:

$$(\hat{\alpha}_{it} \;\; \hat{\beta}_{it})'_{NP} = \left(\sum_{j=t-1}^{t-Th_i} K_{h_i,tj}\mathbf{X}_j\mathbf{X}'_j\right)^{-1} \sum_{j=t-1}^{t-Th_i} K_{h_i,tj}\mathbf{X}_jR_{ij}, \qquad i = 1,\ldots,N,$$

where all elements are already defined and the subscript NP indicates the nonparametric estimator.

Once the smoothness degree h_i is set, the estimator obtained is consistent with the standard rate of convergence in nonparametric settings and has a closed form. Since the role of the bandwidth is to determine the amount of smoothness imposed on the betas and therefore the number of relevant past observations to be taken into account when estimating those betas, it is crucial to select it adequately in advance. If the bandwidth is large, the sub-sample of significantly weighted observations is larger, that is, more past observations are considered relevant in each local estimation. This results in a time series of estimated betas with little variability due to the high degree of smoothness. But if the bandwidth is small the estimation sub-sample is narrowed and the estimated betas have more dispersion. Different bandwidths (h_i) are allowed for the portfolios in order to capture different possible variations and curvatures of the betas. In consequence, the sub-sample size used at any estimation time point is the same when estimating the betas for a given portfolio but may be different for betas from another portfolio.

In regard to the practical issues of choosing the kernel and the bandwidths, it is well known that all kernels are asymptotically equivalent but that this is not the case for the bandwidth value. An optimal bandwidth is such that it minimises an error criterion in order to reach a tradeoff between the squared bias and the variance of the beta estimator. In the context of conditional factor models Ang and Kristensen (2012) and Li and Yang (2011) propose a bandwidth selection criterion for two-sided kernels, considering symmetric sub-samples that take into account not only past observations but also future observations. In this paper, only past observations are taken into account for estimating

conditional betas and the data-driven method considered for selecting the bandwidths simultaneously is based on the proposal of Esteban and Orbe (2010), where the sum of squared residuals for all regressions is minimised together in order to take into account any possible relationships between portfolios.

Finally, note that this nonparametric estimator generalises the rolling OLS estimator since it can be derived as a particular case. If a uniform kernel that weights past observations equally is considered and $h_i = r/T$ is imposed instead of the smoothness degree being selected via a data-driven method, then the estimations obtained by the two estimators match.

2.2. The time-varying beta estimator based on multivariate GARCH models

The literature on financial econometric volatility has provided evidence of fluctuations and high persistence in conditional variance of asset returns and conditional covariance with the market return. Since market betas are ratios of estimated conditional covariances and variances, $\hat{\beta}_{it} = \widehat{cov}_t(R_i, R_m)/\widehat{var}_t(R_m)$, if these second moments are adequately estimated by an MGARCH, then betas are also expected to be accurate estimators.

The estimation procedure for MGARCH models involves maximising the following log-likelihood function for each portfolio i:

$$lnL(\boldsymbol{\theta}_i) = -\frac{1}{2}\sum_{t=1}^{T} ln|\mathbf{H}_{it}| - \frac{1}{2}\sum_{t=1}^{T} \mathbf{y}'_{it}\mathbf{H}_{it}^{-1}\mathbf{y}_{it},$$

where $\boldsymbol{\theta}_i$ is the vector of parameters to be estimated and $\mathbf{y}_{it} = (R_{it}\ R_{mt})'$ is the vector of dependent variables in the mean equation, expressed as $\mathbf{y}_{it} = \boldsymbol{\delta}_i + \boldsymbol{\varepsilon}_{it}$. $\boldsymbol{\delta}_i = (\delta_{i1}\ \delta_{i2})'$ is a bivariate vector of constants and $\boldsymbol{\varepsilon}_{it}$ is a bivariate vector given by $\boldsymbol{\varepsilon}_{it} = \boldsymbol{\mu}_{it}\mathbf{H}_{it}^{1/2}$, where $\boldsymbol{\mu}_{it}$ is a bidimensional i.i.d. normally distributed process with mean zero and identity covariance matrix. The specification of the conditional covariance matrix ($\mathbf{H}_{it}$) depends on the MGARCH structure considered.

This analysis considers two different MGARCH structures widely used in financial literature: BEKK and DCC. The former is the bivariate BEKK (1,1,1) due to Engle and Kroner (1995), which has the advantage that the positive-definite constraint of the conditional covariance matrix is guaranteed by construction. This matrix takes the form:

$$\mathbf{H}_{it} = \mathbf{C}'_i\mathbf{C}_i + \mathbf{A}'_i\boldsymbol{\varepsilon}_{it-1}\,\boldsymbol{\varepsilon}'_{it-1}\mathbf{A}_i + \mathbf{B}'_i\mathbf{H}_{it-1}\mathbf{B}_i, \tag{4}$$

where $\mathbf{C}_i$ is a (2×2) lower triangular coefficient matrix and $\mathbf{A}_i$ and $\mathbf{B}_i$ are (2×2) coefficient matrices. The latter, DCC, is the bivariate dynamic conditional correlation

specification proposed by Engle (2002), where the conditional covariance matrix is decomposed into time-varying correlations and conditional standard deviations, ie.:

$$\mathbf{H}_{it} = \mathbf{D}_{it}\mathbf{R}_{it}\mathbf{D}_{it},$$

where $\mathbf{D}_{it}$ is a (2×2) diagonal matrix containing the conditional standard deviation of each process ε_{it}, obtained from univariate GARCH(1,1) models, $\sigma_{it}^2 = \alpha_{i0} + \alpha_{i1}\varepsilon_{it-1}^2 + \alpha_{i2}\sigma_{it-1}^2$, and the conditional correlation matrix can be written as:

$$\mathbf{R}_{it} = diag(q_{i11,t}^{-1/2}\, q_{i22,t}^{-1/2})\, \mathbf{Q}_{it}\, diag(q_{i11,t}^{-1/2}\, q_{i22,t}^{-1/2})$$

The (2×2) matrix $\mathbf{Q}_{it} = (q_{ijk,t})$ is given by:

$$\mathbf{Q}_{it} = \mathbf{S}_i(1 - \phi_{i1} - \phi_{i2}) + \phi_{i1}(\boldsymbol{\mu}_{it-1}\, \boldsymbol{\mu}_{it-1}') + \phi_{i2}\mathbf{Q}_{it-1},$$

where $\mathbf{S}_i$ is the unconditional correlation matrix of $\boldsymbol{\mu}_{it}$.

The empirical evidence that negative shocks have a larger effect on the volatility of returns than positive shocks is also taken into account in this paper by the estimation of the asymmetric versions of the BEKK and DCC models, denoted by BEKK-A and DCC-A, respectively. In the case of the BEKK-A, the conditional covariance matrix is that of the BEKK model, equation (4), with the following term added:

$$\mathbf{E}_i'\boldsymbol{\nu}_{it-1}\, \boldsymbol{\nu}_{it-1}'\mathbf{E}_i,$$

where $\boldsymbol{\nu}_{it-1} = \boldsymbol{\varepsilon}_{it-1} \odot I_{\varepsilon-1}$, $\odot$ denotes the Hadamard product, $\mathbf{E}_i$ is a (2×2) coefficient matrix, and $I_{\varepsilon-1}$ is an indicator function which takes a value of one for negative residuals, ε_{t-1}, and zero otherwise. In the case of the DCC-A model, the term $e_i\varepsilon_{it-1}^2 I_{t-1}$, with e_i being a coefficient, is added to each of the univariate GARCH(1,1) models that govern the variance of ε_{it}.

Once the conditional covariance matrix is estimated, the time-varying GARCH based beta for portfolio i is calculated as $\hat{\boldsymbol{\beta}}_{it}^{l} = \hat{\mathbf{H}}_{i12t}^{l}/\hat{\mathbf{H}}_{i22t}^{l}$, where $\hat{\mathbf{H}}_{i12t}^{l}$ is the estimated conditional covariance between the ith portfolio returns and the market returns and $\hat{\mathbf{H}}_{i22t}^{l}$ is the estimated conditional variance of the market return for l = BEKK, DCC, BEKK-A, DCC-A conditional covariance matrix structures.

2.3. The Kalman filter time-varying beta estimator

The state-space representation of the market model as in equation (2) enables time-varying coefficients to be estimated through the Kalman filter. The measurement equation is the market model and the transition equations that complete the state-space repre-

sentation determine the changes in the coefficients over time. Therefore, some assumptions about the stochastic behaviour of the conditional betas are needed. Two of the most widely used characterisations of the dynamics of betas are used: the random walk (KF-RW) and the random coefficient (KF-RC).

Under the random walk assumption, betas vary smoothly so their current value is determined by their own previous value plus an error term: $\alpha_{it} = \alpha_{it-1} + \eta_{1it}$ and $\beta_{it} = \beta_{it-1} + \eta_{2it}$. Large variances in the error terms indicate that there is no persistence, so the current beta may be completely different from the previous one. As the variance of the error term of the transition equations decreases less variability is allowed and the betas become more stable. When the variances tend to zero constant betas are obtained.

In the random coefficient model, betas are assumed to vary randomly around a fixed value with some variance: $\alpha_{it} = \alpha_i + \eta_{1it}$ and $\beta_{it} = \beta_i + \eta_{2it}$. The smaller the variance in the error terms of the transition equations, the lower the variations in the betas are, and when the variances tend to zero constant betas are also obtained. As the variances increase more jumps are permitted. In contrast to the random walk, level shifts are not allowed.

The Kalman filter estimation method requires a distribution to be assumed for all stochastic terms in the measurement and transition equations. All errors $(u_{it}, \eta_{1it}, \eta_{2it})$ involved in the estimation process are assumed to be normally distributed with zero mean, to have constant variance and to be uncorrelated from one another. In order to overcome the practical issues of selecting initial parameters, the OLS estimate for each portfolio using the whole sample is chosen for the initial value of the coefficients. Moreover, large enough variances in the error terms in the transition equations are allowed.

3. Data

This analysis uses daily logarithms of returns on 42 stocks traded on the Mexican Stock Exchange between January 2, 2003 and December 31, 2009. The data series have been computed from close daily prices taking into account dividends and splits. The sample is selected on the basis of representative criteria in terms of both market capitalisation and trading volume. The sample basically coincides with the 35 firms included in the reference index, "Índice de Precios y Cotizaciones" (IPC, hereafter). As the composition of this market index is revised annually, this gives a total of 42 firms in the sample period. The proxy for the risk-free asset is the 28-day maturity Treasury Certificate (TC) and data for this proxy are collected from the Banco de Mexico.

To show the representativeness of the selected sample, the table in the Appendix provides the names of the firms selected, their industrial classifications and the percentage of the total trading volume in pesos on the Mexican Stock Exchange at the end of 2009 accounted for by each stock. At that time the market comprised stocks issued by 85 firms, five of which were non domestic companies. Although the sample only contains half of the firms extant, it accounts for 95% of the market in terms of trading volume in

pesos in 2009, as can be seen by adding the weights in the last column of the table in the Appendix.[2] Moreover, the firms selected represent all the different industrial categories.

The individual stocks are sorted and grouped into portfolios. Since the aim of this work is to analyse the appropriateness of alternative beta estimators, it is important that the portfolios in the final sample produce different beta patterns. In that sense, individual betas could be used for sorting and locating stocks in portfolios. However, this would imply, on the one hand, selecting a beta estimation methodology first to analyse the appropriateness of each estimator. On the other hand, in subsequent sections asset pricing tests are used for comparing beta estimators and the results would be subject to the concerns raised by Lewellen, Nagel and Shanken (2010). This is why stocks have been sorted by individual money trading volume. This sorting produces sufficiently different portfolio betas in terms of both level and volatility. The composition of the portfolios is updated monthly by using the volume in pesos of the total trades for each stock during the month and the return of the portfolio is computed daily as the equally weighted average of the returns on stocks in the portfolio. Thus, Portfolio 1 contains the least liquid stocks while the most frequently traded stocks are in Portfolio 6.[3]

Table 1 reports the summary statistics for the returns on the six portfolios, on the market index and on the risk-free asset covering the whole sample period. The mean and the standard deviation are expressed on an annual basis. The beta estimator for each portfolio and its standard error come from the OLS estimation of the market model using the full sample period. Finally, the last row reports the average in time and across stocks within each portfolio of the monthly trading volume in millions of pesos. As can be seen, major differences in trading volume are observed; Portfolio 6 concentrates a large part of the market trading and its stocks have a trading volume 70 times greater than those of Portfolio 1. These liquidity differences do not imply differences in portfolio return volatilities, since standard deviation is similar for all six portfolios, but curiously they do produce increasing mean returns ranging from 14% for Portfolio 1 to 29% for Portfolio 6. Thus, this market seems to not show an illiquidity premium. More importantly, betas are monotonously increasing from Portfolio 1 to Portfolio 6 and also have different levels of standard errors. Therefore, the portfolio formation criterion produces the desirable dispersion in portfolio betas. The distribution of the returns is negatively skewed for the risk-free asset and all portfolios except the fifth and the market index, for which the returns' distributions are symmetric at the 5% significance level. Regarding the kurtosis coefficient, there is a significant positive excess of kurtosis for all cases except for the risk-free asset, for which the coefficient is negative.

2. The same calculation using trading volumes for other years in the sample period gives similar percentages of representativeness.

3. The classification has also been drawn up using trading volume in terms of number of shares and the characteristics of the resulting portfolios are very similar.

Table 1: *Summary Statistics of Returns.*

	Port. 1	Port. 2	Port. 3	Port. 4	Port. 5	Port. 6	IPC	TC
Mean	0.1401	0.2062	0.3263	0.2053	0.2404	0.2914	0.2367	0.0498
Standard deviation	0.2634	0.2182	0.2370	0.2662	0.2579	0.2719	0.2310	0.0006
Skewness	−0.5419	−0.5357	−0.1317	−1.4389	−0.0451	−0.1264	0.1024	−0.2892
Excess Kurtosis	5.2145	3.6191	7.5339	29.5636	5.6640	5.0603	5.3426	−0.4961
Beta	0.6523	0.6949	0.7958	0.9027	0.9667	1.1059		
Standard error	0.0223	0.0152	0.0154	0.0171	0.0133	0.0096		
Volume (millions)	119.87	298.96	497.89	817.24	1682.89	8280.50		

4. Conditional beta estimates

In this section descriptive statistics regarding the nine time series beta estimates obtained by the three methodologies considered are presented and compared. Rolling window OLS is obtained with subsamples of 120 previous observations for all portfolios and denoted by ROLL. The nonparametric estimator uses two alternative kernels: the uniform (NP-U) and the Gaussian (NP-G). The selected bandwidth is 0.1279 for Portfolios 1, 2, 3 and 6 and 0.0896 for Portfolios 4 and 5 when the uniform kernel is used, while for the Gaussian kernel the selected bandwidth is 0.0591 for all portfolios except the fifth, for which is 0.0398. Therefore, although bandwidths are allowed to vary with portfolios, the data-driven values selected indicate that betas have the same smoothness degree for most portfolios and hence the number of relevant past observations is the same. The alternative GARCH specifications produce time series of beta estimates that are denoted as BEKK and DCC for the symmetric versions of the BEKK and DCC models, respectively, and BEKK-A and DCC-A for the corresponding asymmetric versions. This estimation method does not weight the observations according to their temporal neighbourhood but according to the conditional heteroscedasticity structure. Finally, the Kalman filter method is applied with the assumption that the betas follow two alternative stochastic processes: random walk (KF-RW) and random coefficient (KF-RC). In the GARCH and Kalman filter context the total sample information is used, so that series of 1764 daily betas are produced. However, in order to provide a homogeneous, comparable context, the sample of beta estimates is restricted to the period between 17th October, 2003 and 31st December, 2009, with a total of 1564 daily beta estimates for each estimator.

Table 2 presents the mean and the standard deviation of the time series of estimated betas for each portfolio and for all the options considered. The general conclusion is that all estimation methods produce conditional beta series that have very similar mean values, smaller than the point beta estimate from the full sample (see Table 1). If there is any point worth commenting on, it is that KF-RC produces slightly higher mean betas for four out of the six portfolios. Differences between estimates are in

standard deviations. The least volatile estimates are NP-U (for three portfolios) and KF-RC (for two portfolios, including the portfolio 1) while the most volatile estimates are KF-RW (for three portfolios) and BEKK (for two portfolios). The major difference in the volatility pattern of the two Kalman filter beta estimates is therefore noteworthy. Some differences are also found between the statistics of the symmetric and asymmetric GARCH estimates. For example, BEKK-A and DCC-A estimates have slightly higher means than the corresponding symmetric versions and BEKK-A estimates are less volatile than BEKK estimates while DCC-A estimates are more volatile than DCC ones.

The results in Table 2 are confirmed in Figure 1, which shows the time series beta estimates for the two extreme portfolios. Subfigures (a.1) and (a.2) compare ROLL and the two NP estimates, Subfigures (b.1) and (b.2) compare GARCH based estimates, and Subfigures (c.1) and (c.2) compare the beta estimates based on the Kalman filter methodology. All betas move around the same long term mean, the NP methods produce the smoothest betas and changes in the short term are much more pronounced in estimates from GARCH structures and from the Kalman filter method. Subfigures (c.1) and (c.2) show the high short term fluctuation of the estimate from the Kalman filter with random coefficient contrasting with the random walk assumption. However, the rank of this short term dispersion is lower for KF-RC than for KF-RW or for GARCH-based estimates, as the standard deviations in Table 2 indicate. In addition, independently of the estimation methodology, mean betas increase and standard deviations of betas decrease, almost monotonously, from the portfolio containing the least liquid stocks to the portfolio containing the most liquid stocks. Since the most liquid stocks have the highest correlation with the market index, a beta closer to one with lower variability is expected.

Table 2: *Summary Statistics of Beta Estimates.*

Portfolio	Statistic	ROLL	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
1	Mean	0.6431	0.6442	0.6436	0.6321	0.6386	0.6389	0.6479	0.6403	0.6486
	Std. Dv.	0.2154	0.1573	0.1733	0.2910	0.2275	0.2316	0.2453	0.2615	0.1386
2	Mean	0.6979	0.6872	0.6915	0.7115	0.7332	0.7170	0.7389	0.7113	0.7095
	Std. Dv.	0.1379	0.0929	0.1053	0.1566	0.1340	0.1151	0.1399	0.1716	0.0983
3	Mean	0.7543	0.7591	0.7594	0.7492	0.7837	0.7597	0.7909	0.7626	0.7792
	Std. Dv.	0.1179	0.0923	0.1017	0.1366	0.1207	0.1119	0.1272	0.1549	0.0975
4	Mean	0.8273	0.8288	0.8288	0.8058	0.8489	0.8181	0.8581	0.8229	0.8635
	Std. Dv.	0.1518	0.1477	0.1386	0.1643	0.1555	0.1064	0.1634	0.1711	0.1239
5	Mean	0.9180	0.9195	0.9195	0.9108	0.9299	0.9085	0.9285	0.9184	0.9544
	Std. Dv.	0.1125	0.1073	0.1130	0.1438	0.1233	0.1026	0.1200	0.1361	0.0850
6	Mean	1.0742	1.0752	1.0753	1.0722	1.0749	1.0779	1.0795	1.0720	1.1002
	Std. Dv.	0.0671	0.0582	0.0602	0.0688	0.0804	0.0700	0.0882	0.0865	0.0625

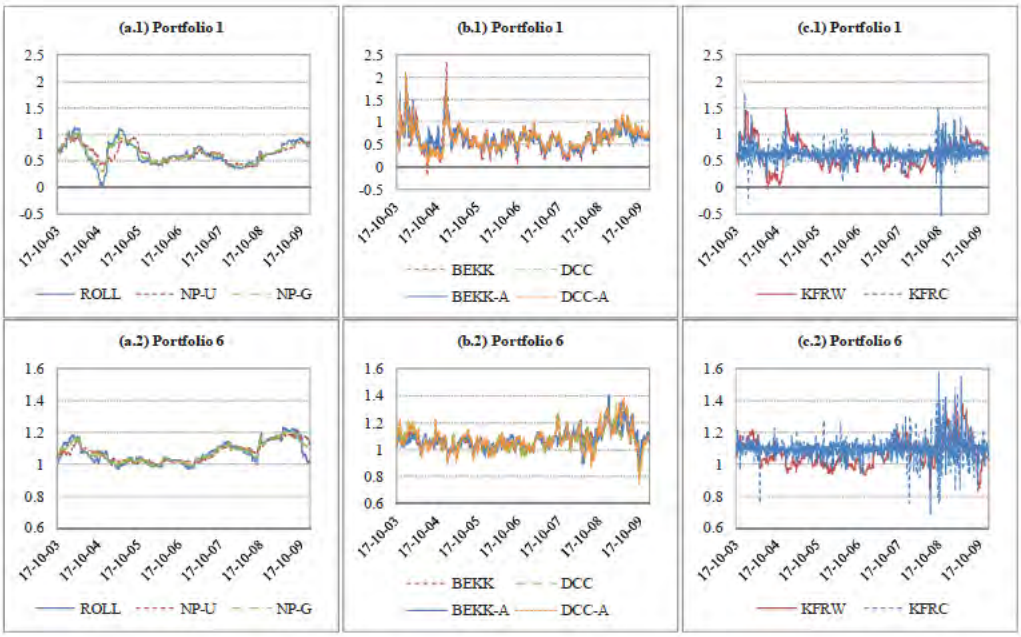

Figure 1: Beta Estimates from Alternative Methodologies.

A more formal comparison between the different beta estimates is carried out using the Kruskal-Wallis test. This is a non-parametric test based on ranked data that returns the p-value for the null hypothesis that two or more samples are drawn from the same population. For each portfolio, the Kruskall-Wallis test is applied to compare the different estimates all together on the one hand, and the estimates in each group on the other hand. The results are consistent for all six portfolios and indicate that the null is rejected when the nine estimates are compared simultaneously. The null is also rejected when the four GARCH based estimates are compared and when the two Kalman filter estimates are compared separately. Only for the group of least-squares-based estimates can the null not be rejected.[4]

In order to gain insight into the similarities of different time-varying beta estimates the correlations between pairs of conditional beta estimates are computed. Table 3 reports the correlations for each portfolio.[5] The results indicate that the pattern is very similar for beta estimates based on minimising some kind of least squares on the one hand, and for beta estimates from GARCH specifications on the other. However, the correlation between any of the estimated betas from each of these groups is much smaller. The different structures assumed for the beta dynamics in the Kalman filter

4. Results are available from the authors upon request.

5. Port.*i*/Port.*j* indicates that correlations for portfolio *j* are in the upper triangular panel while those for portfolio *i* are in the lower triangular panel.

method produce a lower correlation between the estimates. Moreover, the correlation between KF-RW beta estimates and those based on minimising least squares or GARCH structures are high, while the lowest correlations are those between KF-RC estimates and any other. This finding shows that the beta estimation method selected affects the resulting estimates.[6]

Table 3: *Correlations of Alternative Beta Estimates.*

	ROLL	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
Port. 2/Port. 1									
ROLL		0.7980	0.9628	0.4756	0.6081	0.4256	0.5783	0.7770	0.0712
NP-U	0.8213		0.8889	0.3434	0.4207	0.3194	0.4009	0.5228	0.0373
NP-G	0.9581	0.9130		0.5214	0.6395	0.4714	0.6098	0.7867	0.0694
BEKK	0.3530	0.2060	0.3670		0.9376	0.9186	0.9146	0.8242	0.1447
DCC	0.2889	0.1999	0.3073	0.6332		0.8476	0.9742	0.9118	0.1379
BEKK-A	0.3367	0.1932	0.3461	0.7234	0.8010		0.8580	0.7270	0.1568
DCC-A	0.4624	0.3214	0.4950	0.7171	0.8235	0.9069		0.8847	0.1341
KF-RW	0.7370	0.5139	0.7460	0.6349	0.4736	0.6158	0.7302		0.2548
KF-RC	0.0862	0.0345	0.0813	0.0950	0.0883	0.1223	0.1214	0.2773	
Port. 4/Port. 3									
ROLL		0.8323	0.9497	0.5399	0.5816	0.4781	0.5952	0.7330	0.0886
NP-U	0.9720		0.9279	0.4262	0.5287	0.4126	0.5713	0.6181	0.0760
NP-G	0.9686	0.9783		0.5716	0.6487	0.5174	0.6769	0.7781	0.1001
BEKK	0.5317	0.5251	0.5443		0.8092	0.9155	0.8032	0.8777	0.1427
DCC	0.5057	0.4960	0.5221	0.7944		0.7481	0.9119	0.8742	0.1244
BEKK-A	0.5329	0.5202	0.5444	0.8181	0.6872		0.7870	0.7905	0.1576
DCC-A	0.4746	0.4756	0.4981	0.7481	0.9626	0.6920		0.8397	0.1372
KF-RW	0.8193	0.7959	0.8279	0.7376	0.7652	0.6704	0.7181		0.2998
KF-RC	0.1254	0.1293	0.1265	0.1871	0.2005	0.1375	0.1858	0.3030	
Port. 6/Port. 5									
ROLL		0.9593	0.9481	0.3763	0.5037	0.1710	0.4689	0.6926	0.1094
NP-U	0.8479		0.9220	0.3526	0.4789	0.1486	0.4551	0.6634	0.1075
NP-G	0.9549	0.9481		0.4938	0.6416	0.2481	0.6130	0.8372	0.1196
BEKK	0.6086	0.5497	0.6526		0.9068	0.7574	0.8513	0.7485	0.1487
DCC	0.6181	0.5172	0.6485	0.9583		0.6078	0.9404	0.8867	0.1463
BEKK-A	0.5880	0.5646	0.6370	0.7525	0.7144		0.6077	0.4579	0.1575
DCC-A	0.5537	0.4307	0.5736	0.8740	0.9117	0.8211		0.8480	0.1523
KF-RW	0.7236	0.5766	0.7317	0.8807	0.9212	0.7320	0.8655		0.2908
KF-RC	0.1058	0.0942	0.1091	0.1352	0.1218	0.1421	0.1368	0.3031	

6. Similar results are obtained in Faff, Hillier and Hillier (2000) when comparing Kalman filter and GARCH-based beta estimators.

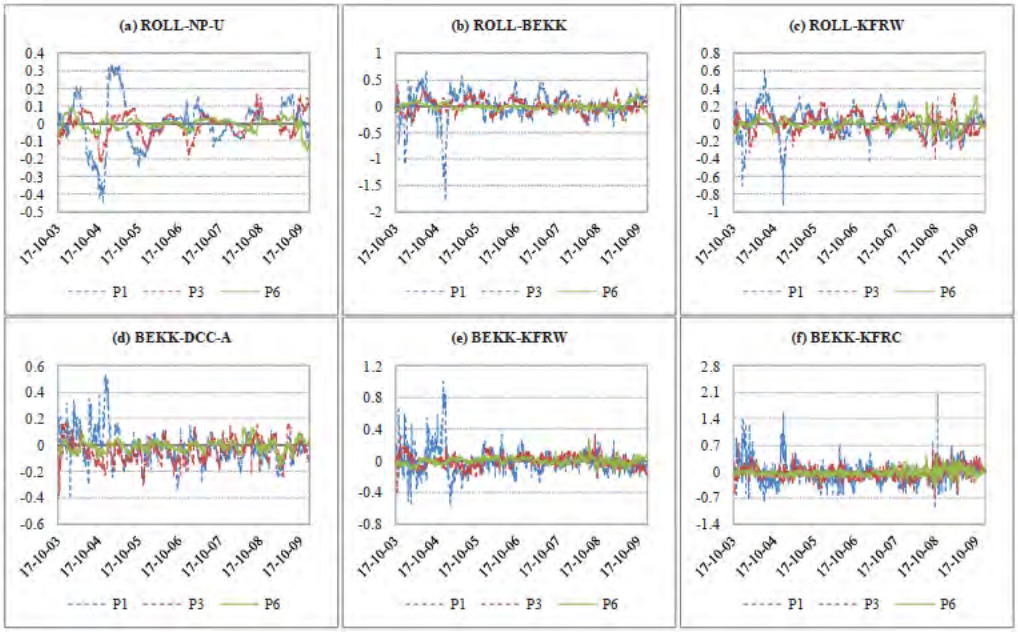

Figure 2: *Differences between Alternative Beta Estimates for Portfolios 1, 3 and 6.*

Figure 2 illustrates the high or low correlation between different estimates by showing the series of the differences between alternative pairs of beta estimates for Portfolios 1, 3 and 6.[7] As mentioned above, the volatility of beta estimates decreases and the mean increases from Portfolios 1 to 6 for all the estimation methodologies, so the largest differences are found between beta estimates in Portfolio 1. The most similar patterns correspond to the rolling and nonparametric estimates (Subfigure (a)) on the one hand, and to the BEKK and DCC-A (Subfigure (d)) on the other. However, major differences arise when beta estimates are obtained using methodologies based on different assumptions. For instance, Subfigure (f) shows that the largest difference is found when BEKK and KF-RC estimates are compared for Portfolio 1. These results are consistent with the correlation coefficients shown in Table 3; the higher the correlation between two beta estimate series the smaller the difference between them.

5. Beta estimator comparison

In this section the accuracy of the different estimators is compared in terms of the utility of time-varying beta estimates for two important financial applications: asset pricing and portfolio management.

7. The differences have been computed and plotted for all pairs of estimates and for all six portfolios. In order to save space, we only provide the most noteworthy cases.

5.1. The asset pricing perspective

This subsection analyses how systematic risk may be assessed more accurately through the use of one beta estimation methodology or another. For this purpose the simplest asset pricing framework is considered: the CAPM. It must be pointed out that this exercise does not set out to test the CAPM and that the analysis presented here could easily be extended to a multi-factor asset pricing model. However, this model offers a simple way of looking at the expected positive relationship between returns and systematic risk that any underlying investor's preferences would imply. In that sense, a beta estimate is more accurate if it is able to improve this relationship.

Next, two different settings for the comparison are considered. The first is based on time series analysis and the second on cross-section analysis.

5.1.1. Time series analysis

The first comparison between beta estimates relies on the appropriateness of the factor model representation. That is, for each portfolio the different beta estimates are compared in terms of fit for the market model. Since time-varying coefficients are estimated, R-squared statistics are not necessarily bounded and they cannot be comparable. Instead, unconditional variance ratios are studied as in Harvey, Solnik and Zhou (2002), among others. Specifically, the proportion of the unconditional variance of returns fitted by the market model, $VR1 = var(\hat{R}_i)/var(R_i)$, is used as a measure of goodness of fit, and the proportion of the unconditional variance of returns that the model fails to explain, $VR2 = var(\hat{u}_i)/var(R_i)$, as a measure of the estimation error. It must be pointed out that computing $\hat{R}_{it}$ and $\hat{u}_{it}$ requires estimates for parameter α_{it} and GARCH models do not provide them. In these cases, an estimation of α_{it} is obtained from the average of the market model where the time variation comes from each daily beta estimate:

$$\hat{\alpha}_{it}^{l} = \bar{R}_i - \hat{\beta}_{it}^{l}\bar{R}_m, \quad i = 1,\ldots,N, \quad t = 1,\ldots,T, \quad l = \text{BEKK, DCC, BEKK-A, DCC-A,}$$

where $\bar{R}_i$ and $\bar{R}_m$ are the mean returns on portfolio i and on the market portfolio, respectively.

Table 4 shows the values of the $VR1$ and $VR2$ criteria for each portfolio and each estimator. The results for the two measures are very similar when the ROLL and NP estimators are compared, since they are both based on the use of rolling least squares. In general, ROLL estimates show a larger fit (larger $VR1$) but also a larger estimation error (larger $VR2$). This could be due to the bandwidth sizes selected. Since the numbers of relevant past observations selected by the data-driven method for the nonparametric estimator are smaller than for the rolling OLS, the smoothness degree imposed is lower and in consequence the estimated betas have a smaller bias but a larger variance. However, all least squares-based methods produce mostly lower values for $VR1$ and similar or higher values for $VR2$ than the rest of estimates. In general terms, according

to the measure of variance explained, the market model is better adjusted when beta estimates come from GARCH structures (especially the asymmetric versions) and only slightly lower values are obtained when using the Kalman filter method with the random coefficient assumption. Moreover, the estimate that produces the lowest adjustment errors is clearly the Kalman filter with random coefficient for all portfolios. Therefore, it seems that the high daily fluctuation of the beta series from this estimation method benefits the time series adjustment of the market model.

Table 4: *Model Fit Criteria.*

Portfolio	Criteria	ROLL	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
1	VR1	0.3531	0.3207	0.3360	0.4107	0.3982	0.4208	0.4165	0.3900	0.4113
	VR2	0.6492	0.6542	0.6465	0.6380	0.6368	0.6310	0.6374	0.5456	0.4462
2	VR1	0.5616	0.5110	0.5329	0.6160	0.5832	0.6184	0.6166	0.5696	0.6068
	VR2	0.4464	0.4498	0.4437	0.4630	0.4592	0.4502	0.4539	0.3657	0.2923
3	VR1	0.6085	0.5730	0.5985	0.6370	0.6413	0.6332	0.6522	0.6392	0.6584
	VR2	0.3848	0.3883	0.3828	0.3725	0.3761	0.3682	0.3739	0.3163	0.2602
4	VR1	0.6383	0.6349	0.6238	0.6651	0.7241	0.6093	0.7293	0.6471	0.6592
	VR2	0.3758	0.3750	0.3740	0.3552	0.3689	0.3678	0.3726	0.3103	0.2334
5	VR1	0.7817	0.7734	0.7827	0.8059	0.7966	0.7634	0.7977	0.7734	0.7796
	VR2	0.2456	0.2451	0.2441	0.2401	0.2391	0.2366	0.2390	0.2060	0.1713
6	VR1	0.8850	0.8729	0.8836	0.8899	0.8947	0.8884	0.9038	0.8967	0.8995
	VR2	0.1109	0.1101	0.1103	0.1096	0.1111	0.1077	0.1101	0.0939	0.0739

The second comparison within this time-series framework employs Jensen's alpha as a measure of the error adjustment of the model: the difference between the observed return and the estimated return. Assuming the CAPM, the Jensen's alpha associated with each beta estimator is computed for each portfolio and period as:

$$\widehat{\alpha}_{it}^{J} = (R_{it} - R_{ft}) - \widehat{\beta}_{it}\,(R_{mt} - R_{ft}), \quad i = 1,\ldots,N, \quad t = 1,\ldots,T,$$

where R_{ft} represents the risk-free rate.

The quadratic sum of these alphas is calculated as a measure of the model misspecification, which allows a comparison to be made between different estimation methods. A large value of the quadratic sum of alphas indicates a poor specification of the model since the estimated returns differ greatly from the observed returns. Table 5 reports this measure. The bottom row shows the total sum of alphas for all portfolios. As expected, though with the exception of Portfolio 4, the quadratic sum of alphas decreases from Portfolio 1 to Portfolio 6 whatever the method used in the estimation of betas. Comparing the different estimates, the misspecification is similar for estimates based on least squares and on GARCH assumptions. The lowest values are obtained for Kalman fil-

ter methods and, specifically, for the case of random coefficient. The quadratic sum of alphas is considerably lower for KF-RC than for the other estimates and for all six portfolios. Consequently, the overall misspecification is also lower for this method.

Table 5: *Quadratic Sum of Jensen's Alphas.*

Portfolio	ROLL	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
1	0.2787	0.2801	0.2777	0.2740	0.2734	0.2707	0.2736	0.2478	0.1936
2	0.1381	0.1391	0.1375	0.1442	0.1428	0.1399	0.1410	0.1201	0.0920
3	0.1406	0.1415	0.1400	0.1370	0.1382	0.1354	0.1373	0.1229	0.0970
4	0.1775	0.1771	0.1774	0.1696	0.1763	0.1756	0.1779	0.1557	0.1124
5	0.1074	0.1074	0.1068	0.1063	0.1058	0.1047	0.1057	0.0957	0.0765
6	0.0542	0.0541	0.0540	0.0542	0.0549	0.0532	0.0544	0.0485	0.0369
Sum	0.8964	0.8993	0.8934	0.8854	0.8913	0.8795	0.8899	0.7907	0.6084

In order to learn whether the differences observed in Table 5 are relevant, a pairwise comparison of Jensen's alphas, in absolute values, associated with two beta estimators is conducted using the Wilcoxon signed rank test. Table 6 reports the median difference between the two series of alphas expressed on an annual basis. Each panel refers to a different portfolio and reports the median difference between the absolute value of alphas from the beta estimate indicated in the first column and the absolute value of alphas from the beta estimate indicated in the first row. For example, in the comparison of ROLL and NP-U for Portfolio 1, -0.0032 indicates that the pricing error is 0.32% lower when the ROLL beta estimate is used. Asterisks indicate that the null that this median difference is zero is rejected. Again, consistent with the evidence in Table 5, the results indicate that lower Jensen's alphas are obtained when betas are estimated by the Kalman filter for all six portfolios. And these time series errors are still lower when the random coefficient structure is assumed. Finally, although not for all the portfolios, some degree of relevance of the asymmetric BEKK estimator is shown when it is compared to the OLS or the non-parametric beta estimators.

Therefore, Tables 4, 5 and 6 provide a consistent conclusion: the lowest adjustment errors for both the market model and the CAPM are obtained when betas are estimated by the Kalman filter and a random coefficient model is assumed. It seems that the variability due to the random coefficient together with the dynamics incorporated into the estimation method are able to produce accurate beta estimates from the time series perspective.

5.1.2. Cross-sectional analysis

In this subsection the estimators are compared in terms of the market risk premium implied by the different estimated betas. Under rational expectations there should be a positive relationship between expected returns and systematic risk. For this purpose, the

Table 6: *Comparison of Jensen's Alphas in Absolute Values. Median Test.*

Port. 1	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
ROLL	−0.0032	−0.0001	0.0075	0.0115∗	0.0114∗	0.0087∗	0.0258∗	0.0813∗
NP-U		0.0046∗	0.0075∗	0.0113∗	0.0163∗	0.0066∗	0.0270∗	0.0899∗
NP-G			0.0045	0.0072∗	0.0106	0.0054	0.0220∗	0.0799∗
BEKK				0.0033∗	0.0015	−0.0006	0.0181∗	0.0634∗
DCC					−0.0021	−0.0022∗	0.0131∗	0.0613∗
BEKK-A						0.0010	0.0146∗	0.0623∗
DCC-A							0.0185∗	0.0561∗
KF-RW								0.0439∗
Port. 2	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
ROLL	−0.0013	0.0007	0.0055	0.0032	0.0076	0.0040	0.0202∗	0.0578∗
NP-U		0.0023∗	0.0074	0.0031	0.0057∗	0.0014	0.0207∗	0.0623∗
NP-G			0.0048	0.0037	0.0048	0.0024	0.0167∗	0.0580∗
BEKK				−0.0020	0.0015	−0.0036	0.0118∗	0.0524∗
DCC					0.0029	−0.0009	0.0173∗	0.0520∗
BEKK-A						−0.0045	0.0113∗	0.0522∗
DCC-A							0.0170∗	0.0467∗
KF-RW								0.0397∗
Port. 3	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
ROLL	0.0030	0.0014∗	0.0042∗	0.0020	0.0055∗	0.0025	0.0173∗	0.0683∗
NP-U		0.0002	0.0036	0.0022	0.0055∗	0.0008	0.0180∗	0.0642∗
NP-G			0.0021	0.0007	0.0049∗	−0.0013	0.0157∗	0.0598∗
BEKK				0.0005	0.0007	−0.0019	0.0121∗	0.0541∗
DCC					0.0003	0.0000	0.0141∗	0.0534∗
BEKK-A						−0.0026	0.0077∗	0.0627∗
DCC-A							0.0188∗	0.0544∗
KF-RW								0.0408∗
Port. 4	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
ROLL	0.0009∗	0.0012∗	−0.0009	−0.0016	0.0017	0.0005	0.0155∗	0.0546∗
NP-U		0.0005	−0.0005	0.0008	0.0017	0.0007	0.0139∗	0.0482∗
NP-G			−0.0006	−0.0031	0.0011	0.0009	0.0135∗	0.0480∗
BEKK				0.0021	0.0014	0.0008	0.0215∗	0.0503∗
DCC					0.0006	0.0004	0.0197∗	0.0490∗
BEKK-A						−0.0022	0.0096∗	0.0543∗
DCC-A							0.0183∗	0.0421∗
KF-RW								0.0306∗

Port. 5	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
ROLL	0.0003	0.0012∗	0.0030	0.0074	0.0083∗	0.0037	0.0179∗	0.0591∗
NP-U		0.0002	0.0000	0.0045	0.0085∗	0.0032	0.0165∗	0.0574∗
NP-G			0.0000	0.0049	0.0045∗	0.0035	0.0172∗	0.0571∗
BEKK				0.0011	0.0031	0.0000	0.0144∗	0.0521∗
DCC					0.0040	−0.0001	0.0132∗	0.0524∗
BEKK-A						−0.0030	0.0060∗	0.0575∗
DCC-A							0.0101∗	0.0534∗
KF-RW								0.0404∗
Port. 6	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
ROLL	0.0006	0.0000	−0.0037	−0.0026∗	−0.0009	−0.0031	0.0038∗	0.0359∗
NP-U		0.0002	−0.0014	−0.0025	0.0006	−0.0010	0.0036∗	0.0297∗
NP-G			−0.0018	−0.0023∗	−0.0008	−0.0018	0.0037∗	0.0331∗
BEKK				−0.0019∗	0.0002	−0.0005	0.0059∗	0.0339∗
DCC					0.0021∗	0.0005	0.0093∗	0.0346∗
BEKK-A						−0.0011∗	0.0036∗	0.0312∗
DCC-A							0.0086∗	0.0318∗
KF-RW								0.0264∗

simple CAPM framework is used, which assumes only one source of systematic risk: the market beta.

Using the Fama and MacBeth (1973) methodology, the following cross-sectional regression is estimated for each day in the sample period:

$$R_{it} - R_{ft} = \gamma_{0t} + \gamma_{1t}\hat{\beta}_{it} + e_{it}, \quad i = 1, \ldots, N, \tag{5}$$

where the beta represents one of the nine alternative estimates. A reasonable beta estimator should produce a positive and significant market risk premium and the more precise the above cross-sectional relationship is, the more accurate the beta estimator is. Additionally, since excess returns are used as dependent variable, an intercept statistically equal to zero indicates a good model fit.

The results from the Fama-MacBeth estimation of the model are presented in Table 7. This table reports the estimates of the intercept and the market risk premium ($\times 10^2$), their t-statistics for individual significance and the corresponding Shanken (1992) adjusted t-statistics. Asterisks indicate that the risk premium is significantly different from zero using both t-statistics at the 5% level. The left panel of the table shows the results when daily portfolio returns and betas are used in the estimation of (5) and one regression is run each day. The right panel provides the results when monthly returns and the beta estimator corresponding to the last day of the previous month are

Table 7: *Cross-Sectional Risk Premium Estimation.*

		Daily frequency		Monthly frequency	
		γ_0	γ_1	γ_0	γ_1
ROLL	Estimate	0.0101	0.0821	0.4088	1.6655
	t-stat.	0.181	1.251	0.320	1.502
	Adj. t-stat.	0.181	1.250	0.311	1.459
NP-U	Estimate	−0.0045	0.1012	−0.2203	2.3191∗
	t-stat.	−0.081	1.568	−0.181	2.210
	Adj. t-stat.	−0.081	1.566	−0.176	2.147
NP-G	Estimate	−0.0054	0.1045	−0.0950	2.2120∗
	t-stat.	−0.098	1.601	−0.075	1.988
	Adj. t-stat.	−0.098	1.599	−0.073	1.930
BEKK	Estimate	0.0329	0.0566	0.9704	1.1540
	t-stat.	0.615	0.891	0.847	1.196
	Adj. t-stat.	0.614	0.890	0.823	1.161
DCC	Estimate	0.0462	0.0436	0.4642	1.6817
	t-stat.	0.852	0.671	0.410	1.682
	Adj. t-stat.	0.851	0.670	0.398	1.634
BEKK-A	Estimate	−0.0495	0.1498∗	0.9704	1.1540
	t-stat.	−0.911	2.242	0.729	1.054
	Adj. t-stat.	−0.910	2.239	0.708	1.024
DCC-A	Estimate	0.0400	0.0512	1.2932	0.6088
	t-stat.	0.735	0.788	1.101	0.564
	Adj. t-stat.	0.734	0.787	1.069	0.548
KF-RW	Estimate	0.0283	0.0736	0.4793	1.6217
	t-stat.	0.571	1.097	0.429	1.782
	Adj. t-stat.	0.570	1.096	0.417	1.730
KF-RC	Estimate	0.0155	0.1181	−0.2193	2.3126∗
	t-stat.	0.314	1.528	−0.155	2.000
	Adj. t-stat.	0.314	1.526	−0.150	1.942

used to reduce the excessive noise that daily observations could introduce into this cross-sectional analysis. In this case, the number of regressions is 75, which is the number of months in the period analysed.

The intercepts are non-statistically different from zero and market risk premia are positive for all beta estimates and for the two data frequencies. However, differences in the value and significance of the risk premia are observed for different beta estimators. At daily frequency, market risk premia are not significant in general. Only for the beta estimated from the asymmetric BEKK method is there a relevant cross-sectional relationship between returns and market betas. The results for the monthly frequency are better and more conclusive. The risk premia associated with betas from GARCH structures are similar and not significant. The cross-sectional relationship clearly improves when non-parametric betas or Kalman filter betas are employed. The risk premium estimate and the t-statistic are very similar when the two NP beta estimators or the KF-RC one are used. In these three cases risk premia are significant at the 5% level.

Thus, the results of this analysis indicate that the estimation of the risk premium depends on the characteristics of the beta estimator. Specifically, the three estimators with the lowest standard deviations are the ones that produce significant risk premia in the relationship between betas and returns at monthly frequency. On the one hand, comparing the standard OLS estimator with the non-parametric estimates, the results suggest that a correct size of the window and the use of weights decaying in time matter with a view to better capturing this cross-sectional relationship. Therefore, an optimal mechanism for choosing the bandwidth is important. On the other hand, the high variability that the Kalman filter produces (but with lower dispersion than GARCH-based methods) is also a good characteristic for having betas more closely related to the cross-section of returns.

5.2. Portfolio management analysis

An important application of betas is their use in portfolio management. Since individual betas are part of the variance of a portfolio, the power of prediction of the different beta estimators can be studied by analysing whether the purpose indicated in the portfolio construction criterion is achieved in the next period.

For each of the estimation methodologies considered, betas for all six portfolios are taken in order to obtain an estimate of the next period covariance matrix, which can then be used to obtain the composition of the overall minimum variance portfolio. Thus, the beta estimators are compared by analysing the variance of the resulting portfolio.

Specifically, according to the market model, for a given month s the covariance matrix of a set of N asset returns is:

$$\mathbf{\Sigma}_s = \sigma^2_{ms}\mathbf{B}_s\mathbf{B}'_s + \mathbf{D}_s,$$

where σ^2_{ms} is the variance of the market return, $\mathbf{B}_s$ is an N-vector of individual betas and $\mathbf{D}_s$ is an $N \times N$ matrix of the idiosyncratic variance-covariances, all of them measured in month s. The variance of the market return is estimated using daily returns within month

s; beta estimates on the last day of month $s-1$ are used as predictors of elements of $\mathbf{B}_s$; and $\mathbf{D}_s$ is estimated as the residual covariance matrix from the market model consistent with these beta estimates employing daily returns within month s:

$$\widehat{\mathbf{D}}_s = \frac{1}{T_d}\widehat{\mathbf{U}}_s'\widehat{\mathbf{U}}_s,$$

Table 8: *Out-of-Sample Comparison for the Prediction of the Global Minimum Variance Portfolio.*

x/y	NP-U	NP-G	BEKK	DCC	BEKK-A	DCC-A	KF-RW	KF-RC
ROLL	48	70.7	76	77.3	70.7	82.7	96	44
	−0.963	2.166∗	13.736∗	12.074∗	7.343∗	12.968∗	21.758∗	−1.731
NP-U		73.3	82.7	90.7	73.3	88	96	49.3
		2.706∗	18.189∗	9.889∗	10.984∗	12.144∗	20.717∗	−0.409
NP-G			78.7	78.7	70.7	82.7	97.3	37.3
			10.860∗	8.420∗	7.666∗	10.333∗	19.290∗	−3.976∗
BEKK				42.7	33.3	45.3	62.7	17.3
				−1.384	−4.339∗	−1.467	3.160∗	−17.213∗
DCC					44	57.3	74.7	18.7
					−2.808∗	0.342	5.541∗	−15.797∗
BEKK-A						65.3	80	17.3
						3.353∗	8.141∗	−16.040∗
DCC-A							77.3	13.3
							4.362∗	−16.368∗
KF-RW								6.7
								−23.543∗

where $\widehat{\mathbf{U}}_s$ is a $T_d \times N$ matrix containing the residuals $\widehat{u}_{isd} = R_{isd} - \widehat{\alpha}_{is-1} - \widehat{\beta}_{is-1}R_{msd}$ for $i = 1,\ldots,N$, $d = 1,\ldots T_d$, where T_d is the number of days in month s and $s = 1,\ldots,S$ with S being the number of months in the sample.

The portfolio formation criterion consists of investing in the minimum variance portfolio, which implies choosing the portfolio weights ($\boldsymbol{\omega}_s$) that solve the following problem:

$$Min\ \boldsymbol{\omega}_s' \boldsymbol{\Sigma}_s \boldsymbol{\omega}_s$$
$$s.t. \quad \boldsymbol{\omega}_s' \mathbf{1}=1$$

This optimisation problem is solved for each month and each beta estimate, then the daily return of the minimum variance portfolio is computed for all the days in the month and its variance is recorded. The most successful beta estimator should lead to portfolios with the lowest variance.

Table 8 provides the results for the comparisons of pairs of series of the variance of the minimum variance portfolio conducted via the Wilcoxon median test. For each comparison x/y, the first number is the percentage of cases in which beta estimation method x produces higher variance than beta estimation method y. Below, the median difference ($\times 10^4$) is reported and an asterisk indicates that the null of equal medians is rejected at the 5% significance level.

The results are quite conclusive: the beta estimate that produces the lowest variance for the next period minimum variance portfolio is the Kalman filter with the random walk assumption. This is the case in the 96/97% of the out-of-sample predictions when it is compared to any least-squares-based estimates and in between 62% and 80% of the predictions when it is compared to GARCH-based estimates. It is also better than the other Kalman filter estimate in 6.7% of the predictions. Moreover, the difference between medians is larger in the cases when the KF-RW estimates is one of the beta estimates in the pair. By contrast, the beta estimated from the Kalman filter with the random coefficient produces the highest variance. On the other hand, GARCH beta estimators are superior to least-squares-based estimators for the purpose of risk hedging in portfolio decisions. Finally, when ROLL and NP estimators are compared the differences in the resulting variance portfolio are not so big but NP-G is significantly better than rolling OLS with both the standard selection of the window size and the optimal window size.

6. Conclusions

This paper compares the performances of three methodologies in estimating time-varying market betas: dynamic estimators based on least squares, time-varying estimators coming from GARCH structures for the conditional variance of the errors of the market model, and Kalman filter estimators. These three methodologies have never previously been compared with one another homogenously.

Specifically, three estimators in the group of least squares are selected: a rolling window OLS and two nonparametric estimators that use uniform and Gaussian kernels, respectively. The advantage of the nonparametric estimators is that they allow the optimal window length to be chosen. In the group of GARCH-based estimators standard DCC and BEKK models and their corresponding asymmetric versions are consid-

ered. In this case the potential benefits of taking into account the returns' conditional heteroscedasticity are examined. Finally, the Kalman filter estimator considering two different specifications for the transition equations is included in the comparison: one imposing a random walk process and the other assuming a random coefficient structure for the dynamics of the beta. Therefore, nine beta estimates are obtained for each of the six portfolios of daily returns for the Mexican stock market in the period 2003-2009. All the descriptive analyses indicate that the time pattern of these nine estimates are substantially different. The distribution of the estimates shows different sample moments for different estimates, especially regarding the standard deviation, and these differences are corroborated by an analysis of the correlations between them and by using the Kruskal-Wallis test.

The accuracy of the estimates is compared under two frameworks: an asset pricing perspective that assumes the CAPM and the mean-variance space for returns for portfolio management purposes. In the first case beta estimates are compared using different measures of the time-series fit of the model and looking at the cross-sectional relationship between mean returns and market betas. In the mean-variance context, the out-of-sample forecasting power of different beta estimates is obtained by comparing the results of the minimum variance portfolio.

The time-series analysis clearly concludes that the Kalman filter estimator that assumes a random coefficient is the best at reducing the adjustment errors in both the market model and the CAPM; moreover this is true for all six portfolios analysed. This estimate has the characteristic of presenting a very high time fluctuation, as GARCH-based estimates do, but a low standard deviation, as the smoothed nonparametric estimates do. This combination seems to be the reason for the good time series adjustment in the daily frequency sample used here.

The Kalman filter with the random coefficient estimate also produces a good fit for the CAPM cross-sectionally. In this case, this estimate and the two nonparametric estimates are the ones for which the relationship between betas and returns are positive and statistically significant. The high volatility in GARCH-based beta estimates has a negative effect on the stability of the relationship between systematic risk and mean returns. Consequently, in estimating the price of risk, dynamic methodologies that produce low dispersion are more appropriate for the prior estimation of systematic risk.

However, highly volatile market betas are appropriate in terms of risk diversification. The Kalman filter with a random walk estimate and the GARCH-based beta estimates are both better than estimates with lower volatility for estimating the composition of the portfolio with the minimum risk.

Given that different conclusions are obtained depending on whether betas or risk premia are estimated, one possible improvement along these lines could be to propose a new estimator that combines the advantages of these different estimators.

Acknowledgements

The authors thank Gonzalo Rubio and two anonymous referees for helpful comments and suggestions that substantially improved the contents of the paper. Financial support is acknowledged from Ministerio de Ciencia e Innovación under research grants ECO2012-35820, ECO2011-29268 and ECO2011-29751, from Generalitat Valenciana under the grant PROMETEO II/2013/015, and from Departamento de Educación, Universidades e Investigación del Gobierno Vasco under research grants IT-783-13 and IT-793-13.

Appendix: Individual stocks data information

Ticker	Firm Name	Sector	Trading Volume (Pesos %)
AMX-L	América Móvil	Telecomunications/Services	23.22
TELMEX-L	Teléfonos de Mexico	Telecomunications/Services	3.49
TELINT-L	Telmex Internacional	Telecomunications/Services	2.09
TELECOM-A1	Carso Global Telecom	Telecomunications/Services	1.89
AXTEL-CPO	Axtel	Telecomunications/Services	1.84
TLEVISA-CPO	Grupo Televisa	Telecomunications/Radio and Television	3.33
TVAZTCA-CPO	TV Azteca	Telecomunications/Radio and Television	1.07
ICH-B	Industrias CH	Materials/Steel	0.21
SIMEC-B	Grupo Simec	Materials/Steel	0.17
GMEXICO-B	Grupo Mexico	Materials/Metals and Mining	7.65
AUTLAND-B	Compañía minera Autland	Materials/Metals and Mining	0.12
CEMEX-CPO	Cemex	Materials/Construction	4.63
MEXCHEM	Mexichem	Materials/Chemical Products	0.93
ASUR-B	Grupo Aeroportuario del Sureste	Industrials/Transportation	0.87
GAP-B	Grupo Aeroportuario del Pacífico	Industrials/Transportation	0.50
OMA-B	Grupo Aeroportuario del Centro Norte	Industrials/Transportation	0.15
GEO-B	Corporación Geo	Industrials/Construction	1.73
URBI	Urbi Desarrollos Urbanos	Industrials/Construction	1.40
HOMEX	Desarrolladora Homex	Industrials/Construction	1.39
ICA	Empresas ICA	Industrials/Construction	1.33
IDEAL-B1	Impulsora del Desarrollo y el Empleo	Industrials/Construction	1.11
ARA	Consorcio Ara	Industrials/Construction	1.10
SARE-B	Sare Holding	Industrials/Construction	0.06
ALFA-A	Alfa	Industrials/Capital Goods	1.43
GCARSO-A1	Grupo Carso	Industrials/Capital Goods	1.02
LAB-B	Genomma Lab Internacional	Health/Medicine Distrib.	1.50
BOLSA-A	Bolsa Mexicana de Valores	Financial Services/Financial Markets	0.24
GFNORTE-O	Grupo Financiero Banorte	Financial Services/Financial Groups	2.04
GFINBUR-O	Grupo Financiero Inbursa	Financial Services/Financial Groups	1.07
COMPART-O	Banco Compartamos	Financial Services/Commercial Banks	0.79
WALMEX-V	Wal-Mart de Mexico	Consumer Staples/Hypermarkets	13.22
SORIANA-B	Organización Soriana	Consumer Staples/Hypermarkets	1.01
COMERCI-UBC	Controladora Comercial Mexicana	Consumer Staples/Hypermarkets	0.07
KIMBER-A	Kimberly-Clark Mexico	Consumer Staples/Household Products	1.06
BIMBO-A	Grupo Bimbo	Consumer Staples/Food	1.00
GRUMA-B	Gruma Sab de C.V.	Consumer Staples/Food	0.51
FEMSA-UBD	Fomento Económico Mexicano	Consumer Staples/Beverages	5.82
GMODELO-C	Grupo Modelo	Consumer Staples/Beverages	1.70
ARCA	Embotelladoras Arcas	Consumer Staples/Beverages	0.54
KOF-L	Coca-cola Femsa	Consumer Staples/Beverages	0.07
ELEKTRA	Grupo Elektra	Consumer Discret./Retails	1.28
GFAMSA-A	Grupo Famsa	Consumer Discret./Retails	0.50

References

Ang, A. and Kristensen, D. (2012). Testing conditional factor models. *Journal of Financial Economics*, 106, 132–156.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21, 79–109.

Bollerslev, T., Engle, R. and Wooldridge, J. (1998). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96, 116–131.

Bollerslev, T. and Zhang, B.Y.B. (2003). Measuring and modeling systematic risk in factor pricing models using high-frequency data. *Journal of Empirical Finance*, 10, 533–558.

Campbell, J. and Vuolteenaho, T. (2004). Bad beta, good beta. *American Economic Review*, 94, 1249–1275.

Carhart, M.M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52, 57–82.

Choudhry, T. (2005). Time-varying beta and the Asian financial crisis: Evidence from Malaysian and Taiwanese firms. *Pacific-Basin Finance Journal*, 13, 93–118.

Choudhry, T. and Wu, H. (2008). Forecasting ability of GARCH vs Kalman filter method: evidence from daily UK time-varying beta. *Journal of Forecasting*, 27, 670–689.

De Santis, G. and Gérard, B. (1998). How big is the premium for currency risk? *Journal of Financial Economics*, 49, 375–412.

Ebner, M. and Neumann, T. (2005). Time-varying betas of German stock returns. *Financial Markets and Portfolio Management*, 19, 29–46.

Engle, R. F. (2002). Dynamic conditional correlation-a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R. F. and Kroner, F. K. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11, 122–150.

Esteban, M. V. and Orbe, S. (2010). A nonparametric approach for estimating betas: the smoothed rolling estimator. *Applied Economics*, 42, 1269–1279.

Faff, R. W., Hillier, D. and Hillier, J. (2000). Time varying beta risk: an analysis of alternative modelling techniques. *Journal of Business Finance and Accounting*, 27, 523–554.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics*, 33, 3–56.

Fama, E. F. and French, K. R. (1997). Industry costs of equity. *Journal of Financial Economics*, 43, 153–193.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return and equilibrium: empirical tests. *Journal of Political Economy*, 81, 607–636.

Ferreira, E., Gil, J. and Orbe, S. (2011). Conditional beta pricing models: a nonparametric approach. *Journal of Banking and Finance*, 35, 3362–3382.

Ferson, W. E. and Harvey, C. R. (1999). Conditioning variables and the cross-section of stock returns. *Journal of Finance*, 54, 1325–1360.

Ghysels, E. and Jacquier, E. (2006). Market beta dynamics and portfolio efficiency. Working Paper, Cornell University, CREF.

Harvey, C. R., Solnik, B. and Zhou, G. (2002). What determines expected international asset returns? *Annals of Economics and Finance*, 3, 249–298.

Jagannathan, R. and Wang, Z. (1996). The conditional CAPM and cross-section of expected returns. *Journal of Finance*, 51, 3–53.

Lettau, M. and Ludvigson, S. (2001). Resurrecting the (C)CAPM: a cross-sectional test when risk premia are time-varying. *Journal of Political Economy*, 109, 1238–1287.

Lewellen, J. and Nagel, S. (2006). The conditional CAPM does not explain asset-pricing anomalies. *Journal of Financial Economics*, 82, 289–314.

Lewellen, J., Nagel, S. and Shanken, J. (2010). A skeptical appraisal of asset-pricing tests. *Journal of Financial Economics*, 96, 175–194.

Li, Y. and Yang, L. (2011). Testing conditional factor models: a nonparametric approach. *Journal of Empirical Finance*, 18, 975–992.

Lin, W. T., Chen, Y. H. and Boot, J. C. (1992). The dynamic and stochastic instability of betas: Implications for forecasting stock returns. *Journal of Forecasting*, 11, 517–541.

Lin, H. J. and Lin, W. T. (2000). A dynamic and stochastic beta and its implications in global capital markets. *International Finance*, 3, 123–160.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47, 261–290.

Mergner, S. and Bulla, J. (2008). Time-varying beta risk of Pan-European industry portfolios: a comparison of alternative modeling techniques. *European Journal of Finance*, 14, 771–802.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41, 867–887.

Moonis, S. A. and Shah, A. (2003). Testing for time-variation in beta in India. *Journal of Emerging Market Finance*, 2, 163–180.

Ng, L. (1991). Tests of the CAPM with time-varying covariances: a multivariate GARCH approach. *The Journal of Finance*, 46, 1507–1520.

Petkova, R. and Zhang, L. (2005). Is value riskier than growth? *Journal of Financial Economics*, 78, 187–202.

Robinson, P. (1989). Nonparametric estimation of time varying parameters. Hackl, P. ed. *Statistical Analysis and Forecasting Economic Structural Change*. New York: Springer-Verlag, 253–264.

Shanken, J. (1990). Intertemporal asset pricing: an empirical investigation. *Journal of Econometrics*, 45, 99–120.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies*, 5, 1–33.

Sharpe, C. (1964). Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance*, 19, 425–442.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. Andersen, T. G., Davis, R. A., Kreiβ, J. and Mikosch, T. eds. *Handbook of Financial Time Series*. Berlin: Springer-Verlag, 201–229.

Wells, C. (1994). Variable betas on the Stockholm exchange 1971–1989. *Applied Financial Economics*, 4, 75–92.

SORT 38 (1) January-June 2014, 43-52

Regression analysis using order statistics and their concomitants

Abdul Rasoul Ziaei[1], Ayyub Sheikhi[2] and Vahid Amirzadeh[2]

Abstract

In this work we derive the exact joint distribution of linear combinations of order statistics and linear combinations of their concomitants and some auxiliary variables in multivariate normal distribution. By extending the results of Sheikhi and Jamalizadeh we investigate some regression equations. Our results generalize those obtained in previous research by Viana, Lee and Loperfido.

MSC: 62G30, 62J05

Keywords: Unified skew normal, L-statistic, exchangeable multivariate normal distribution, order statistic, concomitant.

1. Introduction

The theory of order statistics and their concomitants plays an essential role in statistical inference. An excellent review of development in this field is available in David and Nagaraja (2003). There have been many studies with emphasis on distribution theory. Tsukibayashi (1998) found the moments and the joint distribution of an extreme value and its concomitant. Goel and Hall (1994) discussed the difference between concomitants and order statistics. Yang (1981) studied the linear functions of concomitants of order statistics. Arellano-Valle and Genton (2007) obtained the exact distribution of linear combinations of order statistics from dependent random variables. Viana and Lee (2006) studied the covariance structure of two random vectors $\mathbf{X}_{(n)}$ and $\mathbf{Y}_{[n]}$ in the presence of a random variable Z where $\mathbf{X}_{(n)} = (X_{(1)}, X_{(2)}, \ldots, X_{(n)})^{\mathsf{T}}$ is the vector of order statistics

[1] Islamic Azad University, Dashtestan Branch (IRN), ziaei85@yahoo.com

[2] Department of Statistics, Shahid Bahonar University, Kerman, Iran, sheikhy.a@uk.ac.ir, v_amirzadeh@uk.ac.ir

Received: January 2013

Accepted: November 2013

obtained from $\mathbf{X}_{n\times 1}$ and $\mathbf{Y}_{[n]} = (Y_{[1]}, Y_{[2]}, \ldots, Y_{[n]})^{\mathsf{T}}$ is the vector of concomitants. They also discussed some regression equations between order statistics, concomitants and the covariate variable Z, while Olkin and Viana (1995) studied the covariance structure and several regression models when $(X, Y_1, Y_2)^{\mathsf{T}}$ has a trivariate exchangeable normal distribution. Loperfido (2008) determined the joint distribution of an auxiliary variable X and the maximum of Y_1 and Y_2, i.e. $(X, Y_{(2)})^{\mathsf{T}}$. Sheikhi and Jamalizadeh (2011) found the joint distribution of two linear combinations of order statistics in the presence of a covariate random variable and presented some regression analyses.

We assume that the joint distribution of a covariate p-dimensional random vector $\mathbf{Z}$ and two n-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ follows a $2n+p$ dimensional multivariate normal vector with positive definite covariance matrix, i.e.

$$\begin{pmatrix} \mathbf{Z} \\ \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2n+p}\left(\mu = \begin{pmatrix} \mu_z \\ \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{pmatrix}, \sum = \begin{pmatrix} \Sigma_{\mathbf{zz}} & \Sigma_{\mathbf{xz}}^{\mathsf{T}} & \Sigma_{\mathbf{yz}}^{\mathsf{T}} \\ \Sigma_{\mathbf{xz}} & \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}}^{\mathsf{T}} \\ \Sigma_{\mathbf{yz}} & \Sigma_{\mathbf{xy}} & \Sigma_{\mathbf{yy}} \end{pmatrix}\right) \tag{1}$$

where $\mu_{\mathbf{x}}$, $\mu_{\mathbf{y}}$ and $\mu_{\mathbf{z}}$ are the mean vectors of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ respectively and $\Sigma_{\mathbf{uv}}$ is the covariance matrix of two random vectors $\mathbf{U}$ and $\mathbf{V}$. We assume that all of these covariance matrices are positive definite. The aim of this article is to derive the exact joint distribution of a linear combination of order statistics $(\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)})$ and a linear combination of their concomitants $(\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]})$ in the presence of a p-dimensional random vector $\mathbf{Z}$, where $\mathbf{a} = (a_1, a_2, \ldots, a_n)^{\mathsf{T}}$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)^{\mathsf{T}}$ are arbitrary vectors in $\mathbb{R}^n$. We show that the joint distribution of $\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}\right)^{\mathsf{T}}$ is a mixture of skew-normal distributions. Furthermore, we may explore some regression equations using order statistics, concomitants and covariate variables which generalizes those investigated in Viana and Lee (2006).

Following Arellano Valle and Azzalini (2006), we say that the random vector $\mathbf{X}$ follows a multivariate skew-normal distribution, denoted by $\mathbf{Y} \sim SUN_{d,m}(\xi, \delta, \Omega, \Gamma, \Lambda)$, if its density can be written as

$$f_{\mathbf{Y}}(\mathbf{y}) = \phi_d(\mathbf{y}; \xi, \Omega)\frac{\Phi_m\left(\delta + \Lambda^{\mathsf{T}}\Omega^{-1}(\mathbf{y}-\xi); \Gamma - \Lambda^{\mathsf{T}}\Omega^{-1}\Lambda\right)}{\Phi_m(\delta; \Gamma)} \quad \mathbf{y} \in \mathbb{R}^d \tag{2}$$

where $\delta \in \mathbb{R}^m$, $\xi \in \mathbb{R}^d$, $\Gamma \in \mathbb{R}^{m\times m}$, $\Lambda \in \mathbb{R}^{m\times d}$ and $\Omega \in \mathbb{R}^{d\times d}$ is a positive definite covariance matrix and $\varphi_d(., \xi, \Omega)$ is the density function of a d-dimensional normal with mean vector ξ and covariance matrix Ω and $\Phi_m(.; \sum)$ is the multivariate normal cumulative function with covariance matrix $\sum$.

Let

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} \sim N_{m+d}\left(\begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{\xi} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^{\mathsf{T}} \\ \boldsymbol{\Lambda} & \boldsymbol{\Omega} \end{pmatrix}\right)$$

A d-dimensional random vector $\mathbf{Y}$ is said to have a unified multivariate skew-normal, $\mathbf{Y} \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$ say, if

$$\mathbf{Y} \stackrel{d}{=} \mathbf{V} \mid \mathbf{U} > 0. \tag{3}$$

The density of this random vector can be written as (2).

For more information on multivariate skew-normal distributions and their applications we refer the reader to Azzalini and Dalla Valle (1996), González-Farías et al. (2003), etc. An overview of which can be found in the book edited by Genton (2004) and in the review paper by Azzalini (2005).

The remainder of this paper is organized as follows. In Section 2, we state and prove the main theorem of the paper and deduce some useful corollaries in regression analysis. Section 3 contains some numerical applications of our results.

2. Main results

Consider the following partition of $\mathbf{Y}$ and its corresponding parameters

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{\xi}_2 \end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{21}^{\mathsf{T}} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_1 \\ \boldsymbol{\Lambda}_2 \end{pmatrix} \tag{4}$$

where $\mathbf{Y}_1$ is a vector of dimension $d-1$ and $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ have dimensions $(d-1)\times m$ and $1\times m$ respectively. The two following lemmas are generalizations of those presented in Sheikhi and Jamalizadeh (2011).

Lemma 1 [9]. *If* $\boldsymbol{Y} = \left(\boldsymbol{Y}_1^{\mathsf{T}}, \boldsymbol{Y}_2\right)^{\mathsf{T}} \sim SUN_{d,m}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$, *then*

a) $\mathbf{Y}_1^{\mathsf{T}} \sim SUN_{(d-1),m}(\boldsymbol{\xi}_1, \boldsymbol{\delta}, \boldsymbol{\Omega}_{11}, \boldsymbol{\gamma}, \boldsymbol{\Lambda}_1)$

b) $\mathbf{Y}_2|\mathbf{Y}_1 = \mathbf{y}_1 \sim SUN_{1,m}(\xi_{2.1}, \boldsymbol{\delta}_{2.1}, \omega_{22.1}, \boldsymbol{\Gamma}_{2.1}, \boldsymbol{\Lambda}_{2.1})$

c) $M_{Y_2|\mathbf{Y}_1=\mathbf{y}_1}(t) = \exp(\xi_{2.1}t + \frac{1}{2}t^2\omega_{22.1})\frac{\Phi_m(\boldsymbol{\delta}_{2.1}+\boldsymbol{\Lambda}_{2.1}t;\boldsymbol{\Gamma}_{21.})}{\Phi_m(\boldsymbol{\delta}_{2.1};\boldsymbol{\Gamma}_{2.1})} \quad t \in \mathbb{R}$

where $\xi_{2.1} = \xi_2 + \boldsymbol{\Omega}_{2.1}\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\xi}_1)$, $\boldsymbol{\delta}_{2.1} = \boldsymbol{\delta} + \boldsymbol{\Lambda}_1^{\mathsf{T}}\boldsymbol{\Omega}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\xi}_1)$, $\omega_{22.1} = \omega_{22} - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Omega}_{21}^{\mathsf{T}}$, $\boldsymbol{\Gamma}_{2.1} = \boldsymbol{\Gamma} - \boldsymbol{\Lambda}_1^{\mathsf{T}}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Lambda}_1$ *and* $\boldsymbol{\Lambda}_{2.1} = \boldsymbol{\Lambda}_2 - \boldsymbol{\Omega}_{21}\boldsymbol{\Omega}_{11}^{-1}\boldsymbol{\Lambda}_1$.

Lemma 2 [8]. *If* $\mathbf{Y} = \left(\mathbf{Y}_1^{\mathsf{T}}, Y_2\right)^{\mathsf{T}} \sim SUN_{d,m}(\xi, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$, *then the conditional expectation of* Y_2 *given* $\mathbf{Y}_1 = \boldsymbol{y}_1$ *in (4) is*

$$E(Y_2|\mathbf{Y}_1 = \mathbf{y}_1) = \xi_{2.1} + \frac{G_m(0;\ \boldsymbol{\delta}_{2.1}, \boldsymbol{\Lambda}_{2.1}, \boldsymbol{\Gamma}_{2.1})}{\Phi_m(\boldsymbol{\delta}_{2.1}; \boldsymbol{\Gamma}_{2.1})} \tag{5}$$

where $G_m(\boldsymbol{0};\ \boldsymbol{A},\ \boldsymbol{B},\ \boldsymbol{\Sigma}) = \frac{\partial}{\partial \boldsymbol{t}} \Phi_m\left(\boldsymbol{At} + \boldsymbol{B}; \boldsymbol{\Sigma}\right)|_{\boldsymbol{t}=\boldsymbol{0}}$.

We now define $S(\mathbf{X})$ to be the class of all permutations of components of the random vector **X,** i.e. $S(\mathbf{X})$=$\{\mathbf{X}^{(i)} = \mathbf{P}_i\mathbf{X};\ i = 1, 2, \ldots, n!\}$ where $\mathbf{P}_i$ is an $n \times n$ permutation matrix. We also define the matrix $\boldsymbol{\Delta}$ to be the difference matrix of dimension $n-1$ by n such that its ith row is $\mathbf{e}_{n,i+1}^{\mathsf{T}} - \mathbf{e}_{n,i}^{\mathsf{T}}$, $i = 1, 2, \ldots, n-1$, where $\mathbf{e}_{n,1}, \mathbf{e}_{n,2}, \ldots, \mathbf{e}_{n,n-1}$ are n-dimensional unit basis vectors, i.e. $\boldsymbol{\Delta}X = (X_2 - X_1, X_3 - X_2, \ldots, X_n - X_{n-1})^{\mathsf{T}}$. Further, let $\mathbf{X}^{(i)}$ be the ith permutation of the random vector $\mathbf{X}$. We have $P\left(\boldsymbol{\Delta}X^{(i)} \geq \mathbf{0}\right) = 1 - \Phi_m(-(\boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\boldsymbol{\Delta}^{\mathsf{T}})^{-1/2}\boldsymbol{\Delta}\,\mu_x^{(i)})$ where $\mu_x^{(i)}$ and $\sum_{\mathbf{xx}}^{(i)}$ are the mean vector and covariance matrix of the random vector $\mathbf{X}^{(i)}$, respectively. Hereafter we adopt the notation $G_i(\mathbf{t}, \xi, \sum)$ for $P\left(\boldsymbol{\Delta}X^{(i)} \geq \mathbf{t}\right)$.

Theorem 1 *Under the assumption of model (1) The cdf of random vector* $\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}\right)^{\mathsf{T}}$ *is the mixture*

$$F_{\mathbf{Z},\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)},\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}}(\mathbf{z}, x, y) = \textstyle\sum_{i=1}^{n!} F_{SUN}(\mathbf{z}, x, y; \xi_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i) G_i(\mathbf{t}, \xi, \sigma)$$

where $F_{SUN}(\boldsymbol{.};\ \xi_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i)$ *is the cdf of unified multivariate skew-normal with* $\xi_i = \left(\mu_z, \mathbf{a}^{\mathsf{T}}\,\mu_{\mathbf{x}}^{(i)}, \mathbf{b}^{\mathsf{T}}\,\mu_{\mathbf{y}}^{(i)}\right)^{\mathsf{T}}$, $\boldsymbol{\delta}_i = \boldsymbol{\Delta}\,\mu_{\mathbf{x}}^{(i)}$, $\boldsymbol{\Gamma}_i = \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\boldsymbol{\Delta}^{\mathsf{T}}$, $\boldsymbol{\Lambda}_i = \left(\boldsymbol{\Delta}\sum_{\mathbf{xz}}^{(i)}, \boldsymbol{\Delta}\sum_{\mathbf{xx}}^{(i)}\mathbf{a}, \boldsymbol{\Delta}\sum_{\mathbf{yy}}^{(i)}\mathbf{b}\right)^{\mathsf{T}}$ *and* $\sum_{\mathbf{ux}}^{(i)}$ *is the covariance matrix of random vector U and the ith permutation of the random vector X. Moreover,* Ω_i *is the covariance matrix of* $\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}^{(i)}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}^{(i)}\right)^{\mathsf{T}}$.

Proof. We have

$$\begin{aligned}
F_{\mathbf{Z},\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(2)},\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[2]}}(\mathbf{z}, x, y) &= P\left(\mathbf{Z} \leq \mathbf{z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)} \leq x, \mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]} \leq y\right) \\
&= \sum_{i=1}^{n!} P\left(\mathbf{Z} \leq \mathbf{z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}^{(i)} \leq x, \mathbf{b}^{\mathsf{T}}\mathbf{Y}^{(i)} \leq y | \boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) P\left(\boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) \\
&= \sum_{i=1}^{n!} P\left(\mathbf{Z} \leq \mathbf{z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}^{(i)} \leq x, \mathbf{b}^{\mathsf{T}}\mathbf{Y}^{(i)} \leq y | \boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) G_i(\mathbf{t}, \xi, \Sigma).
\end{aligned}$$

Furthermore,

$$\begin{pmatrix} \boldsymbol{\Delta}\mathbf{X}^{(i)} \\ \mathbf{Z} \\ \mathbf{a}^{\top}\mathbf{X}^{(i)} \\ \mathbf{b}^{\top}\mathbf{Y}^{(i)} \end{pmatrix} \sim N_{n+p+1}\left(\begin{pmatrix} \boldsymbol{\Delta}\boldsymbol{\mu}_{\mathbf{x}}^{(i)} \\ \boldsymbol{\mu}_{\mathbf{z}} \\ \mathbf{a}^{\top}\boldsymbol{\mu}_{\mathbf{x}}^{(i)} \\ \mathbf{b}^{\top}\boldsymbol{\mu}_{\mathbf{y}}^{(i)} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xx}}^{(i)}\boldsymbol{\Delta}^{\top} & \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xz}}^{(i)} & \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xx}}^{(i)}\mathbf{a} & \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{yy}}^{(i)}\mathbf{b} \\ & \boldsymbol{\Sigma}_{\mathbf{zz}} & \boldsymbol{\Sigma}_{\mathbf{zx}}^{(i)}\mathbf{a} & \boldsymbol{\Sigma}_{\mathbf{zy}}^{(i)}\mathbf{b} \\ & & \mathbf{a}^{\top}\boldsymbol{\Sigma}_{\mathbf{xx}}^{(i)}\mathbf{a} & \mathbf{a}^{\top}\boldsymbol{\Sigma}_{\mathbf{xy}}^{(i)}\mathbf{b} \\ & & & \mathbf{b}^{\top}\boldsymbol{\Sigma}_{\mathbf{yy}}^{(i)}\mathbf{b} \end{pmatrix} \right).$$

Now using (3) we immediately conclude that

$$\left(\mathbf{Z}, \mathbf{a}^{\top}\mathbf{X}^{(i)}, \mathbf{b}^{\top}\mathbf{Y}^{(i)}\right)^{\top} | \boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0} \sim SUN_{p+2,n-1}(\boldsymbol{\xi}_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i).$$

This establishes the theorem.

Using the previous theorem and lemma 1, we find the conditional distribution of linear combinations of concomitants given linear combinations of order statistics and covariates.

Corollary 1 Under the assumptions of model (1) the cdf of the random variable $\mathbf{b}^{\top}\mathbf{Y}_{[n]}$ condition on **Z=z** and $\mathbf{a}^{\top}\mathbf{X}_{(n)} = x$ is the mixture

$$F_{\mathbf{b}^{\top}\mathbf{Y}_{[n]} | \mathbf{Z}, \mathbf{a}^{\top}\mathbf{X}_{(n)}}(y | \mathbf{z}, x) = \sum_{i=1}^{n!} F_{SUN}(y | \mathbf{z}, x; \boldsymbol{\xi}_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i) G_i(\mathbf{t}, \boldsymbol{\xi}, \boldsymbol{\Sigma})$$

where $F_{SUN}(y | \mathbf{z}, x; \boldsymbol{\xi}_i, \boldsymbol{\delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i, \boldsymbol{\Lambda}_i)$ is the cdf of conditional unified skew-normal of $\mathbf{b}^{\top}\mathbf{Y}^{(i)} = y$ given **Z=z** and $\mathbf{a}^{\top}\mathbf{X}^{(i)} = x$ and the parameters are as in theorem 1.

The following corollary is obvious via lemma 2.

Corollary 2 Under the assumptions of model (1) the regression equation of $\mathbf{b}^{\top}\mathbf{Y}_{[n]}$ on $\mathbf{Z}$ and $\mathbf{a}^{\top}\mathbf{X}_{(n)}$ is

$$E(\mathbf{b}^{\top}\mathbf{Y}_{[n]} | \mathbf{Z}=\mathbf{z}, \mathbf{a}^{\top}\mathbf{X}_{(n)} = x) = \sum_{i=1}^{n!} \xi_{2.1}^{(i)} + \frac{G_{n-1}^{(i)}(0; \boldsymbol{\delta}_{2.1}, \boldsymbol{\Lambda}_{2.1}, \boldsymbol{\Gamma}_{2.1})}{\Phi_{n-1}(\boldsymbol{\delta}_{2.1}; \boldsymbol{\Gamma}_{2.1})}$$

where the superscript (i) denotes the parameters based on the *i*th permutation of $\mathbf{X}$.

In the remainder of this section, we shall focus on a special case of the multivariate normal distribution. Let the joint distribution of a p-dimensional random vector $\mathbf{Z}$, and two random vectors $\mathbf{X}$ and $\mathbf{Y}$ follow a $2n+p$ dimensional exchangeable multivariate normal random distribution, i.e. its covariance matrix is equicorrelated. Hence we have

$$\begin{pmatrix} \mathbf{Z} \\ \mathbf{X} \\ \mathbf{Y} \end{pmatrix} \sim N_{2n+p}\left(\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{z}} \\ \boldsymbol{\mu}_{\mathbf{x}} \\ \boldsymbol{\mu}_{\mathbf{y}} \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{zz}} & \boldsymbol{\Sigma}_{\mathbf{xz}}^{\top} & \boldsymbol{\Sigma}_{\mathbf{yz}}^{\top} \\ \boldsymbol{\Sigma}_{\mathbf{xz}} & \boldsymbol{\Sigma}_{\mathbf{x}} & \boldsymbol{\Sigma}_{\mathbf{xy}}^{\top} \\ \boldsymbol{\Sigma}_{\mathbf{yz}} & \boldsymbol{\Sigma}_{\mathbf{xy}} & \boldsymbol{\Sigma}_{\mathbf{y}} \end{pmatrix} \right) \tag{6}$$

where

$$\boldsymbol{\mu}_{\mathbf{x}} = \mu_x \mathbf{1}_n, \boldsymbol{\mu}_{\mathbf{y}} = \mu_y \mathbf{1}_n, \boldsymbol{\Sigma}_{\mathbf{xx}} = \sigma_x^2 \left[\rho_x \mathbf{1}_n \mathbf{1}_n^{\mathsf{T}} + (1-\rho_x)\mathbf{I}_n\right],$$
$$\boldsymbol{\Sigma}_{\mathbf{yy}} = \sigma_y^2 \left[\rho_y \mathbf{1}_n \mathbf{1}_n^{\mathsf{T}} + (1-\rho_y)\mathbf{I}_n\right], \boldsymbol{\Sigma}_{\mathbf{xy}} = \rho_{xy}\sigma_x\sigma_y \mathbf{J},$$

where $\mathbf{1}_n = (1,\ldots,1)^{\mathsf{T}}$, $\mathbf{I}_n = diag(1,\ldots,1)$ and $\mathbf{J} = [1]_{n\times n}$. This model is the generalization of that assumed in Viana and Lee (2006).

Sheikhi and Jamalizadeh (2011) found the joint distribution of two linear combinations of order statistics in the presence of a covariate random variable under the exchangeable assumption and presented some regression analyses.

Theorem 2 Under the assumption of model (5),

$$\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}\right)^{\mathsf{T}} \sim SUN_{p+2,n-1}\left(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}\right)$$

where $\boldsymbol{\xi} = \left(\boldsymbol{\mu}_z, \mathbf{a}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{y}}\right)^{\mathsf{T}}, \boldsymbol{\delta} = \mathbf{0}$, $\boldsymbol{\Gamma} = \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{x}}\boldsymbol{\Delta}^{\mathsf{T}}, \boldsymbol{\Lambda} = \left(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xz}}, \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xx}}a, \Delta\boldsymbol{\Sigma}_{\mathbf{yy}}b\right)^{\mathsf{T}}$

and $\boldsymbol{\Omega}$ is the covariance matrix of $\left(\mathbf{Z},\ \mathbf{a}^{\mathsf{T}}\mathbf{X},\ \mathbf{b}^{\mathsf{T}}\mathbf{Y}\right)^{\mathsf{T}}$.

Proof. Since $P\left(\boldsymbol{\Delta}\mathbf{X}^{(i)} \geq \mathbf{0}\right) = \frac{1}{n!}$, $i = 1,\ldots,n!$, by exchangeability we have

$$F_{\mathbf{Z},\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(2)},\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[2]}}(\mathbf{t}_1, t_2, t_3) = P\left(\mathbf{Z} \leq \mathbf{t}_1, \mathbf{a}^{\mathsf{T}}\mathbf{X} \leq t_2, \mathbf{b}^{\mathsf{T}}\mathbf{Y} \leq t_3 | \boldsymbol{\Delta}\mathbf{X} > 0\right).$$

Moreover,

$$\begin{pmatrix} \boldsymbol{\Delta}\mathbf{X} \\ \mathbf{Z} \\ \mathbf{a}^{\mathsf{T}}\mathbf{X} \\ \mathbf{b}^{\mathsf{T}}\mathbf{Y} \end{pmatrix} \sim N_{n+p+1}\left(\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\mu}_z \\ \mathbf{a}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}} \\ \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{y}} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xx}}\boldsymbol{\Delta}^{\mathsf{T}} & \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xz}} & \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{a} & \boldsymbol{\Delta}\boldsymbol{\Sigma}_{\mathbf{xy}}\mathbf{b} \\ & \boldsymbol{\Sigma}_{\mathbf{zz}} & \boldsymbol{\Sigma}_{\mathbf{xz}}^{\mathsf{T}}\mathbf{a} & \boldsymbol{\Sigma}_{\mathbf{yz}}^{\mathsf{T}}\mathbf{b} \\ & & \mathbf{a}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{a} & \mathbf{a}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{xy}}\mathbf{b} \\ & & & \mathbf{b}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{yy}}\mathbf{b} \end{pmatrix}\right).$$

So $\left(\mathbf{Z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}, \mathbf{b}^{\mathsf{T}}\mathbf{Y}\right)^{\mathsf{T}} | (\boldsymbol{\Delta}\mathbf{X} > 0) \sim SUN_{p+2,n-1}(\boldsymbol{\xi}, \boldsymbol{\delta}, \boldsymbol{\Omega}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$, where the parameters are as given in theorem 2.

We may also interested in predicting the concomitants using order statistics and some covariates. The following corollary provides such a regression equation.

Corollary 3 Under the assumptions of model (1) the regression equation of $\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]}$ on $\mathbf{Z}$ and $\mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)}$ is

$$E(\mathbf{b}^{\mathsf{T}}\mathbf{Y}_{[n]} \,|\, \mathbf{Z}{=}\mathbf{z}, \mathbf{a}^{\mathsf{T}}\mathbf{X}_{(n)} = x) = \xi_{2.1} + \frac{G_{n-1}(0;\, A, \boldsymbol{\Omega})}{\Phi_{n-1}\,(\boldsymbol{\delta}; \boldsymbol{\Gamma})}$$

where

$$\xi_{2.1} = \mathbf{b}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{y}} + \frac{1}{\mathbf{b}^{\mathsf{T}}\sum_{\mathbf{yy}}\mathbf{b}}\left[\mathbf{b}^{\mathsf{T}}\sum_{\mathbf{yz}}(\mathbf{Z}-\mathbf{z}) + \mathbf{b}^{\mathsf{T}}\sum_{\mathbf{xy}}\mathbf{a}(\mathbf{a}^{\mathsf{T}}\mathbf{X} - \mathbf{a}^{\mathsf{T}}\boldsymbol{\mu}_{\mathbf{x}})\right]$$

and $G_{n-1}(0;\, A, \boldsymbol{\Omega})$ was defined in lemma 2.

The regression equation of $Y_{[i]}$ on $\mathbf{Z}$ and $X_{(i)}$ may be determined by letting the i-th component of the random vectors $\mathbf{a}$ and $\mathbf{b}$ equal to 1 and other all components equal to zero. Specifically, this regression equation and the regression equation of $Y_{[i]}$ on $\mathbf{Z}$ and X coincide and expressed as

$$Y_{[i]} = \mu_y + \frac{1}{\rho_y\sigma_y^2}\left[\boldsymbol{\sigma}_{y\mathbf{z}}(\mathbf{Z}-\mathbf{z}) + \rho_{xy}\sigma_x\sigma_y(X-\mu_x)\right] + \frac{G_{n-1}(0;\, A, \boldsymbol{\Omega})}{\Phi_{n-1}\,(\boldsymbol{\delta}; \boldsymbol{\Gamma})} \tag{7}$$

where $\boldsymbol{\sigma}_{y\mathbf{z}} = \mathrm{cov}(\mathbf{Z}, Y)$.

Olkin and Viana (1995) discussed that the linear regression of a random variable Z on $\mathbf{X}$ and Z on $\mathbf{X}_{(n)}$ coincide.

3. Numerical results

Viana and Lee (2006) considered the data from a pilot study in which a number of physiological parameters were measured at a specific site on the left and right brain hemispheres of subjects participating in a study conducted at the sleep centre of the University of Illinois at Chicago. They considered the following variables, jointly observed in a sample of $N = 30$ subjects:

(Z) : age;
(X_s) : tissue oxygenation on the right site;
(X_d) : tissue oxygenation on the left site;
(Y_s) : total hemoglobin on the right site;
(Y_d) : total hemoglobin on the left site.

They explored their results by assuming $\mathbf{X}^{\mathsf{T}} = (X_s, X_d)$ and $\mathbf{Y}^{\mathsf{T}} = (Y_s, Y_d)$. In particular, they estimated the covariance matrix of the random vector $(Z, \mathbf{X}_{(2)}^{\mathsf{T}}, \mathbf{Y}_{[2]}^{\mathsf{T}})^{\mathsf{T}}$. In this section we use these data to estimate the distributions obtained in the previous section. We assume that $(Z, \mathbf{X}^{\mathsf{T}}, \mathbf{Y}^{\mathsf{T}})^{\mathsf{T}}$ follows a 5 dimensional exchangeable multivariate normal distribution. The MLE of parameters are as follows:

$$\hat{\mu} = \begin{pmatrix} 50.000 \\ 57.275 \\ 57.275 \\ 39.695 \\ 39.695 \end{pmatrix}, \hat{\Sigma} = \begin{pmatrix} 128.410 & -28.110 & -28.110 & -35.996 & -35.996 \\ -28.110 & 40.016 & 26.090 & 42.909 & 42.909 \\ -28.110 & 26.090 & 40.016 & 42.909 & 42.909 \\ -35.996 & 42.909 & 42.909 & 159.58 & 113.06 \\ -35.996 & 42.909 & 42.909 & 113.06 & 159.58 \end{pmatrix}$$

Hence, $\hat{\rho}_x$=0.651**,** $\hat{\rho}_y$=0.722, $\hat{\rho}_{xz} = 0.392$**,** $\hat{\rho}_{yz} = 0.251$**,** $\hat{\rho}_{xy} = 0.536$. By considering **a** = **b** = **(**0, 1**)**$^{\mathsf{T}}$, theorem 2 implies that $(Z, X_{(2)}, Y_{[2]})^{\mathsf{T}} \sim SUN_{3,1}(\xi, \delta, \Omega, \Gamma, \Lambda)$, where

$$\xi = \begin{pmatrix} 50.000 \\ 57.275 \\ 39.695 \end{pmatrix}, \Omega = \begin{pmatrix} 128.410 & -28.110 & -35.996 \\ -28.110 & 40.016 & 42.909 \\ -35.996 & 42.909 & 159.58 \end{pmatrix}, \Lambda = \begin{pmatrix} 0 \\ 13.965 \\ 0 \end{pmatrix},$$

$$\delta = 0, \Gamma = 27.931.$$

Furthermore, the estimated regression equation of $Y_{[2]}$ on $X_{(2)}$ and Z follows from (6). We readily obtain

$$Y_{[2]} = -16.685 + 0.054Z + 1.034X_{(2)} \ .$$

Also, the linear regression of $Y_{[2]}$ on Z is easily estimated as $Y_{[2]} = 59.909 - 0.287Z$, then as obtained by Viana and Lee (2006). In addition, the regression of $Y_{[2]}$ on $X_{(2)}$ can be expressed as $Y_{[2]} = 24.294 + 0.269X_{(2)}$.

Similarly, we may obtain the joint distribution of $(Z, X_{(1)}, Y_{[1]})$ as well as the regression equation of $Y_{[1]}$ on $X_{(1)}$ and Z by letting **a** = **b** = **(**1, 0**)**$^{\mathsf{T}}$.

4. Conclusion

In this work we find the joint distribution of a linear combination of order statistics and a linear combination of their concomitants in the presence of some covariate random variables as a member of the skew-normal distribution. Some useful special cases of this distribution are investigated as well as some conditional distributions.

The application of our results in density estimation and regression analysis is illustrated by a numerical data set. We hope to extend our results to elliptical distributions in the future.

References

Arellano-Valle, R. B. and Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33, 561–574.

Arellano-Valle, R. B. and Genton, M. G. (2007). On the exact distribution of linear combinations of order statistics from dependent random variables. *Journal of Multivariate Analysis*, 98, 1876–1894.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* 83, 715–726.

David, H. A. and Nagaraja, H.N. (2003). *Order Statistics*, third ed. John Wiley and Sons, New York.

Genton, M. G. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Edited vol., Chapman and Hall/CRC, Boca Raton, FL, 416 pp.

Goel, P. K. and Hall, P. (1994). On the average difference between concomitants and order statistics. *Annals of Probability*, 22, 126–144.

González-Farías, G., Domínguez-Molina, J. A. and Gupta, A. K. (2004b). Additive properties of skew-normal random vectors. *Journal of Statistical Planning and Inference*, 126, 521–534.

Gupta, A. K., Gonzlez-Faras, G. and Domnguez-Molina, J. A. (2004). A multivariate skew normal distribution. *Journal of Multivariate Analysis*, 89, 181–190.

Loperfido, N. (2008). Modeling maxima of longitudinal contralateral observations. *Test* 17, 370–380.

Olkin, I. and Viana, M. (1995). Correlation analysis of extreme observations from a multivariate normal distribution. *Journal of the American Statistical Association*, 90, 1373–1379.

Sheikhi, A. and Jamalizadeh, A. (2011). Regression analysis using order statistics. *Statistical Papers*, 52, 885–892.

Tsukibayashi, S. (1998). The joint distribution and moments of an extreme of the dependent variable and the concomitant of an extreme of the independent variable. *Communications in Statistics-Theory and Methods*, 27(7), 639–651.

Viana, M. A. G. and Lee, H. M. (2006). Correlation analysis of ordered symmetrically dependent observations and their concomitants of order statistics. *Scandinavian Journal of Statistics*, 34(2), 327–340.

Yang, S. (1981). Linear functions of concomitant of order statistics with application to nonparametric estimation of a regression function. *Journal of the American Statistical Association*, 76, 658–662.

SORT 38 (1) January-June 2014, 53-72

Decision making in kidney paired donation programs with altruistic donors

Yijiang Li[1], Peter X.-K. Song[2], Alan B. Leichtman[3], Michael A. Rees[4] and John D. Kalbfleisch[5]

Abstract

In recent years, kidney paired donation has been extended to include living non-directed or altruistic donors, in which an altruistic donor donates to the candidate of an incompatible donor-candidate pair with the understanding that the donor in that pair will further donate to the candidate of a second pair, and so on; such a process continues and thus forms an altruistic donor-initiated chain. In this paper, we propose a novel strategy to sequentially allocate the altruistic donor (or bridge donor) so as to maximize the expected utility; analogous to the way a computer plays chess, the idea is to evaluate different allocations for each altruistic donor (or bridge donor) by looking several moves ahead in a derived look-ahead search tree. Simulation studies are provided to illustrate and evaluate our proposed method.

MSC: 62-Statistics; 90-Operations Research, Mathematical Programming.

Keywords: Altruistic donors, decision analysis, kidney paired donation, look-ahead search tree.

1. Introduction

For patients with end stage renal disease (ESRD), kidney transplantation is a preferred treatment as compared with dialysis for it provides not only a longer survival but also a better quality of life (Evans et al., 1985; Russell et al., 1992; Wolfe et al., 1999). Accord-

[1] Statistician at Google Inc., Mountain View, CA 94043, USA (Email: yijiang1121@gmail.com).

[2] Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA (Email: pxsong@umich.edu).

[3] Professor, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA (Email: aleicht@med.umich.edu).

[4] Professor, Department of Urology, University of Toledo Medical Center, Toledo, Ohio 43614, USA (Email: michael.rees2@utoledo.edu).

[5] Professor, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA (Email: jdkalbfl@umich.edu).

Received: April 2013

Accepted: December 2013

ing to the Organ Procurement and Transplantation Network (OPTN), about 16,760 kidney transplants were performed per year from 2009 to 2012 in the U.S., while during that same period of time the yearly average number of patients added to the waiting list for kidney transplant surpassed 34,100. Part of this gap between supply and demand can be attributed to the unfortunate fact that many patients with kidney failure recruit willing organ donors who, upon evaluation, prove to be ABO blood type and/or Human Leukocyte Antigens (HLA) incompatible. With regard to blood type compatibility, A and B donors can donate to candidates of the same blood type or of type AB; AB donors can donate only to AB candidates; and O donors, known as universal donors, can donate to candidates of any blood type. The HLA incompatibility, on the other hand, is due to the candidate having antibodies against the HLA antigens of a potential donor resulting from prior exposure to donor antigens through pregnancy, transfusion or previous transplant. Both forms of incompatibility can lead to a rapid rejection of the transplanted organ and thus prohibit transplantation.

An evolving strategy, known as kidney paired donation (KPD) (Rapaport, 1986) matches one donor-candidate pair to another pair with a complementary incompatibility, such that the donor of the first pair donates to the candidate of the second, and vice versa; see Figure 1-A and Figure 1-B for illustrations of a two-way exchange and a three-way exchange. Although three-way or higher exchange cycles increase the chance of identifying compatible matches, most KPD programs restrict exchanges to at most three ways for two primary reasons. First, all surgical operations in a cycle must be performed simultaneously to avoid the possibility that one of the donors may renege. This requirement creates substantial logistical difficulties of scheduling, for example, eight surgeons and eight operating rooms at the same time for a four-way exchange. Second, the greater the length of an exchange cycle, the less likely the potential transplants involved will actually occur, for the whole exchange cycle collapses if any of the proposed transplants cannot proceed.

A fundamental problem in managing KPD programs lies in selecting the "optimal" set of kidney exchanges from among the many possible alternatives. This problem has been modeled and analyzed by economists using a game-theoretic approach (Roth et al., 2004). More general approaches have been developed to tackle such a problem via an integer programming (IP) formulation, first proposed by Roth et al. (2007); In this, each

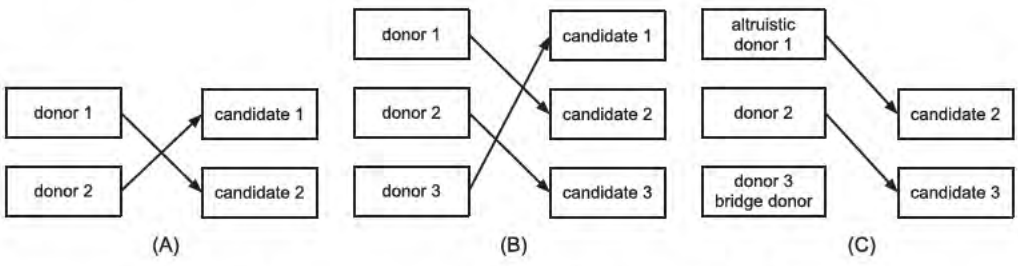

Figure 1: (A): A two-way exchange; (B): A three-way exchange; (C): A NEAD chain.

potential transplant was assigned equal weight, resulting in an allocation strategy that enables the greatest number of transplants to be potentially implemented. Abraham et al. (2007) adopted a more flexible weight assignment in this IP-based formulation and further developed an algorithm to reduce the computational complexity of managing large KPD programs. Li et al. (2013) considered a general utility-based evaluation of potential kidney transplants. Moreover, they explicitly took into account inherent uncertainties in managing KPD programs and exploited possible fall-back or contingent exchanges when the originally planned allocation cannot be fully executed. In a data-driven simulation system, they demonstrate that taking such additional elements into consideration would yield improved allocation strategies.

In recent years, KPD has also been extended to include living non-directed donors (LNDs), or altruistic donors; these are donors who have no designated candidates and decide to donate voluntarily to a stranger. In this context, an altruistic donor may donate to the candidate of an incompatible pair with the understanding that the donor of that pair will become a bridge donor, and further donate to the candidate of a second pair, and so on; such a process continues and thus forms an LND-initiated chain. One advantage to such chains as compared to two-way or higher order exchange cycles is that transplants along the chain do not need to be performed simultaneously (Montgomery et al., 2006; Roth et al., 2006). As a consequence, the donor whose incompatible candidate has received another donor's kidney but has yet to donate could donate later to another candidate; such donors are hence called "bridge donors". For this reason, this LND-initiated chain is sometimes called a non-simultaneous extended altruistic donor (NEAD) chain (Rees et al., 2009). Figure 1-C illustrates a NEAD chain. Kidneys from altruistic donors used to be designated to patients with no living donors and who have therefore been placed on a deceased-donor waiting list. A NEAD chain, however, allows for passing the altruism beyond saving just one patient, to potentially benefitting several patients in the chain; the final donor in an NEAD chain could still donate to the deceased-donor waiting list. The advantage of such chains has already been demonstrated via simulation studies by Gentry et al. (2009) and Ashlagi et al. (2011). In clinical practice, the standard way of incorporating LND and bridge donors into the optimization of a KPD is to consider chains up to a given length along with cycles in the optimization for each match run. Thus, at regular intervals, the KPD pool is examined and a set of chains segments and/or a set of cycles are chosen using the integer programming approach, and those chosen are implemented if possible.

In this paper, we consider a different strategy for developing a NEAD chain under uncertainties in a KPD program with one altruistic donor. We also discuss in general some possible extensions of this strategy to incorporate multiple altruistic donors. Analogous to the way a computer plays chess, we propose an approach to sequentially allocating an altruistic donor (or a bridge donor) so as to maximize the expected utility over a certain given number of moves. The idea is to evaluate different allocation options available for each altruistic donor (or bridge donor) by looking several moves ahead along a derived look-ahead search tree. With these options in mind, we proceed with the

next allocation of the altruistic or bridge donor that has the highest evaluation. This is the first step in developing an approach that would alternate between optimizing the use of LND and bridge donors and assigning cycles, each in an optimum way. This approach would then be compared with the standard simultaneous maximization over chains and cycles as described above.

The rest of the paper is organized as follows: in Section 2, we introduce a graph representation for a KPD program with altruistic donors. With this representation, we define the *optimal policy* in the context of managing a KPD program with one altruistic donor. This optimal policy can be obtained in general by following a standard decision-tree analysis, which we briefly illustrate in Section 3. The computation associated with this decision-tree based approach, however, is very expensive for large KPD programs. To address this issue, we propose, in Section 4, a more efficient and practical approach which sequentially extends a NEAD chain according to the utility calculated along a look-ahead search tree. Section 5 provides simulation studies to illustrate and evaluate our proposed strategy. In Section 6, we conclude with some discussion on possible extensions to incorporate multiple altruistic donors.

2. Problem formulation

In this section, we describe a graph representation for KPD programs that includes incompatible pairs as well as altruistic donors. We then define the optimal policy in the management of a KPD program with a single altruistic donor.

2.1. Graph representation

We represent a KPD program as a *directed graph*, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where the *vertex set*, $\mathcal{V} \equiv \mathcal{V}(\mathcal{G}) = \{1, 2, \cdots, m, m+1, \cdots, n\}$, consists of m altruistic donors and $n-m$ incompatible donor-candidate pairs, where $m \leq n$. We denote by, $\mathcal{V}_a \equiv \mathcal{V}_a(\mathcal{G}) = \{1, 2, \cdots, m\}$, the collection of altruistic donors, and $\mathcal{V}_p \equiv \mathcal{V}_p(\mathcal{G}) = \mathcal{V} \setminus \mathcal{V}_a$, the set of incompatible pairs. The *edge set*, $\mathcal{E} \equiv \mathcal{E}(\mathcal{G})$, is a *binary relation* on $\mathcal{V}$, consisting of ordered pairs of vertices in $\mathcal{V}$. An edge from i to j, denoted as (i, j), implies that the donor in pair i (or the altruistic donor i) is predicted to be compatible with the candidate in pair j. Such a prediction is based on a *virtual crossmatch* test, which involves computer cross-checking for blood type compatibility as well as comparing preexisting candidate antibodies against donor HLA antigens. Before a predicted compatible transplant can be further considered for an actual surgical operation, the compatibility must be confirmed by a more labour-intensive *laboratory crossmatch* test to assure histocompatibility; this involves incubating the serum of a candidate with the white blood cells of a prospective donor. Figure 2 illustrates a graph representation for a two-way exchange, a three-way exchange, and a NEAD chain, corresponding respectively to scenarios (A)-(C) in Figure 1.

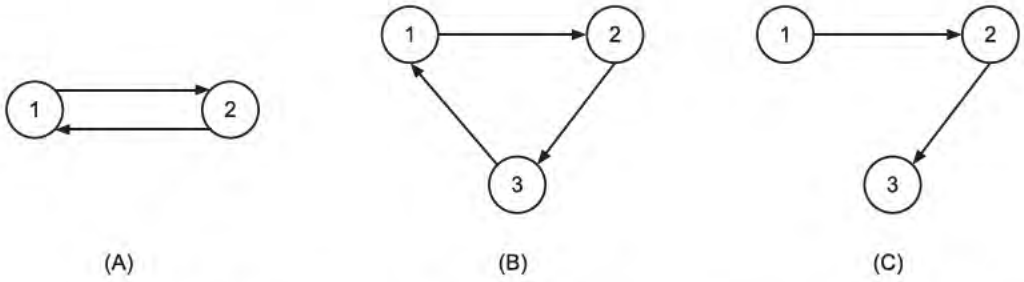

Figure 2: *(A): A graph representation of a KPD program with a two-way exchange cycle, where* $\mathcal{V}_p = \{1,2\}$ *and* $\mathcal{E} = \{(1,2),(2,1)\}$*; (B): A graph representation of a KPD program with a three-way exchange cycle, where* $\mathcal{V}_p = \{1,2,3\}$ *and* $\mathcal{E} = \{(1,2),(2,3),(3,1)\}$*; (C): A graph representation of a NEAD chain, where* $\mathcal{V}_a = \{1\}$*,* $\mathcal{V}_p = \{2,3\}$ *and* $\mathcal{E} = \{(1,2),(2,3)\}$*; donor* 3 *at the end of the chain becomes a bridge donor.*

The virtual crossmatch test is necessary because in practice the laboratory crossmatch test cannot be undertaken on all possibly compatible donors and candidates due to labour and resource limitations. Further, even if the laboratory crossmatch result is negative (non-reactive), an actual transplant operation may not occur due to other friction including, for example, refusal or illness or death of the candidate or the donor. To incorporate such stochastic features, we associate with each edge, $e = (i, j)$, a probability (denoted as p_e or p_{ij}) that e, if chosen, could result in an actual transplant operation (Li et al., 2013). Throughout the rest of the paper, we use the term "*is viable*" to indicate that an edge could lead to an actual transplant.

In addition, we associate with each edge (or potential transplant) a general utility (Li et al., 2013). Such utilities are often rule-based and determined by various attributes such as degree of sensitization of the candidate against the potential donor pool, or time since enrolment in the KPD. These utilities could also be based on predicted medical outcomes such as the estimated graft or patient survival, or the incremental years of recipient life that would accrue with a kidney transplant as opposed to a candidate's remaining on dialysis; see Wolfe et al. (2008). For each potential transplant $e = (i, j)$, we denote such an assigned utility as u_e or u_{ij}.

In this paper, our attention is not on the estimation of edge utilities and probabilities. It is worth noting though that research along this line is important and needed in the practical management of a KPD program; see more discussion on this aspect in Wolfe et al. (2008), Schaubel et al. (2009), and Li et al. (2013).

2.2. The optimal policy

One difficulty with selecting a long NEAD chain and then arranging transplants accordingly is that in practice this long chain can rarely be fully implemented. This is because the chain would break as soon as one transplant cannot proceed as planned. In this paper,

we propose to extend a NEAD chain sequentially in a near optimal way by selecting one potential transplant recipient at a time. In subsequent discussion, we note how this can be used as the basis of more general approaches.

Consider a KPD program with only one altruistic donor, i.e. $m = 1$ and $\mathscr{V}_a = \{1\}$. This naturally implies $(i,1) \notin \mathscr{E}$ for all $i \in \mathscr{V}$, as altruistic donors don't have designated candidates. For $j \in \mathscr{V}$ such that $j = 1$ or $(1,j) \in \mathscr{E}$, let $\mathscr{G}(j) \equiv (\mathscr{V}_j, \mathscr{E}_j)$ be a *subgraph* of $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where

$$\mathscr{V}_j = \{v \in \mathscr{V} : v \text{ is accessible from } j\},$$
$$\mathscr{E}_j = \{(v_1, v_2) \in \mathscr{E} : v_1 \in \mathscr{V}_j, v_2 \in \mathscr{V}_j, v_2 \neq j\}.$$

In this paper, a vertex j is said to be *accessible* from a vertex i if $i = j$ or if there exists a set of edges in $\mathscr{E}$, denoted as $\{(i_k, i_{k+1}), k = 0, 1, \cdots, n\}$ such that $i_0 = i$ and $i_{n+1} = j$. In general terms, $\mathscr{G}(j)$ represents the resulting KPD graph if the transplant according to $(1,j) \in \mathscr{E}$ is arranged and j becomes a bridge donor.

Managing a KPD program with one altruistic donor could then be viewed as a sequential decision problem, in which we start with $U = 0$ and $\mathscr{G} = \mathscr{G}(1)$, and then repeat the following steps until $|\mathscr{V}(\mathscr{G})| = 1$:

1. choose one edge from $A \equiv \{(1,j) : (1,j) \in \mathscr{E}\}$, say $(1,b)$.
2. if $(1,b)$ is viable, update

$$U \leftarrow U + u_{1b},$$
$$\mathscr{G} \leftarrow \mathscr{G}(b),$$
$$1 \leftarrow b;$$

if $(1,b)$ is not viable, update the KPD pool

$$\mathscr{G} \leftarrow \mathscr{G}_{-b}(1), \text{ where } \mathscr{G}_{-b} = (\mathscr{V}, \mathscr{E} \setminus \{(1,b)\}).$$

Step (i) is carried out to implement a *policy* that would be used to manage the KPD program by specifying what action from A to take at each loop; two sample policies are,

$$b = \underset{j:(1,j)\in A}{\operatorname{argmax}}\, u_{1j}$$

$$b = \underset{j:(1,j)\in A}{\operatorname{argmax}}\, u_{1j} p_{1j}.$$

These correspond to greedy algorithms that look at the next step only and manage to optimize the utility or the expected utility of that step. They may, of course, be very poor strategies since they ignore any subsequent implications of possible next steps.

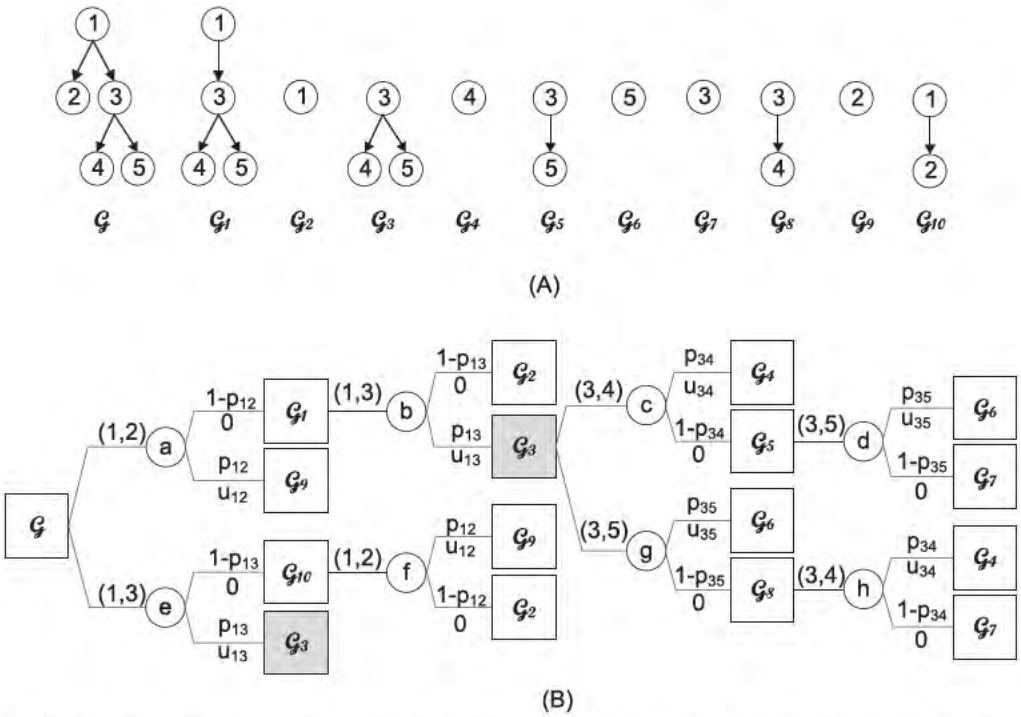

Figure 3: (A): A KPD program $\mathcal{G}$ with one altruistic donor and four incompatible pairs as well as various subgraphs of $\mathcal{G}$; (B): A standard decision tree analysis for a KPD program $\mathcal{G}$ as in (A), with squares representing decision nodes and circles indicating chance nodes; the decision node $\mathcal{G}_3$ (which is shaded) appears twice in the tree and hence is only drawn once.

For any given policy on $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the value of U after the algorithm terminates can be interpreted as the cumulative claimed utility. This value, which we denote by U_∞, is random; and its expectation could be used to evaluate the policy from which it arose. Among all policies defined in the above way, the *optimal policy* refers to the one that attains the highest value of $E(U_\infty)$. This way of defining the optimal policy provides a formal framework that will prove convenient in later discussions, even though in general one can rarely follow this optimal policy through until the iterative procedure ends. This is an important issue, arising due to various practical concerns, that we will revisit in Section 4.2.

Figure 3-A provides an illustrative example, where $\mathcal{G}$ represents a KPD program with four incompatible pairs (vertices 2, 3, 4 and 5) and one altruistic donor (vertex 1). Starting from $\mathcal{G}$, the action space is $A = \{(1,2),(1,3)\}$ and suppose we proceed by selecting $(1,2)$. If it is viable, this would lead to $\mathcal{G}(2)$, denoted as $\mathcal{G}_9$ in Figure 3-A, and the resulting value of U_∞ is u_{12}; if $(1,2)$ is not viable, we end up with $\mathcal{G}_1$, at which the updated action space becomes $A = \{(1,3)\}$. We then continue by selecting $(1,3)$, and if it is not viable, we stop at $\mathcal{G}_2$; if $(1,3)$ is viable, we then proceed to $\mathcal{G}_3$, at which the updated action space becomes $A = \{(3,4),(3,5)\}$; and we continue this process by

selecting one allocation from A. In this paper, we assume that edges in a KPD graph have an independence relationship. Though this assumption can be relaxed, it is a reasonable one when pair withdrawal (due to factors such as pregnancy, illness, or death) does not occur frequently; see Li et al. (2013) for related discussion.

3. Decision tree analysis for KPD

The optimal policy introduced in the previous section can be obtained by conducting a standard decision tree analysis, which we briefly illustrate below using a small example. The computation associated with such an analysis, however, can be rather complicated for large problems. We will return to this computational issue in Section 4, and present an alternative and more efficient approach to analyzing policies and optimizing the allocations. Note that a general mathematical framework derived from theories of Markov decision processes (MDPs) can be used to rigorously formulate the problem of managing KPD programs with altruistic donors (Li, 2012). However, solving for the optimal policy is computationally dificult for large or even moderate KPD problems, which poses a serious impediment to the development of practical algorithms based on this MDP framework. We briefly describe the MDP formulation in this section by using a particular example. In Section 4, we describe an alternative and more efficient way of analyzing the KPD that takes account of the fall-back options.

The structure of $\mathscr{G}$ in Figure 3-A cannot be used directly for a standard decision tree analysis due to the existence of various fall-back options; for example, if edge $(1,3)$ is selected but not viable, we could fall back to $(1,2)$. The complete analysis is instead provided by a derived decision tree (oriented from left to right) as shown in Figure 3-B, where squares represent *decision nodes* and circles indicate *chance nodes*. Each decision node is followed in this tree by a fixed number of chance nodes associated with all actions available at that decision node. Each chance node is then followed by two decision nodes corresponding to the two possible outcomes of choosing that chance node: one outcome is that the chosen transplant $e \in \mathscr{E}$ is viable, resulting in a utility of u_e, whereas the other is that e is not viable, for which zero utility is generated. These two utilities are associated with the edges from the chance node to the two corresponding decision nodes. For example, in Figure 3-B, starting from the decision node $\mathscr{G}$, two actions are available, either arrange a transplant according to edge $(1,2)$ leading to chance node a or according to edge $(1,3)$ leading to chance node e. In the case where $(1,2)$ is chosen, associated with the chance node a are two possible outcomes, $\mathscr{G}_1$ and $\mathscr{G}_9$, which occur with probabilities $1-p_{12}$ and p_{12} respectively. If $\mathscr{G}_9$ occurs, we claim a utility of u_{12}, and zero utility is generated if $\mathscr{G}_1$ occurs, for which we continue on this analysis from chance node b.

The *expected value* (EV) associated with a chance node or a decision node is calculated alternately in a backward direction along the tree from the right to the left. Precisely, (i) the EV at a leaf decision node is 0 (this could be set to some non-zero

number to represent the potential value associated with the corresponding bridge donor; see more discussion on this in Section 6); (ii) the EV at a chance node is computed by taking a weighted average of the sums of the utilities along the edges originating at this chance node and the EVs at the corresponding successor decision nodes; (iii) the EV at a non-leaf decision node is calculated by taking the maximum of the EVs of its children nodes.

For example, in Figure 3-B, the EVs at decision nodes $\mathscr{G}_5$ and $\mathscr{G}_8$ are $EV[\mathscr{G}_5] = EV[d] = p_{35}u_{35}$ and $EV[\mathscr{G}_8] = EV[h] = p_{34}u_{34}$ respectively. The EVs at chance nodes c and g are $EV[c] = p_{34}u_{34} + (1-p_{34})EV[\mathscr{G}_5]$ and $EV[g] = p_{35}u_{35} + (1-p_{35})EV[\mathscr{G}_8]$ respectively. This indicates that $EV[c] \geq EV[g]$ if and only if $u_{34} \geq u_{35}$, and the action taken at $\mathscr{G}_3$ is therefore $(3,4)$ or $(3,5)$ depending on which one has the larger edge utility. The EV at node $\mathscr{G}_3$ is then calculated as

$$\begin{aligned} EV[\mathscr{G}_3] &= \max\{EV[c], EV[g]\} \\ &= \max\{p_{34}u_{34} + (1-p_{34})p_{35}u_{35}, p_{35}u_{35} + (1-p_{35})p_{34}u_{34}\}. \end{aligned} \tag{1}$$

After computing EVs associated with all decision and chance nodes in this way, the optimal policy at each decision node is to adopt the action associated with the chance node that has the maximum EV. This procedure starts from the root decision node, that is from the altruistic donor.

4. A look-ahead search tree-based strategy

The structure of the derived decision tree in Figure 3-B is much more complicated than the structure of $\mathscr{G}$ itself in Figure 3-A. As a result, the standard decision tree analysis as introduced in Section 3 results in substantial computational difficulties when the KPD graph is large. In this section, we address this issue by presenting a more efficient and practical approach that relies on evaluating different allocations for each altruistic donor (or bridge donor) according to a derived look-ahead search tree.

4.1. Identifying the optimal policy via a search tree

Consider first a KPD program, $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}_a = \{1\}, \mathscr{V}_p = \{2,3,\cdots,n\}$, and $\mathscr{E} = \{(1,i) : i = 2,3,\cdots,n\}$. Without loss of generality, assume $u_{12} \geq u_{13} \geq \cdots \geq u_{1n}$. For this specific KPD program, the optimal policy to follow at $\mathscr{G}$ is to try transplant $(1,2)$, and if it fails then try $(1,3)$, then $(1,4)$ and so forth. The associated EV of this policy is

$$EV[\mathscr{G}] = \sum_{k=2}^{n} \left\{ u_{1k}p_{1k} \prod_{i=2}^{k-1}(1-p_{1i}) \right\}. \tag{2}$$

Based on this fact, we could then select the optimal action to take from $\mathscr{G}$ directly and hence avoid explicitly constructing a decision tree and calculating the EV associated with each node of the tree, as would be required for the standard decision analysis in Section 3. This observation is very useful as we can see, for example, by applying formula (2) at the decision node $\mathscr{G}_3$ in Figure 3-B. This would lead to the optimal action of taking $(3,4)$ or $(3,5)$ depending on which one has the larger utility; and the EV at $\mathscr{G}_3$ is therefore computed as

$$EV[\mathscr{G}_3] = 1_{[u_{34} \geq u_{35}]} \{p_{34}u_{34} + (1-p_{34})p_{35}u_{35}\} + 1_{[u_{34} < u_{35}]} \{p_{35}u_{35} + (1-p_{35})p_{34}u_{34}\}. \tag{3}$$

Note that formula (3) is exactly equal to the one calculated via a standard decision analysis as in formula (1), but this latter approach requires calculating EVs at additional nodes $\mathscr{G}_5$ and $\mathscr{G}_8$.

Consider now a KPD program, $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where $\mathscr{V}_a = \{1\}$, let $A \equiv \{(1,j) : (1,j) \in \mathscr{E}\}$ and $u^*_{1j} \equiv u_{1j} + EV[\mathscr{G}(j)]$, for all $(1,j) \in A$. Without loss of generality, we assume $A = \{(1,j) : j = 2,3,\cdots,l\}$ and $u^*_{12} \geq u^*_{13} \geq \cdots \geq u^*_{1l}$. Then the optimal decision to take at $\mathscr{G}$ is to attempt transplant $(1,2)$; and if it fails, try $(1,3)$ and then $(1,4)$, and so on; the associated EV is

$$EV[\mathscr{G}] = \sum_{k=2}^{l} \left\{ p_{1k}u^*_{1k} \prod_{i=2}^{k-1} (1-p_{1i}) \right\}. \tag{4}$$

Based on this result, we could evaluate various choices in A by u^*_{1j} and then proceed with the one having the largest value. We repeatedly apply this procedure from terminal nodes up to sequentially form a NEAD chain, with formula (4) evaluating the expected utility in this process.

To identify the optimal action to take at $\mathscr{G}$, we recursively apply formula (4), which in fact does not require calculating EVs associated with all decision nodes and chance nodes, but only a fraction of them. These required nodes can then be organized according to their dependence relationship as in (4) to form a *search tree*. In this tree, the node on the left hand side of (4) is the *parent* while the nodes on the right hand side denote *children*; and edges connecting them represent the corresponding actions. The structure of this tree therefore allows us to compute EVs associated with its nodes recursively in a backward manner from the leaf nodes to the root.

Figure 4 provides an example of a search tree and illustrates calculating EVs associated with its nodes; the search tree in this figure only involves 5 nodes, much less than that of the decision tree in Figure 3-B. The optimal action to take at $\mathscr{G}$ in this example is edge $(1,3)$ if $u^*_{12} = u_{12}$ is smaller than $u^*_{13} = u_{13} + EV[\mathscr{G}_3]$ or edge $(1,2)$ if otherwise; the EV at $\mathscr{G}$ is therefore computed as

$$EV[\mathscr{G}] = 1_{[u^*_{13} \geq u^*_{12}]} \{p_{13}u^*_{13} + (1-p_{13})p_{12}u^*_{12}\} + 1_{[u^*_{13} < u^*_{12}]} \{p_{12}u^*_{12} + (1-p_{12})p_{13}u^*_{13}\}.$$

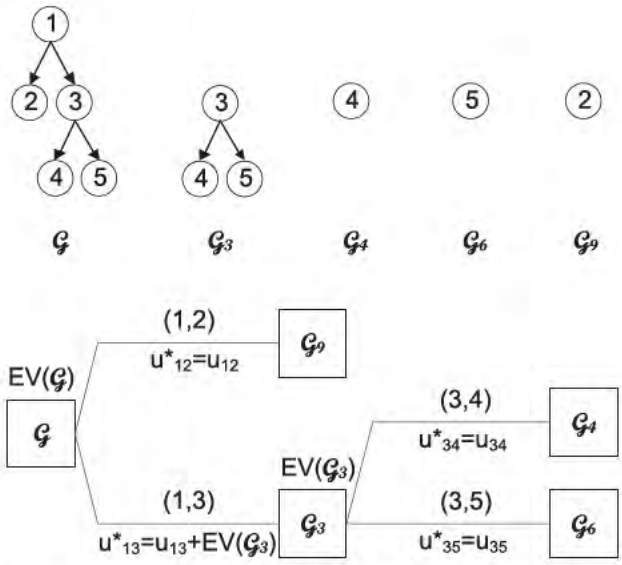

Figure 4: *A search tree-based analysis for a KPD program $\mathcal{G}$.*

Clearly, the decision analysis of this search tree is much simpler than that from a standard decision tree as in Figure 3-B, although both lead to the same result.

In general, the search tree associated with a KPD program can be constructed by an algorithm based on the classic depth-first search (DFS). We developed such an algorithm that also computes the EVs while performing a DFS on the KPD graph. The optimal policy is then determined by the following iterative algorithm:

1. for $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, construct the corresponding search tree by following a DFS-based algorithm, and compute the EV associated with each node of this search tree; this is done recursively from the terminal nodes up to the root node.
2. update the current action space $A \equiv \{(1,j) : (1,j) \in \mathcal{E}(\mathcal{G})\}$, and calculate $u^*_{1j} = u_{1j} + EV[\mathcal{G}(j)]$ for $(1,j) \in A$.
3. choose $(1,b) \in A$ with $b = \text{argmax}_{j:(1,j)\in A}\, u^*_{1j}$.
4. if $(1,b)$ is viable, update

$$\mathcal{G} \leftarrow \mathcal{G}(b), \text{ i.e. update the KPD graph,}$$
$$1 \leftarrow b, \text{ i.e. set the bridge donor } b \text{ as the new altruistic donor;}$$

   if $(1,b)$ is not viable, update the KPD pool

$$\mathcal{G} \leftarrow \mathcal{G}_{-b}(1), \text{ where } \mathcal{G}_{-b} = (\mathcal{V}, \mathcal{E} \setminus \{(1,b)\}).$$

5. go back to (ii) until $|\mathcal{V}(\mathcal{G})| = 1$.

For completeness, we also briefly describe here a slightly more complicated example in which the KPD itself is not a tree (as it is in Figure 3-A above). The KPD in this example is obtained by adding edges $(2,4)$ and $(4,3)$ to the KPD graph $\mathscr{G}$ in Figure 3-A. Note that vertex 4 can be reached in two distinct ways and gives rise to two distinct subgraphs. Specifically, if $(1,3)$ and $(3,4)$ are transplanted, $\mathscr{G}(4)$ only has vertex 4; if $(1,2)$ and $(2,4)$ are transplanted, $\mathscr{G}(4)$ has vertices 4 and 3 and also contains edge $(4,3)$.

4.2. A depth-k search tree

Although the search tree-based approach allows for a much more efficient analysis than does a standard decision tree analysis, constructing such a search tree is computationally very expensive for a general large KPD graph; in fact, the computation is extensive even without the effort entailed in computing EVs associated with nodes along that tree. This unfortunate fact poses a substantial difficulty in identifying the optimal policy when the KPD program is large.

Further, a more important issue is that the optimal policy (even if it could be computed) would most likely not be implementable in practice. This is mainly because the practical process of initiating and extending a NEAD chain would require a relatively long period of time, during which the KPD pool would constantly be updated and evolve as new pairs arrive and/or existing pairs withdraw or candidates die; in addition, candidates in the pool may also be transplanted via exchanges among incompatible pairs or deceased donor kidneys from the waiting list, since these would typically be arranged in parallel with the NEAD chain mechanism. Thus, assessing strategies by looking a long way down the tree from the root node is often not that useful in practice.

To address such problems, we propose to proceed by first deriving a subtree, which we call a *depth-k search tree*, from the original search tree. Such a subtree can be readily obtained by the same DFS-based algorithm as introduced in Section 4.1, by simply restricting the depth of the search from the root node to k. We then follow the recursive relationship as in formula (4) to calculate the EVs associated with corresponding nodes in the subtree, beginning this calculation from the leaf nodes at depth k and working up through the tree to the root node. The EV of a leaf node is set to zero or some reasonable measurement of the value of the corresponding bridge donor (see Section 6 for more discussion). At this stage the iterative procedure presented in Section 4.1 can be applied, but with one modification that – if the chosen action $(1,b)$ is viable, we regenerate a depth-k search tree rooted at $\mathscr{G}(b)$ and compute EVs associated with the nodes of this new tree.

Instead of the optimality that exists only in a rather idealized scenario, the policy obtained from the depth-k search tree provides a more practical evaluation of potential bridge donors and a greatly reduced computational complexity. Further, our simulation results (see Section 5) suggest that the allocation strategy derived from a search tree performs reasonably well for a moderate depth of, say, 3 or 4. It is useful to note that

this strategy is feasible in the context of the search tree approach of this section, but would still be very complicated to implement using the standard decision tree analysis; for example, the depth-1 search tree constructed according to formula (2) provides the same analysis as the one via a standard decision tree of depth $2n+1$.

5. Simulation studies

So far in this paper, we have explored a look-ahead search tree-based approach to manage a KPD program with one altruistic donor. This approach sequentially extends a NEAD chain by selecting one potential bridge donor at a time, taking into consideration the operational uncertainties and the long-term consequences associated with various possible selections. In this section, we provide simulation results of applying such an allocation strategy to manage a simulated KPD program.

5.1. Simulating incompatible pairs and altruistic donors

We simulate incompatible pairs and altruistic donors as in Li et al. (2013). For an incompatible pair, we simulate its candidate and donor separately from their own population distributions. Candidates are sampled at random (with replacement) from a database of incompatible pairs, which is derived from the University of Michigan KPD program. This database currently consists of 115 transplant candidates, each having at least one willing but incompatible donor. We are in the process of incorporating additional databases from other KPD programs for the purpose of reflecting a broader candidate variation. On the other hand, donors are simulated by sampling their blood types and HLA haplotypes respectively. Blood type is drawn from its U.S. population distribution: O, 44%; A, 42%; B, 10%; and AB, 4% (Stanford Blood Center, 2010), and HLA haplotypes are sampled according to a population frequency table derived from a public database on potential bone marrow donors (Maiers et al., 2007).

We consider a simulated donor-candidate pair as an incompatible pair, and hence include it in the KPD pool, if either their blood types mismatch or the donor's HLA haplotypes overlap with some of the candidate's antibody specificities. Finally, an altruistic donor is generated in the same way as we have described above for generating a donor in an incompatible pair.

5.2. Simulation setup

In Section 2.1, each potential transplant (which is predicted to be compatible by a virtual crossmatch test) is assigned a probability to reflect the inherent uncertainty in the system and a general utility to quantify the rule-based or outcome-based evaluation

of that potential transplant. As we have mentioned, estimation of these probabilities and utilities is an important aspect in the practical management of a KPD program. This also forms an independent line of research in parallel with the work of developing KPD allocation strategies. For illustrative purpose, our approach here is to obtain these utilities and probabilities according to certain simplified probability distributions, and then use them to study the method proposed in Section 4.2.

We perform a total of $3,000$ simulations. In all simulations, edge probabilities are generated from a uniform distribution, $U(0.1,0.5)$, which suggests an average success rate of 30% for a predicted compatible (by virtual crossmatch test) transplant. This rate is in line with the early experience at the University of Michigan KPD program and the Alliance for Paired Donation, though current success rate are somewhat higher. For edge utilities, we fix them at 1 for $1,000$ simulations, draw them from uniform $U(10,20)$ and $U(10,30)$ respectively for the remaining $2,000$ simulations (with $1,000$ each). For each simulation, we execute an allocation strategy based on a depth-k search tree for k equal to 1, 2, 3, 4, and 5 respectively. We then record important performance measures such as cumulative claimed utilities and cumulative number of transplants. Note that when k is equal to 1, the allocation strategy simply corresponds to selecting, among all possible choices available for the altruistic donor, the one that has the largest edge utility.

5.3. Simulation results

First, we report on the cumulative number of transplants achieved in simulated KPD programs with one altruistic donor and 100 incompatible donor-candidate pairs. We compare the average number of transplants across different values of k and under the three utility generating distributions. Table 1 provides summary comparison, in which we observe a consistent pattern where the number of transplants performed increases with k. This is true regardless of which distribution is used to generate edge utilities.

Table 1: *Summary of the average number of transplants performed (denoted by N) and the average cumulative utilities claimed (denoted by U_∞), by implementing a depth-k search tree-based allocation strategy on a simulated KPD program with one altruistic donor and* 100 *incompatible pairs. Edge utilities are generated from $U(1,1)$, $U(10,20)$, and $U(10,30)$; and edge probabilities are generated from $U(0.1,0.5)$. The summary is calculated over* $3,000$ *rounds of simulations, with* $1,000$ *simulations for each utility generating distribution. Note that for the choice $U(1,1)$, the claimed utility equals the number of completed transplants.*

	$u_e = 1$	$u_e \sim U(10,20)$		$u_e \sim U(10,30)$	
depth-k	mean N	mean N	mean U_∞	mean N	mean U_∞
$k=1$	3.16	3.18	55.99	3.18	80.22
$k=2$	8.17	6.65	112.21	6.19	149.58
$k=3$	8.70	7.87	128.93	7.63	176.78
$k=4$	8.74	8.26	133.54	7.99	181.92
$k=5$	8.89	8.41	134.45	8.29	185.43

Table 2: *Three correlation matrices for the total number of transplants performed in a depth-k search tree-based allocation strategy across different values of k. The entry at the ith row and the jth column represents the correlation between the total number of transplants when $k = i$ and that when $k = j$ when managing the same simulated KPD program (with one altruistic donor and 100 incompatible pairs). Matrix on the left: $u_e = 1$; matrix in the middle: $u_e \sim U(10,20)$; matrix on the right: $u_e \sim U(10,30)$.*

1	.34	.34	.33	.31
–	1	.67	.67	.66
–	–	1	.79	.73
–	–	–	1	.81
–	–	–	–	1

1	.27	.29	.26	.29
–	1	.50	.51	.48
–	–	1	.66	.58
–	–	–	1	.72
–	–	–	–	1

1	.29	.27	.29	.29
–	1	.46	.48	.48
–	–	1	.58	.57
–	–	–	1	.72
–	–	–	–	1

Another observation is that the extra benefit in the number of transplants through increasing k is diminishing as k gets large. For example, when edge utilities are generated from $U(10,20)$, increasing the value of k from 1 to 4 would almost triple the total number of transplants (on average from 3.18 to 8.26); however, further increasing k (from 4 to 5) appears to have very limited effects.

In terms of comparing the cumulative claimed utility, Table 1 also demonstrates similar patterns to those observed above for comparing the number of transplants. These results suggest that $k = 3$ or 4 would provide a satisfactory solution in practice. Further investigation, however, with data from more KPD programs would be useful. Notice that when edge utilities are fixed at 1, the cumulative claimed utility is the same as the cumulative number of transplants.

Finally, we take a look at the correlation matrix among five variables; each variable represents the number of transplants performed when k is equal to each one of the five values. We anticipate that the correlation between variable 4 (the number of transplants achieved when $k = 4$) and variable 5 would be higher than the correlation between variable 1 and variable 5. Table 2 exactly unveils such a pattern in three correlation matrices (with each one corresponding to one utility generating distribution). Similar observations are also noted in the correlation matrices for the cumulative claimed utility.

6. Concluding remarks

In this paper, we have studied the problem of managing a KPD program with one altruistic donor. One important yet challenging part of this problem is to recognize various friction (as discussed in Section 2.1) inherent in the system and to guide the decision-making process by taking into account these uncertainties. Realizing the fact that a long pre-specified NEAD chain in practice can almost never be implemented as planned, we propose to initiate and extend such a chain in a sequential way by selecting potential transplant recipients one at a time. Each selection is made keeping in mind the associated long-term consequences so as to maximize the expected gain over a certain given number of moves. In order to do this efficiently and practically, we

construct a depth-k search tree for a KPD graph using a DFS-based algorithm. We then evaluate various choices available for each altruistic donor (or bridge donor) according to the calculation performed along that search tree, and recommend the choice with the greatest expected utility.

In the process of extending a NEAD chain, the bridge donor at the end of the current chain might be incompatible with a majority of the candidate population, which would significantly prolong the waiting time for that bridge donor to be matched with a present or future candidate. Furthermore, this long waiting time for the hard-to-match bridge donor may make him/her more likely to withdraw from the KPD pool and so terminate the NEAD chain. To partially avoid this unfortunate circumstance, some KPD programs have not allowed a blood type AB donor to become a bridge donor. Actually, this issue can be partially addressed in our proposed sequential allocation strategy, which incorporates the long-term consequences associated with each pair choice, and would in general tend to avoid choices that could lead to a hard-to-match bridge donor. As we have briefly mentioned in Section 3 and Section 4.2, one way to further address this issue is to assign each possible bridge donor a reasonable *base utility*. This base utility represents the potential "contribution" from a bridge donor; a hard-to-match bridge donor would be assigned a small base utility and an easy-to-match one would be given a larger value.

Although this paper has focused on managing a KPD program with one altruistic donor, the proposed approach can be generalized to incorporate multiple altruistic donors. One way to achieve this is to construct a depth-k search tree for each altruistic donor and use this tree to evaluate various allocations options available for that altruistic donor according to the method in Section 4.2. Among all allocations possible for these altruistic donors, we select a disjoint collection such that the overall expected utility can be maximized. More specifically, let $\mathscr{V}_a \equiv \{1,2,\cdots,m\}$ be m altruistic donors in a general KPD program. We denote by $\mathscr{A} \equiv \cup_{i=1}^{m}\{(i,j) \in \mathscr{E}\}$ the possible allocations available for these m altruistic donors. Each potential transplant, $(i,j) \in \mathscr{A}$, can be evaluated by its expected utility, which is calculated as $u^*_{ij} = u_{ij} + EV[\mathscr{G}_{[i]}(j)]$, where $\mathscr{G}_{[i]} \equiv \mathscr{G}(i)$. We then select from $\mathscr{A}$ a disjoint collection of edges (or transplants), in the sense that no two edges can share a common vertex, so as to maximize the sum of expected utilities. For those selected transplants, viable ones would result in actual operations and generate new bridge donors, and altruistic donors and incompatible pairs involved in non-viable transplants are recycled back to the KPD pool.

The above way of allocating multiple altruistic donors can be arranged in parallel with the selection of exchange sets (or cycles) among incompatible pairs. Let $\mathscr{S}_r$ be the collection of all exchange sets of size up to r among $(n-m)$ incompatible pairs (Li et al., 2013). Typically, in clinical KPD programs, one chooses $r = 3$, although larger exchanges could be considered as well. For each $S \in \mathscr{S}_r$, let EU_S represent its expected utility and let Y_S be a decision variable equal to 1 if S is selected and 0 if not; for each $(i,j) \in \mathscr{A}$, Z_{ij} is another decision variable whose value is 1 if (i,j) is chosen for a transplant and 0 otherwise; the expected utility of this potential transplant (i,j) is given

by u^*_{ij} as discussed above. By adopting a formulation that is similar to the one proposed in Li et al. (2013), we then manage such a KPD program by solving the following IP problem:

$$\max_{\{Y_S\},\{Z_{ij}\}} \left\{ \sum_{S \in \mathscr{S}_r} Y_S EU_S + \sum_{(i,j)\in\mathscr{A}} Z_{ij} u^*_{ij} \right\}, \tag{5}$$

$$\text{subject to} \sum_{S \in \mathscr{S}_r(l)} Y_S + \sum_{(i,j)\in\mathscr{A}(l)} Z_{ij} \leq 1, \forall l \in \mathscr{V}, \tag{6}$$

where, in (6), $\mathscr{S}_r(l)$ represents the exchange sets in $\mathscr{S}_r$ that contain l and $\mathscr{A}(l)$ similarly denotes a subset of transplants in $\mathscr{A}$ that involve l. Various extensions of this approach would be possible to allow fall back options for the exchanges and for the assignment of the altruistic donors. Such approaches are discussed in Li et al. (2013). This approach, and extensions of it, provide an alternative to the simultaneous selection of chains and cycles as is traditionally done in the match runs of a KPD; see for example, Roth et al. (2006), Gentry et al. (2009), and Ashlagi et al. (2011).

The mechanism of a NEAD chain allows the altruism from a single altruistic donor to benefit a potentially large number of patients, but it does so exclusively for patients recruiting a willing but incompatible living donor. This mechanism excludes patients without a designated living donor and who are therefore placed on a deceased-donor waiting list. Among those patients who would benefit from this NEAD chain, approximately 73% of them are white; whereas those who would not benefit from this mechanism form a 52% non-white population (Segev et al., 2008). On the other hand, not all altruistic (or bridge) donors are well suited for initiating and extending chains among a pool of incompatible pairs. For example, consider a KPD pool in which an altruistic donor may not be in a good position (because of either incompatibility or poor utility) to be matched up with any candidate. In this case, rather than placing this altruistic donor in a waiting "mode" for a potentially long time, redirecting him/her to a deceased-donor waiting list, where a compatible patient with potentially good transplant outcomes might be identified rather easily, appears a more suitable alternative.

To decide whether an altruistic donor is better off initiating a NEAD chain or donating directly to someone on a deceased-donor waiting list, we could evaluate an altruistic donor by the utility expected to be achieved if this donor is chosen to initiate a chain. To be precise, we may first perform the calculation according to formula (4) as in Section 4.1 to evaluate the expected utility for each altruistic donor, i.e. $\{EV[\mathscr{G}(i)] : i \in \mathscr{V}_a\}$. The result from this evaluation could then be used to assess the suitability of assigning each altruistic donor to a deceased-donor waiting list; a relatively high value of $EV[\mathscr{G}(i)]$ would recommend reserving altruistic donor i for extending a NEAD chain while a comparatively low value of $EV[\mathscr{G}(i)]$ would indicate a transplant to someone waiting for a deceased-donor kidney. It is worth noting that different ways of assigning edge utilities and probabilities could be adopted in calculating $\{EV[\mathscr{G}(i)]$

: $i \in \mathcal{V}_a\}$. This would provide extra benefit in allowing more control over what kidneys in general are distributed to a deceased-donor waiting list. For example, if each edge is assigned an equal utility while the edge probability remains representing the likelihood of that edge being viable, then altruistic (or bridge) donors who are less compatible with candidates in the current KPD pool would be more likely to be directed to a deceased-donor waiting list. In addition to evaluating an altruistic donor against the current KPD pool, it is often rational in practice to perform the evaluation against the deceased-donor waiting list. For example, Blood Type O or B candidates frequently wait a year or more longer on the deceased-donor waiting list before receiving a kidney transplant than do Blood Type A or AB candidates. Therefore a Blood Type O or B altruistic or bridge donor might be argued to have higher utility. So might someone who matches to a pediatric candidate. Similarly an altruistic or bridge donor that might match to a sensitized wait-listed candidate might have a higher utility, but a lower probability of progressing to transplant.

Acknowledgements

This work was supported in part by a grant from the National Institutes of Health (NIH) CTSA at the University of Michigan 2UL1TR000433-06 and by the NIH grant 1R01-DK093513. The authors thank the support from the Michigan Institute for Clinical and Health Research (MICHR), Michigan School of Public Health, Scientific Registry of Transplant Recipients, and National Institute of General Medical Sciences (NIGMS). Last but not least, the authors would like to thank the editor and referees whose comments helped to improve this manuscript.

References

Abraham, D. J., Blum, A. and Sandholm, T. (2007). Clearing algorithms for barter exchange markets: Enabling nationwide kidney exchanges. *Ec'07: Proceedings of the Eighth Annual Conference on Electronic Commerce*, 295–304.

Ashlagi, I., Gilchrist, D. S., Roth, A. E. and Rees, M. A. (2011). Nonsimultaneous chains and dominos in kidney-paired donation-revisited. *American Journal of Transplantation*, 11(5): 984–994.

Evans, R. W., Manninen, D. L., Garrison, L. P., Hart, L. G., Blagg, C. R., Gutman, R. A., Hull, A. R. and Lowrie, E. G. (1985). The quality of life of patients with end-stage renal disease. *New England Journal of Medicine*, 312(9): 553–559.

Gentry, S. E., Montgomery, R. A., Swihart, B. J. and Segev, D. L. (2009). The roles of dominos and nonsimultaneous chains in kidney paired donation. *American Journal of Transplantation*, 9(6): 1330–1336.

Li, Y. (2012). Optimization and simulation of kidney paired donation programs (doctoral dissertation). *University of Michigan, Ann Arbor*.

Li, Y., Kalbfleisch, J. D., Song, P. X.-K., Zhou, Y., Leichtman, A. B. and Rees, M. A. (2013). Optimal decisions for organ exchanges in a kidney paired donation program. *Statistics in Biosciences*, 6(1), 85–104.

Maiers, M., Gragert, L. and Klitz, W. (2007). High-resolution hla alleles and haplotypes in the united states population. *Human Immunology*, 68(9): 779–788.

Montgomery, R. A., Gentry, S. E., Marks, W. H., Warren, D. S., Hiller, J., Houp, J., Zachary, A. A., Melancon, J. K., Maley, W. R., Rabb, H., Simpkins, C. and Segev, D. L. (2006). Domino paired kidney donation: a strategy to make best use of live non-directed donation. *The Lancet*, 368(9533): 419–421.

Rapaport, F. T. (1986). The case for a living emotionally related international kidney donor exchange registry. *Transplantation Proceedings*, 18: 5–9.

Rees, M. A., Kopke, J. E., Pelletier, R. P., Segev, D. L., Rutter, M. E., Fabrega, A. J., Rogers, J., Pankewycz, O. G., Hiller, J., Roth, A. E., Sandholm, T., Ünver, M. U. and Montgomery, R. A. (2009). A nonsimultaneous, extended, altruistic-donor chain. *New England Journal of Medicine*, 360(11): 1096–1101.

Roth, A. E., Sönmez, T. and Ünver, M. U. (2004). Kidney exchange. *Quarterly Journal of Economics*, 119(2): 457–488.

Roth, A. E., Sönmez, T. and Ünver, M. U. (2007). Efficient kidney exchange: Coincidence of wants in markets with compatibility-based preferences. *American Economic Review*, 97(3): 828–851.

Roth, A. E., Sönmez, T., Ünver, M. U., Delmonico, F. L. and Saidman, S. L. (2006). Utilizing list exchange and nondirected donation through 'chain' paired kidney donations. *American Journal of Transplantation*, 6(11): 2694–2705.

Russell, J. D., Beecroft, M. L., Ludwin, D. and Churchill, D. N. (1992). The quality of life in renal transplantation-a prospective study. *Transplantation*, 54(4): 656–660.

Schaubel, D. E., Wolfe, R. A., Sima, C. S. and Merion, R. M. (2009). Estimating the effect of a time-dependent treatment by levels of an internal time-dependent covariate: Application to the contrast between liver wait-list and posttransplant mortality. *Journal of the American Statistical Association*, 104(485): 49–59.

Segev, D. L., Kucirka, L. M., Gentry, S. E. and Montgomery, R. A. (2008). Utilization and outcomes of kidney paired donation in the united states. *Transplantation*, 86(4): 502–510.

Wolfe, R. A., Ashby, V. B., Milford, E. L., Ojo, A. O., Ettenger, R. E., Agodoa, L. Y. C., Held, P. J. and Port, F. K. (1999). Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New England Journal of Medicine*, 341(23): 1725–1730.

Wolfe, R. A., McCullough, K. P., Schaubel, D. E., Kalbfleisch, J. D., Murray, S., Stegall, M. D. and Leichtman, A. B. (2008). Calculating life years from transplant (lyft): Methods for kidney and kidney-pancreas candidates. *American Journal of Transplantation*, 8(4p2): 997–1011.

SORT 38 (1) January-June 2014, 73-88

The asymptotic relative efficiency and the ratio of sample sizes when testing two different null hypotheses

Guadalupe Gómez*, [1] and Moisés Gómez-Mateu[1]

Abstract

Composite endpoints, consisting of the union of two or more outcomes, are often used as the primary endpoint in time-to-event randomized clinical trials. Previously, Gómez and Lagakos provided a method to guide the decision between using a composite endpoint instead of one of its components when testing the effect of a treatment in a randomized clinical trial. Consider the problem of testing the null hypotheses of no treatment effect by means of either the single component or the composite endpoint. In this paper we prove that the usual interpretation of the asymptotic relative efficiency as the reciprocal ratio of the sample sizes required for two test procedures, for the same null and alternative hypothesis, and attaining the same power at the same significance level, can be extended to the test procedures considered here for two different null and alternative hypotheses. A simulation to study the relationship between asymptotic relative efficiency and finite sample sizes is carried out.

MSC: 62N03, 62P10

Keywords: Asymptotic relative efficiency, composite endpoint, logrank test, sample size, simulation, survival analysis.

1. Introduction

In clinical trials research, one of the most important issues that investigators have to solve at the design stage of the study is the appropriate choice of the primary endpoint. Composite endpoints (CE) consisting of the union of two or more outcomes are commonly used as primary endpoints. For example, in the cardiovascular area the

* Corresponding author e-mail: lupe.gomez@upc.edu

[1] Departament d'Estadística i I.O., Universitat Politècnica de Catalunya, Jordi Girona 1–3, 08034. Barcelona, Spain.

Received: May 2013

Accepted: January 2014

relevant endpoint of death is often combined with other additional endpoints such as myocardial infarction, stroke or hospitalization. Pros and cons on the use of CE have been extensively discussed (Freemantle et al., 2003; Ferreira-González et al., 2007, among many others). One of the main advantages of using a CE relies in the fact that by means of a CE the problem of multiplicity is adequately addressed and the bias associated with competing risks (Wittkop et al., 2010) is avoided. Also, with a CE the number of observed events will be higher and, hopefully, the power of the test will increase. However, as it has been discussed (Montori et al., 2005) and shown in Gómez and Lagakos (2013), adding inappropriate components to the relevant endpoint might actually lead to a decrease in the power of the test statistic, consequently having a larger chance to fail in detecting a real effect of the treatment under study.

Gómez and Lagakos (2013) developed a methodology to help to decide when it is worthwhile to base the analysis on the composite endpoint $\mathscr{E}_* = \mathscr{E}_1 \cup \mathscr{E}_2$ where $\mathscr{E}_1$ and $\mathscr{E}_2$ are two candidate relevant endpoints to evaluate the effect of a treatment instead of sticking to one of them, $\mathscr{E}_1$, say. In order to do so, they compared how more efficient than $\mathscr{E}_1$ would $\mathscr{E}_*$ be to justify its use. Let H_0 be the null hypothesis of no treatment effect evaluated on $\mathscr{E}_1$ and denote by H_a an alternative hypothesis, for instance, claiming to delay the event $\mathscr{E}_1$. Analogously, define H_0^* and H_a^* the null and alternative hypotheses if the treatment effect is to be evaluated on $\mathscr{E}_*$. Since when comparing two treatment groups based on time-to-event endpoints, the primary analysis would be based, very commonly, on a logrank test, their method considers the logrank test Z to test H_0 versus H_a and the logrank test Z_* to test H_0^* versus H_a^*. The asymptotic relative efficiency (ARE) of Z_* versus Z is the measure proposed to choose between $\mathscr{E}_1$ and $\mathscr{E}_*$, with values larger than 1 in favour of $\mathscr{E}_*$. This relative measure can be computed as $(\mu_*/\mu)^2$ where μ and μ_* are, respectively, the asymptotic means of Z and Z_*, under alternative contiguous hypotheses to H_0 and H_0^*. The purpose of this paper is to prove that the usual interpretation of the ARE, as the ratio of sample sizes, n and n_*, needed to attain the same power for a given significance level, still holds even though two different sets of hypothesis (H_0 versus H_a and H_0^* versus H_a^*) are compared.

To clarify the purpose of our investigation consider the following. If we were to test H_0 versus H_a with two different test statistics S_n and T_m, Pitman's relative efficiency would be defined as the ratio m/n, where n and m are the required sample sizes for S_n and T_m, respectively, to attain the same power for a given significance level. Furthermore, if both S_n and T_m are asymptotically normal with unit variance and means μ_S and μ_T, it can be proved that Pitman's ARE corresponds to the square of the ratio of the noncentrality parameters, that is $(\mu_S/\mu_T)^2$. Gómez and Lagakos' method compares the logrank statistics: Z and Z_* derived for two different set of hypotheses H_0 versus H_a and H_0^* versus H_a^* and do so using, as definition of the ARE, the ratio $(\mu_*/\mu)^2$ where μ and μ_* are, respectively, the asymptotic means of Z and Z_*, under alternative contiguous hypotheses to H_0 and H_0^*.

This paper is organized as follows. In Section 2 the notation, assumptions and main results from Gómez and Lagakos' paper are introduced. Section 3 establishes

the limiting relationship between ARE and sample sizes and proves that the usual interpretation of the ARE as the ratio of sample sizes holds. Section 4 presents a simulation to study under which conditions and for finite sample sizes, the relationship $\mathrm{ARE}(Z_*, Z) = (\mu_*/\mu)^2 = n/n_*$ holds where n and n_* are the needed sample sizes for Z and Z_*, respectively, to attain the same power for a given significance level. Section 5 concludes the paper with a discussion.

2. Notation, the logrank test and the asymptotic relative efficiency

2.1. The logrank tests for the relevant and for the composite endpoints

Assume that we have a two-arm study involving random assignment to an active ($X = 1$) or control treatment ($X = 0$) aiming to prove the efficacy of the new active treatment. The effect of treatment is to be evaluated on the time $T_1^{(j)}$ to a relevant event $\mathscr{E}_1$, where the superscript j indicates the treatment group ($j = 0$ for the control group and $j = 1$ for the treatment group). Let $\lambda_1^{(j)}(t)$ denote the hazard function of $T_1^{(j)}$ ($j = 0, 1$). The null hypothesis of no effect is given by $H_0 : \mathrm{HR}_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t) = 1$ and the alternative that the new treatment improves survival by $H_a : \mathrm{HR}_1(t) < 1$. The logrank test Z is used to test that the new treatment improves survival.

Assume now that an additional endpoint $\mathscr{E}_2$ is considered as component of the primary endpoint and the composite endpoint $\mathscr{E}_* = \mathscr{E}_1 \cup \mathscr{E}_2$ is to be used, instead, to prove the efficacy of the new treatment. The effect of treatment would then be evaluated on the time $T_*^{(j)}$ to $\mathscr{E}_*$ where $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$ and $T_2^{(j)}$ stands for the time to $\mathscr{E}_2$ ($j = 0, 1$). Let $\lambda_2^{(j)}(t)$ and $\lambda_*^{(j)}(t)$ denote, respectively, the hazard functions of $T_2^{(j)}$ and $T_*^{(j)}$ ($j = 0, 1$). The treatment effect on $\mathscr{E}_*$ would then be tested with the logrank test Z_* to compare $H_0^* : \mathrm{HR}_*(t) = \lambda_*^{(1)}(t)/\lambda_*^{(0)}(t) = 1$ versus $H_a^* : \mathrm{HR}_*(t) < 1$.

Observation of endpoints $\mathscr{E}_1$ and $\mathscr{E}_2$ depends on whether or not they include a terminating event and yield four different situations referred, in Gómez and Lagakos (2013), as Cases 1, 2, 3 and 4. In this paper we assume that the additional endpoint does not include a terminating event, which corresponds to Case 1 when neither the relevant nor the additional endpoint includes a terminating event, and Case 3, when the relevant endpoint includes a terminating event.

Schoenfeld (1981) studies the asymptotic behaviour of the logrank statistic and proves that under the null hypothesis of no treatment difference, the logrank is asymptotically $N(0, 1)$ and, under a sequence of alternatives contiguous to the null, the logrank is asymptotically normal with unit variance and finite mean. Gómez and Lagakos apply Schoenfeld's results and proceed as follows. They consider $\lambda_1^{(0)}(t)$ as fixed and define a sequence of alternatives $H_{a,n}$ consisting of instantaneous hazard functions close enough to $\lambda_1^{(0)}(t)$, for instance taking $\lambda_{1,n}^{(1)}(t) = \lambda_1^{(0)}(t) e^{g(t)/\sqrt{n}}$ for some $g(t)$ function. These sequence of alternatives, formulated equivalently as $\mathrm{HR}_{1,n}(t) = e^{g(t)/\sqrt{n}}$, include pro-

portional hazard alternatives, i.e, taking $g(t) = \beta$ for a fixed real value β. Logrank Z is asymptotically $N(0,1)$ under the null hypothesis of no treatment difference (H_0 : $\mathrm{HR}_1(t) = 1$) and asymptotically normal with unit variance and mean μ given in equation (1) under the sequence of alternatives $H_{a,n} : \mathrm{HR}_{1,n}(t) = e^{g(t)/\sqrt{n}} < 1$. Analogously, fix $\lambda_*^{(0)}(t)$ and define $H_0^* : \mathrm{HR}_*(t) = 1$ and the sequence of alternatives $H_{a,n}^* : \mathrm{HR}_{*,n}(t) = e^{g_*(t)/\sqrt{n}} < 1$ for a given function $g_*(t)$. It follows that Z^* is asymptotically $N(0,1)$ under H_0^* and asymptotically normal with unit variance and mean μ_* given in equation (2) under the sequence $H_{a,n}^*$. The asymptotic means of Z and Z^* are given by

$$\mu = \frac{\int_0^\infty g(t)p(t)[1-p(t)]\mathrm{Pr}_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)dt}{\sqrt{\int_0^\infty p(t)[1-p(t)]\mathrm{Pr}_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)dt}}, \tag{1}$$

$$\mu_* = \frac{\int_0^\infty g_*(t))p_*(t)[1-p_*(t)]\mathrm{Pr}_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]\mathrm{Pr}_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)dt}}, \tag{2}$$

where $U = \min\{T_1, C\}$ (in Cases 1 and 3) and $U_* = \min\{T_*, C\}$ denote the observed outcome; C denotes the censoring time; $p(t) = \mathrm{Pr}_{H_0}\{X = 1 | U \geq t\}$ and $p_*(t) = \mathrm{Pr}_{H_0^*}\{X = 1 | U_* \geq t\}$ are the null probabilities that someone at risk at time t is in treatment group 1; $\mathrm{Pr}_{H_0}\{U \geq t\}$ and $\mathrm{Pr}_{H_0^*}\{U_* \geq t\}$ are the null probabilities that someone is still at risk at time t and $\mathrm{Pr}_{H_0}\{U \geq t\}\lambda_1^{(0)}(t)$ and $\mathrm{Pr}_{H_0^*}\{U_* \geq t\}\lambda_*^{(0)}(t)$ correspond to the probabilities, under the null hypothesis, of observing events $\mathscr{E}_1$ and $\mathscr{E}_*$, respectively, by time t.

2.2. Asymptotic relative efficiency

Efficiency calculations throughout the paper will assume that end-of-study censoring at time τ ($\tau = 1$ without loss of generality) is the only non-informative censoring cause for both groups; this assumption indirectly implies that the censoring mechanism is the same for both groups. It is as well assumed that the hazard functions $\lambda_1^{(j)}(t)$ and $\lambda_2^{(j)}(t)$ ($j = 0, 1$) are proportional, that is, $\mathrm{HR}_1(t) = \mathrm{HR}_1$ and $\mathrm{HR}_2(t) = \mathrm{HR}_2$, for all t, where $\mathrm{HR}_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t)$ and $\mathrm{HR}_2(t) = \lambda_2^{(1)}(t)/\lambda_2^{(0)}(t)$ are the hazard ratios between $T_1^{(0)}$ and $T_1^{(1)}$ and between $T_2^{(0)}$ and $T_2^{(1)}$, respectively. Note that although we are assuming that the hazard functions $\lambda_1^{(j)}(t)$ and $\lambda_2^{(j)}(t)$ ($j = 0, 1$) are proportional, this does not imply the proportionality of hazards $\lambda_*^{(0)}(t)$ and $\lambda_*^{(1)}(t)$ for the composite endpoint T_* (see Figure 1).

To assess the difference in efficiency between using logrank test Z, based on the relevant endpoint $\mathscr{E}_1$, and logrank test Z_*, based on the composite endpoint $\mathscr{E}_*$, Gómez and Lagakos base their strategy on the behaviour of the asymptotic relative efficiency (ARE) of Z_* versus Z. The ARE is a measure of the relative power of two tests that can

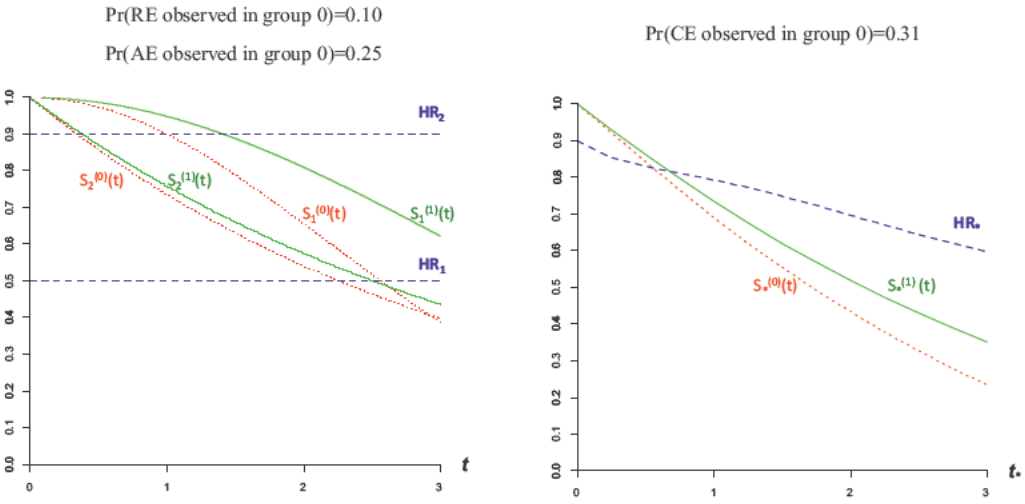

Figure 1: *Survival and hazard ratio for the relevant endpoint (RE), T_1, for the additional endpoint (AE), T_2 and for the composite endpoint (CE), $T_* = \min\{T_1, T_2\}$. $T_1 \sim$ Weibull with shape parameter $\beta_1 = 2$ (increasing hazard) for treatment groups 0 and 1 and $T_2 \sim$ Weibull with shape parameter $\beta_2 = 1$ (constant hazard) for treatment groups 0 and 1. Scale parameters for T_1 and T_2 have been calculated such that* $\Pr\{T_1$ *observed in group 0*$\}$*=0.1,* $\Pr\{T_2$ *observed in group 0*$\}$*=0.25,* $\mathrm{HR}_1 = 0.5$*,* $\mathrm{HR}_2 = 0.9$ *and Spearman's* $\rho(T_1, T_2) = 0.45$ *assuming Frank's copula between T_1 and T_2. Considering the RE as a terminating event (case 3), in this setting* $\mathrm{ARE}(Z_*, Z) = 0.21$.

be interpreted, when the two tests are for the same null and alternative hypothesis, as the ratio of the required sample sizes to detect a specific treatment effect to attain the same power for a given significance level (Lehmann and Romano, 2005). In this case, a value of ARE$=$ 0.6 would mean that we only need 60% as many cases to reach a given power if we use $\mathscr{E}_1$ as we would need if we used $\mathscr{E}_*$. Whenever the tests under consideration, Z and Z_*, are asymptotically N(0,1) under H_0 and H_0^*, respectively, and asymptotically normal with variance 1 under a sequence of contiguous alternatives to the null hypothesis, a different definition for Pitman's relative efficiency as the square of the ratio of the non-centrality parameters μ and μ_* is appropriate

$$\mathrm{ARE}(Z_*, Z) = \left(\frac{\mu_*}{\mu}\right)^2, \tag{3}$$

where μ and μ_* are to be replaced by expressions (1) and (2).

Before providing the expression that is being used to evaluate the ARE, and for the sake of clarity, we enumerate the assumptions that have been taken into account:

- End-of-study censoring at time τ is the only non-informative censoring cause for both groups.
- The additional endpoint does not include a terminating event.

- The hazard ratios between $T_1^{(0)}$ and $T_1^{(1)}$ and between $T_2^{(0)}$ and $T_2^{(1)}$ are proportional, that is, $\mathrm{HR}_1(t) = \lambda_1^{(1)}(t)/\lambda_1^{(0)}(t) = \mathrm{HR}_1$ and $\mathrm{HR}_2(t) = \lambda_2^{(1)}(t)/\lambda_2^{(0)}(t) = \mathrm{HR}_2$ for all t.
- Effect of treatment on $\mathscr{E}_1$ is tested establishing $H_0 : \mathrm{HR}_1 = 1$ versus a sequence of alternatives $H_{a,n} : \lambda_{1,n}^{(1)}(t) = \lambda_1^{(0)}(t)e^{g(t)/\sqrt{n}}$ for some $g(t)$ function. Note that $g(t)/\sqrt{n} = \log\{\lambda_{1,n}^{(1)}(t)/\lambda_1^{(0)}(t)\}$.
- Effect of treatment on $\mathscr{E}_*$ is tested establishing $H_0^* : \mathrm{HR}_*(t) = 1$ versus a sequence of alternatives $H_{a,n}^* : \mathrm{HR}_{*,n}(t) = e^{g_*(t)/\sqrt{n}} < 1$ for a given function $g_*(t)$. Note that $g_*(t)/\sqrt{n} = \log\{\mathrm{HR}_{*,n}(t)\}$.

Under the above assumptions expression (3) becomes

$$\mathrm{ARE}(Z_*, Z) = \frac{\left(\int_0^1 \log\left\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\right\} f_*^{(0)}(t)dt\right)^2}{\left(\log\left\{\mathrm{HR}_1\right\}\right)^2 \left(\int_0^1 f_*^{(0)}(t)dt\right)\left(\int_0^1 f_1^{(0)}(t)dt\right)}, \tag{4}$$

where $f_1^{(0)}(t)$ and $f_*^{(0)}(t)$ are the density functions of $T_1^{(0)}$ and $T_*^{(0)}$, respectively.

Remark The density function $f_*^{(0)}(t)$ is the density of the $T_*^{(0)} = \min\{T_1^{(0)}, T_2^{(0)}\}$, computed from the joint density between $T_1^{(0)}$ and $T_2^{(0)}$, which itself is built from the marginals of $T_1^{(0)}$ and $T_2^{(0)}$ by means of a bivariate copula.

3. Relationship between ARE and sample sizes

We start establishing that if the hazard ratios for $T_1^{(j)}$ $(j = 0, 1)$ and for $T_2^{(j)}$ $(j = 0, 1)$ approach the unity as n gets large, so does the hazard ratio of the minimum $T_*^{(j)}$ between $T_1^{(j)}$ and $T_2^{(j)}$ $(j = 0, 1)$.

Lemma 1 *Given two sequences of hazard ratios* $\{\mathrm{HR}_{1,n}(t) = \lambda_{1,n}^{(1)}(t)/\lambda_1^{(0)}(t)\}$ *and* $\{\mathrm{HR}_{2,n}(t) = \lambda_{2,n}^{(1)}(t)/\lambda_2^{(0)}(t)\}$, *both converging uniformly to 1 as* $n \to \infty$, *the sequence corresponding to the hazard ratio of* $T_*^{(j)} = \min\{T_1^{(j)}, T_2^{(j)}\}$, *namely* $\{\mathrm{HR}_{*,n}(t) = = \lambda_{*,n}^{(1)}(t)/\lambda_*^{(0)}(t)\}$, *tends to* 1 *as* $n \to \infty$. *In particular, this lemma holds whenever* $\log(\lambda_{k,n}^{(1)}(t)/\lambda_k^{(0)}(t)\}) = O(n^{-1/2})$, *which in turn, is true if* $\log(\lambda_{k,n}^{(1)}(t)/\lambda_k^{(0)}(t)\}) = = g_k(t)/\sqrt{n}$, *for any bounded real function* $g_k(t)$ $(k = 1, 2)$.

Proof 1 It follows immediately that for fixed t, $\lim_{n\to\infty} \lambda_{1,n}^{(1)}(t) = \lambda_1^{(0)}(t)$ and $\lim_{n\to\infty} \lambda_{2,n}^{(1)}(t) = \lambda_2^{(0)}(t)$. Furthermore, it follows that the corresponding densities and

survival functions $f_{1,n}^{(1)}(t)$, $f_{2,n}^{(1)}(t)$, $S_{1,n}^{(1)}(t)$ and $S_{2,n}^{(1)}(t)$, converge to $f_1^{(0)}(t)$, $f_2^{(0)}(t)$, $S_1^{(0)}(t)$ and $S_2^{(0)}(t)$, respectively. Taking into account that the survival function of the minimum, $S_{*,n}^{(1)}(t)$ is expressed in terms of the marginal survival functions $S_{1,n}^{(1)}(t)$ and $S_{2,n}^{(1)}(t)$ of $T_1^{(1)}$ and $T_2^{(1)}$ via a copula C, that is,
$S_{*,n}^{(1)}(t) = C(S_{1,n}^{(1)}(t), S_{2,n}^{(1)}(t))$, it remains to prove that $\lim_{n\to\infty} S_{*,n}^{(1)}(t) = S_*^{(0)}(t)$. This result will imply that

$\lim_{n\to\infty} f_{*,n}^{(1)}(t) = f_*^{(0)}(t)$, $\lim_{n\to\infty} \lambda_{*,n}^{(1)}(t) = \lambda_*^{(0)}(t)$ and hence the sequence $\mathrm{HR}_{*,n}(t) \to 1$ as $n \to \infty$, as we wanted to prove.
The convergence of $S_{*,n}^{(1)}(t)$ to $S_*^{(0)}(t)$ is guaranteed by the convergence of $S_{1,n}^{(1)}(t)$ and $S_{2,n}^{(1)}(t)$ to $S_1^{(0)}(t)$ and $S_2^{(0)}(t)$, respectively, together with the fact that bivariate copulas C are bivariate distribution functions with uniform marginals. The reader is referred to Lindner and Szimayer (2005) for the corresponding technical proofs. □

Proposition 1 *Consider two test procedures ϕ_n and ϕ_n^* to test $H_0 : \mathrm{HR}_1(t) = 1$ against $H_{a,n} : \mathrm{HR}_{1,n}(t) < 1$ and $H_0^* : \mathrm{HR}_*(t) = 1$ against $H_{a,n}^* : \mathrm{HR}_{*,n}(t) < 1$, respectively. Let n and n_* be the sample sizes required for ϕ_n and ϕ_n^*, respectively, to have power at least Π at level α. Assume the sequences $\phi = \{\phi_n\}$ and $\phi^* = \{\phi_n^*\}$ are based on the logrank statistics Z and Z^*, respectively, converging, to Normal $(\mu, 1)$ and Normal $(\mu_*, 1)$ with μ and μ_* given in (1) and (2), under sequences of local alternatives $\mathrm{HR}_{k,n}(t)$ $(k = 1, 2)$ converging uniformly to 1 as $n \to \infty$. Given $0 < \alpha < \Pi < 1$,*

$$\lim_{\substack{\mathrm{HR}_{1,n}(t)\to 1 \\ \mathrm{HR}_{2,n}(t)\to 1}} \frac{n}{n_*} = \mathrm{ARE}(Z_*, Z).$$

The usual interpretation of the ARE as the reciprocal ratio of the sample sizes holds even when two different sets of hypotheses (H_0 versus $H_{a,n}$ and H_0^* versus $H_{a,n}^*$) are tested. As a consequence of this proposition, the interpretation of the ARE is the following. If $\mathrm{ARE}(Z_*, Z) = 0.7$, then, asymptotically, we only need 70% as many cases to attain a given power if we use Z as we would need if we used Z_*.

Proof 2 By Lemma 1, uniform convergence to 1 of $\{\mathrm{HR}_{1,n}(t)\}$ and $\{\mathrm{HR}_{2,n}(t)\}$ imply that $\lim \mathrm{HR}_{*,n}(t) \to 1$. Under the sequence of contiguous alternatives to the null $H_{a,n}$: $\{\mathrm{HR}_{1,n}(t) = \lambda_{1,n}^{(1)}(t)/\lambda_1^{(0)}(t)\} \to 1$ and $H_{a,n}^* : \{\mathrm{HR}_{*,n}(t) = \lambda_{*,n}^{(1)}(t)/\lambda_*^{(0)}(t)\} \to 1$, both Z and Z^* are asymptotically $N(\mu, 1)$ and $N(\mu_*, 1)$, respectively. The power function for a one-sided test with size α is therefore given, respectively, by

$$\Pi_1 = \lim_{n\to\infty} \mathrm{Prob}\{Z < z_{1-\alpha} | H_{a,n}\} = 1 - \Phi(-z_{1-\alpha} + \mu)$$

$$\Pi_* = \lim_{n\to\infty} \mathrm{Prob}\{Z_* < z_{1-\alpha} | H_{a,n}^*\} = 1 - \Phi(-z_{1-\alpha} + \mu_*) \tag{5}$$

where Φ is the distribution function of the standard normal and $z_{1-\alpha}$ is the standard normal quantile corresponding to the left tail probability α. It immediately follows that $\Pi_1 = \Pi_*$ is equivalent to $\mu = \mu_*$.

The equivalence of powers ($\Pi_1 = \Pi_*$) implies that $\mu = \mu_*$, given by (1) and (2). Equivalently

$$\left(\frac{\mu_*}{\mu}\right)^2 = 1 \quad \Longleftrightarrow \quad \left(\frac{\frac{\int_0^\infty g(t)p(t)[1-p(t)]\Pr_{H_0}\{U\geq t\}\lambda_1^{(0)}(t)dt}{\sqrt{\int_0^\infty p(t)[1-p(t)]\Pr_{H_0}\{U\geq t\}\lambda_1^{(0)}(t)dt}}}{\frac{\int_0^\infty g_*(t)p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_*\geq t\}\lambda_*^{(0)}(t)dt}{\sqrt{\int_0^\infty p_*(t)[1-p_*(t)]\Pr_{H_0^*}\{U_*\geq t\}\lambda_*^{(0)}(t)dt}}} \right)^2 = 1. \tag{6}$$

Since

$$p(t) = \frac{\Pr_{H_0}\{U \geq t | X = 1\}\pi}{\Pr_{H_0}\{U \geq t\}} = \frac{\Pr_{H_0}\{U^{(j)} \geq t\}\pi}{\Pr_{H_0}\{U \geq t\}}$$

where $\pi = \Pr_{H_0}\{X = 1\}$, we have

$$p(t)(1-p(t))\Pr_{H_0}\{U \geq t\} = \frac{\Pr_{H_0}\{U^{(1)} \geq t\}\pi \Pr_{H_0}\{U^{(0)} \geq t\}(1-\pi)}{\Pr_{H_0}\{U^{(0)} \geq t\}(1-\pi) + \Pr_{H_0}\{U^{(1)} \geq t\}\pi}.$$

Based on the stated assumptions, because $T_1^{(j)}$ is right-censored by the end-of-study at time τ, and under the null hypothesis of no effect ($S_1^{(0)}(t) = S_1^{(1)}(t)$), we have $\Pr_{H_0}\{U^{(j)} \geq t\} = S_1^{(0)}(t)1\{[0,1]\}(t)$, for $j = 0, 1$. Replacing in (1), the noncentrality parameter μ becomes

$$\mu = \frac{\sqrt{\pi(1-\pi)}\int_0^1 g(t)S_1^{(0)}(t)\lambda_1^{(0)}(t)dt}{\sqrt{\int_0^1 S_1^{(0)}(t)\lambda_1^{(0)}(t)dt}} = \frac{\sqrt{\pi(1-\pi)}\int_0^1 g(t)f_1^{(0)}(t)dt}{\sqrt{\int_0^1 f_1^{(0)}(t)dt}}$$

where $f_1^{(0)}(t)$ is the marginal density function for $T_1^{(0)}$. Analogously, it can be seen that

$$\mu_* = \frac{\sqrt{\pi(1-\pi)}\int_0^1 g_*(t)f_*^{(0)}(t)dt}{\sqrt{\int_0^1 f_*^{(0)}(t)dt}}$$

where $f_*^{(0)}(t)$ is the density function for $T_*^{(0)}$. The reader is addressed to the online supporting material of Gómez and Lagakos paper for other technical details.

If we would replace $g(t)$ and $g_*(t)$ by $\sqrt{n}\log\left(\frac{\lambda_{1,n}^{(1)}(t)}{\lambda_1^{(0)}(t)}\right) = \sqrt{n}\log(\mathrm{HR}_1)$ and $\sqrt{n_*}\log\left(\frac{\lambda_{*,n}^{(1)}(t)}{\lambda_*^{(0)}(t)}\right)$, respectively, equality (6), after cancelling $\pi(1-\pi)$, becomes equal to

$$\lim_{\substack{\mathrm{HR}_{1,n}(t)\to 1\\ \mathrm{HR}_{2,n}(t)\to 1}} \frac{\sqrt{n_*}}{\sqrt{n}} \frac{\frac{\int_0^1 \log\left\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\right\} f_*^{(0)}(t)dt}{\sqrt{\int_0^1 f_*^{(0)}(t)dt}}}{\log(\mathrm{HR}_1)\sqrt{\int_0^1 f_1^{(0)}(t)dt}} = 1$$

which in turn is equivalent to

$$\lim_{\substack{\mathrm{HR}_{1,n}(t)\to 1\\ \mathrm{HR}_{2,n}(t)\to 1}} \frac{n}{n_*} = \frac{\left(\int_0^1 \log\left\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\right\} f_*^{(0)}(t)dt\right)^2}{(\log(\mathrm{HR}_1))^2 \left(\int_0^1 f_*^{(0)}(t)dt\right)\left(\int_0^1 f_1^{(0)}(t)dt\right)} \tag{7}$$

and it follows that $\mathrm{ARE}(Z_*, Z) = \lim_{\substack{\mathrm{HR}_{1,n}(t)\to 1\\ \mathrm{HR}_{2,n}(t)\to 1}} \frac{n}{n_*}$, as we wanted to prove. □

Note that (7) implies

$$\frac{\left(\int_0^1 \log\left\{\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)\right\} f_*^{(0)}(t)dt\right)^2}{(\log(\mathrm{HR}_1))^2 \left(\int_0^1 f_*^{(0)}(t)dt\right)^2} = \lim_{\substack{\mathrm{HR}_{1,n}(t)\to 1\\ \mathrm{HR}_{2,n}(t)\to 1}} \frac{n\left(\int_0^1 f_1^{(0)}(t)dt\right)}{n_*\left(\int_0^1 f_*^{(0)}(t)dt\right)} \approx \frac{\text{expected number } \mathscr{E}_1}{\text{expected number } \mathscr{E}_*}$$

and whenever $\lambda_*^{(1)}(t)/\lambda_*^{(0)}(t)$ is approximately constant and equal to HR_*, we would have

$$\frac{\left(\frac{1}{\log(\mathrm{HR}_1)}\right)^2}{\left(\frac{1}{\log(\mathrm{HR}_*)}\right)^2} = \lim_{\substack{\mathrm{HR}_{1,n}(t)\to 1\\ \mathrm{HR}_{2,n}(t)\to 1}} \frac{n\left(\int_0^1 f_1^{(0)}(t)dt\right)}{n_*\left(\int_0^1 f_*^{(0)}(t)dt\right)} \approx \frac{\text{expected number } \mathscr{E}_1}{\text{expected number } \mathscr{E}_*}$$

4. Simulation

4.1. Simulation

Our next aim is to simulate data to empirically check how close we are to the limiting relationship $n/n_* = \mathrm{ARE}(Z_*, Z)$ when $\Pi_1 = \Pi_*$ for different finite sample sizes. To

conduct the simulations we will assume, as Gómez and Lagakos did, that $T_1^{(j)}$ and $T_2^{(j)}$ follow Weibull distributions. Weibull distributions are chosen for their wide use in the field of survival analysis due to its flexibility, allowing decreasing, constant and increasing hazard rates. The corresponding shape and scale parameters are denoted by β_k and $b_k^{(j)}$ ($j = 0,1$, $k = 1,2$) (shape parameters for both groups are taken equal so that the assumption of the proportionality of the hazard ratios holds). To establish the bivariate distribution of $(T_1^{(0)}, T_2^{(0)})$ we consider Frank's Archimedean survival copula, again as Gómez and Lagakos did. Other choices of copulas would be possible, although main conclusions and recommendations will not differ (Plana-Ripoll and Gómez, 2014). Frank's copula depends on an association parameter θ between $T_1^{(0)}$ and $T_2^{(0)}$ which is biunivocally related to Spearman's rank correlation ρ. Different scenarios will be simulated according to several choices of $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, \mathrm{HR}_1, \mathrm{HR}_2, \rho)$ where $p_1^{(0)}$ and $p_2^{(0)}$ are the probability of observing events $\mathscr{E}_1$ and $\mathscr{E}_2$, respectively, for treatment group 0, HR_1 and HR_2 are relative treatment hazard ratios for $T_j^{(1)}$ versus $T_j^{(0)}$ ($j = 1,2$, respectively) and ρ is Spearman's rank correlation between $T_1^{(0)}$ and $T_2^{(0)}$.

Given a set of values for $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, \mathrm{HR}_1, \mathrm{HR}_2, \rho)$, for a given power Π and a significance level α, the simulation steps are the following:

1. **Computations for the relevant endpoint $\mathscr{E}_1$.** The scale parameters $b_1^{(0)}$ and $b_1^{(1)}$ and the probability $p_1^{(1)}$ of observing the relevant endpoint in group 1 are derived as:

$$b_1^{(0)} = \frac{1}{(-\log(1-p_1^{(0)}))^{1/\beta_1}}$$

$$b_1^{(1)} = \frac{b_1^{(0)}}{\mathrm{HR}_1^{(1/\beta_1)}}$$

$$p_1^{(1)} = 1 - e^{-(1/b_1^{(1)})^{\beta_1}}$$

2. **Computations for the additional endpoint $\mathscr{E}_2$.** The scale parameters $b_2^{(0)}$ and $b_2^{(1)}$ and the probability $p_2^{(1)}$ of observing the additional endpoint in group 1 are derived as:

$$b_2^{(0)} = \begin{cases} \frac{1}{(-\log(1-p_2^{(0)}))^{1/\beta_2}} & \text{for Case 1} \\ * & \text{for Case 3} \end{cases}$$

$$b_2^{(1)} = \frac{b_2^{(0)}}{\mathrm{HR}_2^{(1/\beta_2)}}$$

$$p_2^{(1)} = 1 - e^{-(1/b_2^{(1)})^{\beta_2}}$$

$*$ For Case 3, $b_2^{(0)}$ is found as the solution of equation $p_2^{(1)} = \int_0^1 \int_u^1 f_{(1,2)}^{(0)}(u,v;\rho)dvdu$, where $f_{(1,2)}^{(0)}(\cdot,\cdot;\rho)$ is the joint density between $T_1^{(0)}$ and $T_2^{(0)}$ and ρ is Spearman's ρ coefficient between $T_1^{(0)}$ and $T_2^{(0)}$.

3. **Computation of sample sizes n and n_***

   (a) Compute n (per group) following Freedman (1982) formulas as follows

$$n = \frac{E}{p_1^{(0)} + p_1^{(1)}} \tag{8}$$

   where

$$E = \frac{(HR_1+1)^2(z_{1-\alpha}+z_\Pi)^2}{(HR_1-1)^2} \tag{9}$$

   (b) Compute $\text{ARE}(Z_*, Z)$ based on $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, \text{HR}_1, \text{HR}_2, \rho)$.

   (c) Compute $n_* = n/\text{ARE}(Z_*, Z)$.

   (d) Compute $N = \max\{n, n_*\}$.

4. **Simulation of $T_1^{(0)}, T_1^{(1)}, T_2^{(0)}, T_2^{(1)}, T_*^{(0)}, T_*^{(1)}$**
   Simulate 1000 samples of size N for the 4 endpoints $T_k^{(j)}$ from Weibull $(b_k^{(j)}, \beta_k)$ $(j = 0, 1,\ k = 1, 2)$. Compute $T_*^{(j)} = min\{T_1^{(j)}, T_2^{(j)}\}$.

5. **Computation of empirical powers $\hat{\Pi}_1$ and $\hat{\Pi}_*$**
   For each sample of size n (n_*), compute the logrank statistic Z (Z_*) to compare the treatment effect between $T_1^{(0)}$ and $T_1^{(1)}$ ($T_*^{(0)}$ and $T_*^{(1)}$). For a given significance level α, the rejection region comprises all observed Z (Z_*) such that $Z < z_{1-\alpha}$ ($Z_* < z_{1-\alpha}$) where $z_{1-\alpha}$ is the standard normal quantile corresponding to the left tail probability α. The empirical powers, denoted by $\hat{\Pi}_1$ ($\hat{\Pi}_*$), are calculated as the proportion of samples for which $Z < z_{1-\alpha}$ ($Z_* < z_{1-\alpha}$).
   We note here that whenever $n_* < n$, we only use, for each sample, the first n_* simulated values to compute $\hat{\Pi}_*$, while when $n < n_*$, we only use the first n simulated values to compute $\hat{\Pi}_1$.

6. **Comparison between $\hat{\Pi}_1$ and $\hat{\Pi}_*$**
   For each scenario $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, \text{HR}_1, \text{HR}_2, \rho)$, we compare the differences between the two empirical powers $\hat{\Pi}_1$ and $\hat{\Pi}_*$ obtained from the 1000 simulations.

Table 1: *Values of parameters β_1, β_2, p_1, p_2, HR_1, HR_2 and ρ used for the simulations. There are 624 different configurations, excluding those yielding sample sizes larger than 1100 and* $\mathrm{ARE}(Z_*, Z) > 10$.

Parameters						
$\beta_1 = \beta_2$	0.5	1	2			
(p_1, p_2)	(0.05, 0.01)	(0.05, 0.15)	(0.05,0.35)	(0.1, 0.01)	(0.1, 0.15)	(0.1,0.35)
(p_1, p_2)	(0.15, 0.01)	(0.15, 0.15)	(0.15,0.35)	(0.35, 0.01)	(0.35, 0.15)	(0.35,0.35)
ρ	0.15	0.45	0.75			
$(\mathrm{HR}_1, \mathrm{HR}_2)$	(0.5, 0.3)	(0.5, 0.7)	(0.5, 0.9)	(0.6, 0.3)	(0.6, 0.7)	(0.6, 0.9)
$(\mathrm{HR}_1, \mathrm{HR}_2)$	(0.7, 0.3)	(0.7, 0.7)	(0.7, 0.9)	(0.8, 0.3)	(0.8, 0.7)	
Total number of cases	624					

4.2. Results

We have set $\Pi = 0.9$ and $\alpha = 0.05$ (other values would not provide additional information). We have chosen meaningful values for $(\beta_1, \beta_2, p_1^{(0)}, p_2^{(0)}, \mathrm{HR}_1, \mathrm{HR}_2, \rho)$, based on those arising in cardiovascular clinical trials (Gómez, Gómez-Mateu, Dafni, 2014) (see Table 1). We restrict our simulation study to 624 scenarios corresponding to $\mathrm{ARE}(Z_*, Z) \leq 10$ and sample sizes smaller than 1100 patients per group. These scenarios yield $\mathrm{ARE}(Z_*, Z)$ values between 0.20 and 9.93, sample sizes, n, for the relevant endpoint between 142 and 1081, and, n_*, for the composite endpoint between 53 and 1077 (see Table 2). Similar results were obtained for Case 1, when neither the relevant nor the additional endpoint includes a terminating event, and for Case 3 when the relevant endpoint includes a terminating event, and we only discuss here Case 1.

Table 2: *Computed values of n, n_* and* $\mathrm{ARE}(Z_*, Z)$ *in step 3 of the simulation based on the parameter values given in Table 1.*

	min	median	max
n	142	509	1081
n_*	53	398	1077
$\mathrm{ARE}(Z_*, Z)$	0.2	1.04	9.93

The empirical powers $\hat{\Pi}_1$ in our simulation study resulted in powers between 0.87 and 0.94, with a median of 0.91. A slightly higher median was found for scenarios with low hazard ratios. This finding is acknowledged as well by Freedman (1982).

Table 3 provides the percentiles for the absolute value differences between $\hat{\Pi}_*$ and $\hat{\Pi}_1$. We observe that in 75% of the cases the difference is smaller than 2.3%, and among cases with ARE as large as 3 the difference shrinks to 1.9%. There are, however, few instances, where this difference can be as large as 6%, and they deserve a closer look.

Table 3: Percentiles of $|\hat{\Pi}_* - \hat{\Pi}_1|$ *as a function of ARE values, where* w_i *indicates the corresponding percentile.*

	min	$w_{0.1}$	$w_{0.25}$	$w_{0.5}$	$w_{0.75}$	$w_{0.9}$	max
For all ARE	0	0.002	0.004	0.010	0.023	0.036	0.062
$ARE(Z_*,Z) \leq 3$	0	0.002	0.004	0.008	0.019	0.033	0.062
$ARE(Z_*,Z) > 3$	0.001	0.009	0.016	0.026	0.038	0.046	0.062

Figure 2 plots the differences $\hat{\Pi}_* - \hat{\Pi}_1$ as a function of the $ARE(Z_*,Z)$ values. The behaviour is remarkably different when $ARE(Z_*,Z) \leq 3$ or $ARE(Z_*,Z) > 3$. Whenever $ARE(Z_*,Z) \leq 3$, $\hat{\Pi}_*$ fluctuates around $\hat{\Pi}_1$, within a range of 4%. However, when $ARE(Z_*,Z) > 3$, corresponding mostly to scenarios where treatment has an stronger effect on the additional endpoint than on the relevant endpoint ($HR_2 \leq HR_1 - 0.2$) and the anticipated number of events in the control group is larger for the additional endpoint than for the relevant ($p_2^{(0)} \geq p_1^{(0)}$), the empirical power $\hat{\Pi}_*$ of the logrank test based on the CE never achieves the same power as the logrank test for the relevant endpoint would get. In these cases the interpretation of the $ARE(Z_*,Z)$ as the ratio of the sample sizes, n/n_*, is not as straightforward. Nevertheless, this does not mean that the recommendation of using the CE does not have to be followed since larger values for n_* needed to attain the same power as n does, would reduce the ARE value but not as much as to cross the "1" border that would imply to use the relevant endpoint instead of the CE.

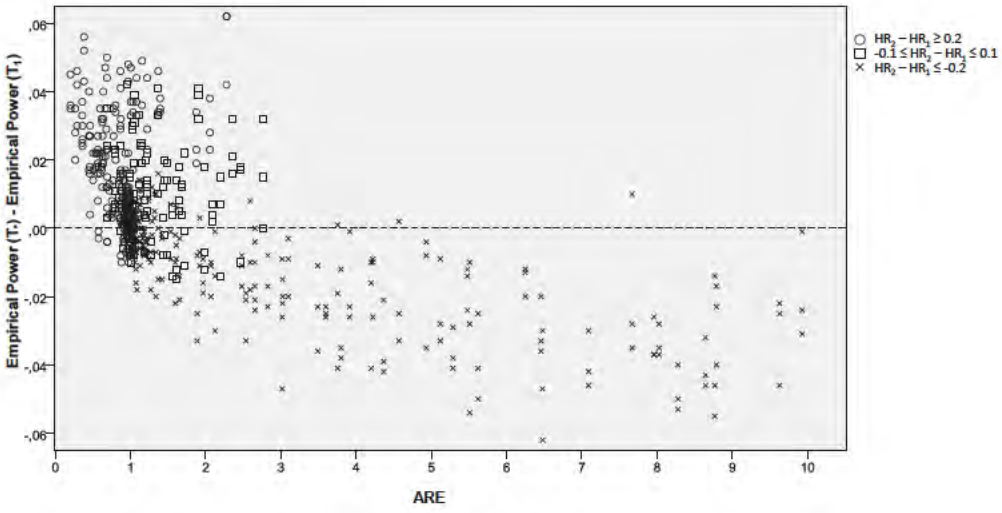

Figure 2: Differences between empirical powers $\hat{\Pi}_* - \hat{\Pi}_1$ *as function of* $ARE(Z_*,Z)$ *and in terms of* $HR_2 - HR_1$.

If we analyze the differences between $\hat{\Pi}_*$ and $\hat{\Pi}_1$ as a function of the differences between the two hazard ratios ($\mathrm{HR}_2 - \mathrm{HR}_1$), we observe that when the two hazard ratios are very close, the two empirical powers are as well very close. Whenever $\mathrm{HR}_2 - \mathrm{HR}_1 \leq -0.2$, not only $\mathrm{ARE}(Z_*, Z)$ values tend to be higher, but also $\hat{\Pi}_* < \hat{\Pi}_1$. (see Figure 2).

Taking into account that absolute differences between powers smaller than 5% could be considered irrelevant, we conclude that the asymptotic relationship $\mathrm{ARE}(Z_*, Z) = n/n_*$ is valid in the majority of scenarios.

All computations in this paper have been implemented in R and are available on request to either author.

5. Discussion

Pitman's relative efficiency is defined as the limiting ratio of sample sizes to give the same asymptotic power under sequences of local alternatives. Given two asymptotically standard normal tests S_n and T_m under the same null and alternative hypotheses, the alternative definition $\mathrm{ARE} = (\mu_S/\mu_T)^2$ where $\sqrt{n}\mu_S$ and $\sqrt{m}\mu_T$ are the respective means under local alternatives, can be used because the equality of the powers holds if $\frac{m}{n} = (\frac{\mu_S}{\mu_T})^2$.

Gómez and Lagakos' method uses the alternative definition of ARE to develop all the computations for the two corresponding logrank tests. Our goal has been to check that the relationship between $(\mu_S/\mu_T)^2$ and the ratio of sample sizes still held when the two hypotheses under test were not the same (H_0 versus H_a and H_0^* versus H_a^*).

It is important to keep in mind that these two hypotheses tests are by no means equivalent, for instance, to check whether treatment has a beneficial effect, we might use $\mathscr{E}_1$ or we might add endpoint $\mathscr{E}_2$ and use $\mathscr{E}_*$. As it is shown in Gómez (2011), even if we assume that the times to $\mathscr{E}_1$ and to $\mathscr{E}_2$ are independent, a beneficial effect on $\mathscr{E}_*$ can occur simultaneously with a beneficial effect on $\mathscr{E}_1$ and a harmful effect on $\mathscr{E}_2$ and not finding a beneficial effect on the composite event $\mathscr{E}_*$ is no guarantee of not having some effect on the individual events $\mathscr{E}_1$ or $\mathscr{E}_2$.

The main result of this paper proves that $\mathrm{ARE}(Z_*, Z)$ coincides with n/n_*, being n and n_* the sample sizes needed to detect specific alternatives HR_1 and HR_2 to attain power Π and for the same significance level α. Therefore, we can use and interpret ARE in its usual way.

The simulation study has been conducted in such a way that for fixed values n and $\mathrm{ARE}(Z_*, Z)$, the sample size n_* is calculated as $n_* = n/\mathrm{ARE}(Z_*, Z)$. Hence an approximate equality of the empirical powers $\hat{\Pi}_1$, of logrank test Z for H_0 versus $H_{a,n}$, and of $\hat{\Pi}_*$ of logrank test Z_* for H_0^* versus $H_{a,n}^*$, indicates that the relationship $\mathrm{ARE}(Z_*, Z) = n/n_*$ holds. Main results from our simulations show that the absolute differences between $\hat{\Pi}_1$ and $\hat{\Pi}_*$ are most of the times less than 2.5%, hence the usual interpretation between (n, n_*) and $ARE(Z_*, Z)$ holds for finite sample sizes.

For those scenarios under which $ARE(Z_*, Z) > 3$, we observe that the empirical power of the test based on $\mathscr{E}_*$ never achieves the empirical power that the logrank test based on $\mathscr{E}_1$ would get. Consequently, larger values of n_* would be needed to attain the same power as n does. In these instances, even though the relationship $\mathrm{ARE}(Z_*, Z) = n/n_*$ is not necessarily true, the recommendation to use the composite endpoint $\mathscr{E}_*$ instead of the relevant endpoint $\mathscr{E}_1$ will still be valid because very rarely a value of $ARE(Z_*, Z) > 3$ would go down to less than 1. However, caution will be needed if one wants to use the relationship $\mathrm{ARE}(Z_*, Z) = n/n_*$ to compute the required sample size n_* if $ARE(Z_*, Z) > 3$. In these cases, a different formulation should be seek.

References

Ferreira-González, I., Permanyer-Miralda, G., Busse, J.W., Bryant, D.M., Montori, V.M., Alonso-Coello, P., Walter, S.D. and Guyatt, G.H. (2007). Methodologic discussions for using and interpreting composite endpoints are limited, but still identify major concerns. *Journal of Clinical Epidemiology*, 60, 651–657.

Freedman, L.S. (1982). Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine*, 1, 121–129.

Freemantle, N., Calvert, M., Wood, J., Eastaugh, J. and Griffin, C. (2003). Composite outcomes in Randomized Trials. Greater precision but with greater uncertainty? *Journal of the American Medical Association*, 289, 2554–2559.

Gómez, G. (2011). Some theoretical thoughts when using a composite endpoint to prove the efficacy of a treatment. *International Workshop on Statistical Modelling. Proceedings of the 26th International Workshop on Statistical Modelling*, València 14–21. http://hdl.handle.net/2117/22571. Last accessed 19 May 2014.

Gómez, G., Gómez-Mateu, M. and Dafni, U. (2014). Informed Choice of Composite End Points in Cardiovascular Trials. *Circulation. Cardiovascular Quality and Outcomes*, 7, 170–178.

Gómez, G. and Lagakos, S.W. (2013). Statistical considerations when using a composite endpoint for comparing treatment groups. *Statistics in Medicine*, 32, 719–738.

Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd Ed. Springer.

Lindner, A.M. and Szimayer, A. (2005). A limit theorem for copulas. *Sonderforschungsbereich 386, Paper 433*. http://epub.ub.uni-muenchen.de/1802. Last accessed 19 May 2014.

Montori, V.M., Permanyer-Miralda, G., Ferreira-González, I., Busse, J.W., Pacheco-Huergo, V., Bryant, D., Alonso, J., Akl, E.A., Domingo-Salvany, A., Mills, E., Wu, P., Schnemann, H.J., Jaeschke, R. and Guyatt, G.H. (2005). Validity of composite end points in clinical trials. *British Medical Journal*, 330, 594–596.

Plana-Ripoll, O. and Gómez, G. (2014). Extension of the ARE method to select the Primary Endpoint in a Randomized Clinical Trial. *Submitted*.

Schoenfeld, D. (1981). The Asymptotic Properties of Nonparametric Tests for Comparing Survival Distributions. *Biometrika*, 68, 316–319.

Wittkop, L., Smith, C., Fox, Z., Sabin, C., Richert, L., Aboulker, J.P., Phillips, A., Chêne, G., Babiker, A. and Thiébaut, R. on behalf of NEAT-WP4. (2010). Methodological issues in the use of composite endpoints in clinical trials: examples from the HIV field. *Clinical Trials*, 7, 19–35.

SORT 38 (1) January-June 2014, 89-102

Testing extreme value copulas to estimate the quantile

Zuhair Bahraoui, Catalina Bolancé and Ana M. Pérez-Marín*

Abstract

We generalize the test proposed by Kojadinovic, Segers and Yan which is used for testing whether the data belongs to the family of extreme value copulas. We prove that the generalized test can be applied whatever the alternative hypothesis. We also study the effect of using different extreme value copulas in the context of risk estimation. To measure the risk we use a quantile. Our results have been motivated by a bivariate sample of losses from a real database of auto insurance claims.

MSC: MSC62-07 MSC62F05.

Keywords: Extreme value copula, extreme value distributions, quantile.

1. Introduction

Let S be the sum of k dependent random variables $X_1,\ldots,X_k$, i.e. $S = X_1 + \cdots + X_k$. The distribution of S depends on the multivariate distribution, i.e. on the relationship between the random variables X_j, $j = 1,\ldots,k$ (see Sarabia and Gómez-Déniz, 2008, for a review about the methods of construction of multivariate distributions). Analyzing the distribution of S is essential in finance and insurance for quantifying the risk of loss. In this regard, there are studies that have analyzed the stochastic behaviour of the sum of dependent risks and the way in which the dependency between these marginal risks may affect the total risk of loss (see, Denuit et al., 1999; Kaas et al., 2000; Cossette et al., 2002; Bolancé et al., 2008b). The aim of this paper is to analyze the test proposed by Kojadinovic et al. (2011) that allows to test whether or not our data have been generated

* bahraouizuhair@gmail.com, bolance@ub.edu and amperez@ub.edu
Department of Econometrics, Riskcenter-IREA, University of Barcelona, Av. Diagonal, 690, 08034 Barcelona, Spain.
Received: December 2013
Accepted: February 2014

by an extreme value copula. We conclude that weak convergence of the test statistic is true for any of the alternative hypothesis. Using a real data base, we have analyzed how the error in the selection of the copula can affect the risk estimate. Throughout this paper we simplify the notation to the bivariate case.

As noted by Fisher (2000), copulas are interesting for statisticians for two basic reasons: firstly, because of their application in the study of nonparametric measures of dependence and, secondly, as a starting point for constructing multivariate distributions that capture dependency structures, even when the marginals follow extreme value distributions (EVD). Also, we know that the choice of the marginals may be crucial to model the dependency behaviour of variables. According to Nelsen (2006), when coupling the marginals in the joint distribution, the copula captures the link between the two behaviours. The relationship between the joint distribution and the marginals is established in the fundamental theorem proposed by Sklar (1959). This theorem shows that a bivariate cumulative distribution function (CDF) H of a random vector of variables (X_1, X_2) with marginal cumulative distribution functions (CDFs) F_1 and F_2 includes a copula C according to the following expression:

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)) \forall x_1, x_2 \in \mathbb{R}. \tag{1}$$

Due to the fact that the joint distribution (and therefore the dependency structure) is unknown, specific tests for choosing the best copula are necessary. This has been the motivation for developing tests for the adequacy of copulas. It is worth mentioning the paper by Genest and Rivest (1993) on inference for bivariate Archimedean copulas, the test proposed in Scaillet (2005) on the positive quadrant dependence hypothesis and, finally, the test of symmetry in bivariate copulas introduced in Quessy et al. (2012).

Regarding the inference for extreme value copulas, we can mention the test proposed in Genest et al. (2011) based on a Cramér-von Mises statistic and the test analyzed in Ghorbal et al. (2009) based on an U-statistic. However, Kojadinovic et al. (2011) uses the $max - stable$ property to test the adequacy of an extreme value copula that is also based on the Cramér-von Mises statistic. In our study we find a similar result for the bivariate case and we obtain the weak convergence of the statistic proposed in the general case.

In Section 2, first, we present our main result and, second, we describe three examples of copulas which are extreme value copulas: Gumbel, Galambos and Hüsler-Reiss. In Section 3 we describe a real database of auto insurance claims which we use in the empirical application. In Section 4 we report the results of our empirical study, firstly we apply the test described in Section 2 and, secondly, we calculate the quantile using different extreme value copulas and compare these results with those obtained when using a widely known non extreme value copula, such as a Gaussian copula. We use two alternative marginal distributions and we compare them: the log-normal, that is a EVD Type I (Gumbel), and the Champernowne distribution, which converges to a Pareto in

the tail and therefore is an EVD Type II (Frechet). We also note that the Champernowe distribution looks more like a log-normal near 0. We conclude in Section 5.

2. Test for extreme value copulas

We know that the class of extreme value copulas corresponds to the class of $max-stable$ copulas (see, for example, Segers, 2012). A copula is $max-stable$ if for every positive real number r and all u_1, u_2 in $[0,1]$, $C(u_1,u_2)=C^r(u_1^{1/r},u_2^{1/r})$. Then we formulate the null hypothesis and its alternative as:

$$\begin{cases} H_0^r : C(u_1,u_2)=C^r(u_1^{1/r},u_2^{1/r}), \ \ \forall u_1,u_2\in[0,1],\forall r>0 \\ H_1^r : C(u_1,u_2)\neq C^r(u_1^{1/r},u_2^{1/r}), \quad \exists u_1,u_2\in[0,1],\exists r>0 \end{cases}.$$

Specifically we need to test the $max-stable$ hypothesis,

$$\begin{cases} H_0 : \bigcap_{r>0} H_0^r \\ H_1 : \bigcup_{r>0} H_1^r, \end{cases}$$

in practice we only can test H_0^r for some values of r. From Kojadinovic et al. (2011), it seems that $r<1$ is not so good, so they consider only values of r greater than 1.

Let (X_{i1},X_{i2}), $\forall i=1,\ldots,n$ be a bivariate sample of n independent and identically distributed observations. We consider the functions:

$$\begin{aligned} \mathbb{D}_n^r(u_1,u_2) &= \sqrt{n}\left(C_n(u_1,u_2)-C_n^r(u_1^{1/r},u_2^{1/r})\right) \\ \mathbb{D}^r(u_1,u_2) &= \sqrt{n}\left(C(u_1,u_2)-C^r(u_1^{1/r},u_2^{1/r})\right), \end{aligned}$$

where $C_n(u_1,u_2)$ is the empirical copula defined as:

$$C_n(u_1,u_2) \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\mathbf{I}(\hat{F}_{1_n}(X_{i1})\leq u_1,\hat{F}_{2_n}(X_{i2})\leq u_2), \quad u_1,u_2\in[0,1]^2, \qquad (2)$$

where $I(\cdot)$ is an indicator function that takes value 1 if the condition in brackets is true and 0 otherwise. $\hat{F}_{1_n}$ and $\hat{F}_{2_n}$ are the empirical marginal cumulative distribution functions. To test the $max-stable$ property we need to analyze if we can use $\mathbb{D}_n^r(u_1,u_2)$ as an estimator of $\mathbb{D}^r(u_1,u_2)$. Then we find the convergence to a Gaussian process of the difference $\mathbb{D}_n^r(u_1,u_2)-\mathbb{D}^r(u_1,u_2)$.

We use the result by Fermanian et al. (2004) for the weak convergence of the empirical copula process C_n to a Gaussian process $\mathbb{G}$ in the space of all bounded real-

valued functions on $[0,1]^2$, i.e. $l^\infty([0,1]^2)$, which is expressed as follows:

$$\sqrt{n}\,(C_n(u_1,u_2)-C(u_1,u_2)) \rightsquigarrow \mathbb{G}(u_1,u_2) \tag{3}$$

$$= \mathbb{B}(u_1,u_2)-\partial_1 C(u_1,u_2)\mathbb{B}(u_1,1)-\partial_2 C(u_1,u_2)\mathbb{B}(1,u_2), \tag{4}$$

where $\partial_j C(u_1,u_2), j=1,2$ are the partial derivatives of the function C respect to u_j and $\rightsquigarrow$ indicates weak convergence and $\mathbb{B}$ is a Brownian bridge on $[0,1]^2$ with covariance functions:

$$E[\mathbb{B}(u_1,u_2)\mathbb{B}(u_1',u_2')] = C(u_1\wedge u_1', u_2\wedge u_2') - C(u_1,u_2)C(u_1',u_2'),$$

where $\wedge$ is the minimum.

Proposition 1 *If the partial derivatives of a copula $C(u_1,u_2)$ are continuous then for any $r>0$ we have:*

$$\mathbb{D}_n^r(u_1,u_2)-\mathbb{D}^r(u_1,u_2) \rightsquigarrow \mathbb{C}^r(u_1,u_2) = \mathbb{G}(u_1,u_2) - rC^{r-1}(u_1^{1/r},u_2^{1/r})\mathbb{G}(u_1^{1/r},u_2^{1/r}), \tag{5}$$

in $l^\infty([0,1]^2)$. The result in (5) is true under H_0^r and H_1^r.

Kojadinovic et al. (2011) proved the weak convergence under H_0^r of $\mathbb{D}_n^r(u_1,u_2)$ towards the same process defined in Proposition 1. We have proved that the weak convergence of the difference $\mathbb{D}_n^r(u_1,u_2)-\mathbb{D}^r(u_1,u_2)$ is true under H_0^r and H_1^r.

Proof 1 In order to prove the result in Proposition 1 we consider the function:

$$\Gamma : C(u_1,u_2) \longrightarrow \Gamma(C(u_1,u_2)) = C^r(u_1^{1/r},u_2^{1/r}), r>0.$$

Γ is a differentiable function as defined by Hadamard (see, Ren, 1995). We use the Delta functional method to analyze the weak convergence of $\Gamma(C(u_1,u_2)) = C^r(u_1^{1/r},u_2^{1/r})$. To find the Hadamard derivative of Γ that is denoted by Γ', we consider the function:

$$\begin{aligned} h(t) &= \Gamma((C+t\Delta)(u_1,u_2)) - \Gamma(C(u_1,u_2)) \\ &= (C+t\Delta)^r(u_1^{1/r},u_2^{1/r}) - C^r(u_1^{1/r},u_2^{1/r}), \end{aligned}$$

where $t\Delta$ is a function representing a difference, namely, t is a real value and Δ is a fixed perturbation. Then we calculate Γ' as the derivative of function h at $t=0$. Namely, $\Gamma'(\Delta)$ if the first derivative of function $\Gamma(C(u_1,u_2)) = C^r(u_1^{1/r},u_2^{1/r})$ with respect to t evaluated at $t=0$.

Using the expression of the Pascal triangle:

$$(a+b)^n = \sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) a^{n-k} b^k,$$

we obtain that:

$$\begin{aligned} h(t) &= \sum_{k=0}^{r} \left(\begin{array}{c} r \\ k \end{array} \right) C^{r-k}(u_1^{1/r}, u_2^{1/r}) t^k \Delta^k(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r}) \\ &= \left(\begin{array}{c} r \\ 0 \end{array} \right) C^r(u_1^{1/r}, u_2^{1/r}) + \left(\begin{array}{c} r \\ 1 \end{array} \right) C^{r-1}(u_1^{1/r}, u_2^{1/r}) t\Delta(u_1^{1/r}, u_2^{1/r}) \\ &+ \sum_{k=2}^{r} \left(\begin{array}{c} r \\ k \end{array} \right) C^{r-k}(u_1^{1/r}, u_2^{1/r}) t^k \Delta^k(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r}). \end{aligned}$$

If we differentiate at $t = 0$, we obtain:

$$\frac{\partial h(t)}{\partial t}|_{t=0} = \Gamma'(\Delta) = rC^{r-1}(u_1^{1/r}, u_2^{1/r})\Delta(u_1^{1/r}, u_2^{1/r}).$$

The result in Proposition 1 is obtained by observing that:

$$\mathbb{D}_n^r(u,v) - \mathbb{D}^r(u,v) = \sqrt{n}\left((C_n(u_1,u_2) - C(u_1,u_2)) - (C_n^r(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r})) \right).$$

Using the convergence of the empirical copula given by Fermanian et al. (see Fermanian et al. (2004)) we obtain:

$$\sqrt{n}\,(C_n(u_1,u_2) - C(u_1,u_2)) \rightsquigarrow \mathbb{G}(u_1,u_2),$$

and, finally, applying the Delta functional method, we obtain:

$$\sqrt{n}\left(C_n^r(u_1^{1/r}, u_2^{1/r}) - C^r(u_1^{1/r}, u_2^{1/r}) \right) \rightsquigarrow \Gamma'(\mathbb{G}(u_1,u_2)).$$

□

Under the hypothesis H_0 we have that $\mathbb{D}^r(u_1,u_2) = 0$ and in this case $\mathbb{D}_n^r(u_1,u_2)$ weakly converges to process (5).

For hypothesis testing given a fixed r, we use a Cramér-von Mises statistic:

$$S_n^r = \int_0^1 \int_0^1 \left(\mathbb{D}_n^r(u_1,u_2)\right)^2 du_1 du_2. \tag{6}$$

As proposed by Kojadinovic et al. (2011) for a range of values of r, $r_1,\ldots,r_p$, the following statistic can be considered:

$$S_n^{r_1,\ldots,r_p} = \sum_{i=1}^{p} S_n^{r_i}. \tag{7}$$

To calculate the critical values we use the method proposed by Van der Vaart (2000), consisting on generating independent copies of S_n^r. The procedure is as follows:

1. If $\partial_j C(u_1,u_2), j=1,2$ are continuous on $[0,1]^2$ then N independent copies of $\mathbb{D}_n^r$, $\mathbb{D}_n^{r,(1)},\ldots,\mathbb{D}_n^{r,(N)}$ can be generated, such that

$$(\mathbb{D}_n^r, \mathbb{D}_n^{r,(1)},\ldots \mathbb{D}_n^{r,(N)}) \rightsquigarrow (\mathbb{D}^r, \mathbb{D}^{r,(1)},\ldots \mathbb{D}^{r,(N)}),$$

   where $\mathbb{D}^{r,(1)},\ldots,\mathbb{D}^{r,(N)}$ are independent copies of $\mathbb{D}^r$.

2. If $\partial_j C(u_1,u_2), j=1,2$ are continuous on $[0,1]^2$ then, $(S_n^{r,(1)}, S_n^{r,(2)},\ldots,S_n^{r,(N)})$ can be calculated by using a numerical approximation of formula (6) (see, Kojadinovic et al., 2011), then:

$$(S_n^r, S_n^{r,(1)}, S_n^{r,(2)},\ldots,S_n^{r,(N)}) \rightsquigarrow (S^r, S^{r,(1)}, S^{r,(2)},\ldots,S^{r,(N)}),$$

   where $(S^{r,(1)}, S^{r,(2)},\ldots,S^{r,(N)})$ are independent copies of S^r.

3. Obtain the p-value as:

$$\frac{1}{N}\sum_{k=1}^{N} \boldsymbol{I}(S_n^{r,(k)} \geq S_n^r).$$

The Van der Vaart method is implemented in the software R with the function `evTestC()` included in the package `copula` (see, Hofert et al., 2013).

2.1. Three examples of extreme value copulas

In the application presented in next section, we compare three examples of extreme value copulas: Gumbel, Galambos and Hüsler-Reiss, which are described in this section.

The functional form of Gumbel copula (see, Gumbel, 1958) is given by:

$$C_\theta(u_1,u_2) = \exp\left(-\left[(-\ln(u_1))^\theta + (-\ln(u_2))^\theta\right]^{1/\theta}\right),$$

where $\theta \in [1, +\infty)$ is the parameter controlling the dependency structure. Note that, the dependence is perfect when $\theta \to \infty$, while independence corresponds to the case when $\theta = 1$. For the Gumbel copula, it is well known that lower tail dependence is $\lambda_L = 0$ and upper tail dependence is $\lambda_U = 2 - 2^{\frac{1}{\theta}}$, i.e. the Gumbel copula has upper tail dependence.

The Galambos copula was proposed by Galambos (1975) and has the following form:

$$C(u_1, u_2) = u_1 u_2 \exp\left(\left[(-\ln(u_1))^{-\theta} + (-\ln(u_2))^{-\theta}\right]^{-1/\theta}\right),$$

where the range of θ is $[0, \infty)$ and the upper tail dependence is $\lambda_U = 2 - 2^{\frac{1}{\theta}}$.

Another example of extreme value copulas is the Hüsler-Reiss copula that was developed by Hüsler and Reiss (1989). Its functional form is given by:

$$C(u_1, u_2) = \exp\left(-\hat{u_1}\Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta \ln\left(\frac{\hat{u_1}}{\hat{u_2}}\right)\right] - \hat{u_2}\Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta \ln\left(\frac{\hat{u_2}}{\hat{u_1}}\right)\right]\right),$$

where the range of θ is $[0, \infty)$ and Φ is cdf of the standard Gaussian, $u_1 = -\ln(\hat{u_1})$ and $u_2 = -\ln(\hat{u_2})$.

3. The data

Our example is motivated by a problem in the context of insurance. We assume that when there is an accident, the total cost to be paid to a policyholder is the sum of two components: (1) the material damage and (2) the bodily injury compensation. The insurance company is interested in evaluating the risk of a given claim exceeding a certain amount. So the right-tail quantiles are important to understand the risk that an accident claim is very costly.

We work with a random sample of 518 observations containing two types of costs: Cost1, representing property damages and compensation of the loss, and Cost2, which corresponds to the expenses related to medical care and hospitalization. In general, the cost of bodily injuries is covered by the National Institute of Health, however the insured has to bear the cost of some medical expenses and rehabilitation, technical assistance, drugs, etc., including compensation for pain, suffering and loss of income.

Bodily injury claims typically take years to be settled. Nevertheless, all the claims in our sample were already settled in 2002, according to the company, (see, Bolancé et al., 2008b). Finally, we should mention that the compensation may include payments to third parties that have been damaged in one way or another.

In Table 1 we summarize the descriptive statistics of the sample for Cost1, Cost2 and the Total Cost. The variables Cost1 and Cost2 are always positive, and there is a big

difference between the corresponding maximum and minimum values. Furthermore, we observe that the variables described in Table 1 have right skewness. In Figure 1 we show the histograms representing the shape of the distributions associated with the variables Cost1 and Cost2.

The K-Plot (related to Kendall Plot, see, Genest and Boies, 2003) is a visual method that allows us to analyze in a descriptive way if our bivariate data have been generated by an extreme value copula. In Figure 2 we show the K-Plot, that compare the order

Table 1: Descriptive statistics.

Cost	Average	Std.Dev.	Skewness	Min	Max	Median
Cost1	182.80	686.80	15.65	13.00	137900.00	677.00
Cost2	283.92	863.17	8.04	1.00	11855.00	88.00
Total Cost	211.20	752.00	15.27	32.00	149800.00	789.00

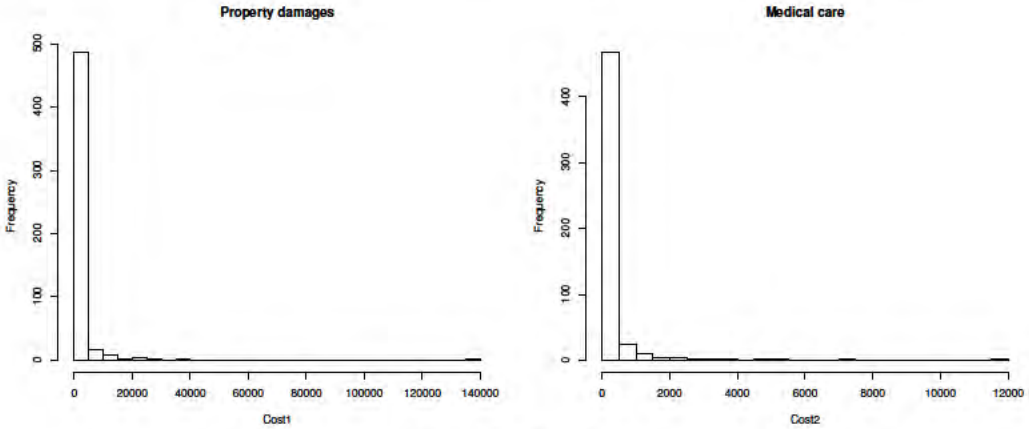

Figure 1: Histograms.

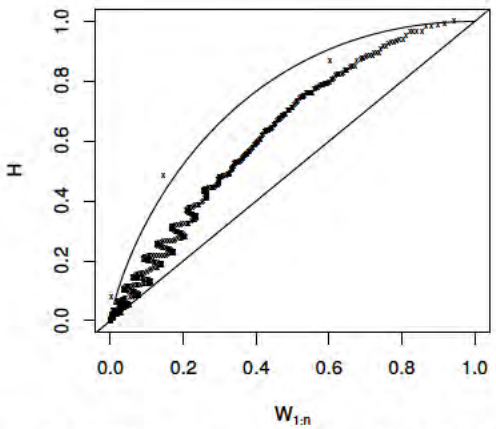

Figure 2: K-Plot associated with to copula of (Cost1, Cost2).

in real data (H, pseudo-observations generated by the bivariate empirical distribution) with the order supposing that the data have been generated by the independence copula (W, expected pseudo-observations). We note that costs have a positive association (as shown in the values of the K-plot above the diagonal, which indicates independence). Almost all points are between the straight line and the boundary curve indicating perfect positive dependence. It seems that for larger values of W, the data are closed to the case of a perfect positive dependence. This means that the higher the severity of the claim, the higher is the correlation between the medical costs and compensation.

4. Results

In this section we report the results that we have obtained in an empirical application of the methodology that we have presented. In order to estimate the total risk of loss, our goal is to determine the dependency structure between the data corresponding to a sample of claims provided by a major insurance company which operates in Spain. To test if our data are generated by an extreme value copula we calculate the value of the Cramér-Von Mises statistic in (7), firstly with $r = 3,4,5$. We have estimated the significance level of the test statistic using the method proposed by Van der Vaart (2000). In total, we generated 1000 independent copies of $S_n^{3,4,5}$. The results are shown in Table 2 and allow us to conclude that the analyzed bivariate data are generated by an extreme value copula.

Table 2: *Cramér-Von Mises statistic.*

Statistic	Estimation	p-value
$S_n^{3,4,5}$	0.2680	0.1773

Table 3: *Copula estimation results.*

	Gaussian	t-Student*	Gumbel	Galambos	Hüsler-Reiss
Parameters	0.5905	0.5981	1.7397	1.0208	1.4946
Standard Errors	0.02485	0.02718	0.07538	0.07689	0.09059
AIC	−212.3695	−217.0000	−246.3839	−243.3305	−237.8542
BIC	−208.1195	−208.5000	−242.1339	−239.0805	−233.6042
CIC	−208.1195	−208.5000	−242.1339	−239.0805	−233.6042
Kendall Tau = 0.4252. *d.f. = 9.6442					

We estimate the parameters of the three extreme value copulas described in Section 2.1: Gumbel, Galambos and Hüsler-Reiss.

In Table 3 we show the estimated parameters for these three copulas together with those obtained for the Gaussian and the t-Student copulas. To estimate the dependence parameter of Gaussian, Gumbel, Galambos and Hüsler-Reiss copulas we have used the inversion of Kendall's tau method (Itau). To estimate the dependence parameter and the degree of freedom of the t-Student copula we have used maximum likelihood estimation (MLE). For selecting the copula we have used two known statistical information criterion, the Akaike Information Criterion $AIC = -2\log L(\theta) + 2k$ and the Bayesian Information Criterion $BIC = -2\ln L(\theta) + k\ln(n)k$, where k is the number of parameters to be estimated and L the value of the likelihood function. Also, we have used the copula information criterion *CIC* propose by Gronneberg and Hjort (2014). The corresponding results are presented in Table 3. We observe that BIC and CIC values are very similar and we conclude that the Gumbel copula is the one that best reflects the dependence structure of our data.

Once the dependency structure is estimated, the next step is to estimate the marginal distribution functions. Considering the histograms in Figure 1, we chosed to use two EVD. Namely, we compare the log-normal distribution, that is a EVD Type I (Gumbel), with the modified Champernowne distribution[1], which converges to a Pareto in the tail and therefore it is an EVD Type II (Frechet); besides the Champernowe distribution looks more like a log-normal near 0. Furthermore, the Champernowne distribution have been analyzed in the context of semiparametric estimation of EVD (see, for example, Bolancé, 2010; Bolancé et al., 2008a; Alemany et al., 2013). In Table 4 we show the results for the maximum likelihood estimation of the marginal distributions. We can see that for Cost1, Log-normal and Champernowne have similar AIC and BIC, however for Cost2 Champernowne provides lower values of AIC and BIC.

Table 4: *Maximum likelihood estimation of marginal distributions.*

	Log-normal	Champernowne
CDFs	$\int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt,\ x \geq 0$	$\frac{(x+c)^\delta - c^\delta}{(x+c)^\delta + (H+c)^\delta - 2c^\delta},\ x \geq 0$
X_1 = Cost1	$\mu = 6.4437, \sigma = 1.3349,$ $AIC = 8448.8950$ and $BIC = 8452.7190$	$\delta = 1.3271, H = 677, c = 0$ $AIC = 8448.163$ and $BIC = 8453.899$
X_2 = Cost2	$\mu = 4.3755, \sigma = 1.5189,$ $AIC = 9425.1340$ and $BIC = 9428.9590$	$\delta = 1.1622, H = 88, c = 0$ $AIC = 6443.7150$ and $BIC = 6449.4510$

1. The cdf of the modified Champernowne distribution is:

$$F(x) = \frac{(x+c)^\delta - c^\delta}{(x+c)^\delta + (H+c)^\delta - 2c^\delta},\ x \geq 0,$$

with parameters $\delta > 0$, $H > 0$ and $c \geq 0$. The estimation of transformation parameters is performed using the maximum likelihood method described in Buch-Larsen et al. (2005).

For evaluating the risk of total loss we estimate the quantile of S at confidence level α ($q_\alpha(S)$). We use the Monte Carlo simulation method and the procedure is as follows:

1. We generate the pseudo-random sample $\left(\hat{U}_{1i}, \hat{U}_{2i}\right), \forall i = 1, \ldots, r$, from the bivariate copulas whose estimated parameters are shown in Table 3.
2. Using the inverse of the marginal CDFs we calculate $\left(\hat{X_{1i}} = F_1^{-1}(\hat{U}_{1i}), \hat{X_{2i}} = \right.$ $\left. = F_2^{-1}(\hat{U}_{2i})\right), \forall i = 1, \ldots, l$, where the sample volume l is large.
3. We calculate $\hat{S}_i = \hat{X}_{1i} + \hat{X}_{2i}, \forall i = 1, \ldots, l$ and we estimate $q_\alpha(S)$ empirically from the generated pseudo-sample. We generate $l = 10,000$ samples.

In Table 5 we show the results of the estimations of q_α for $\alpha = 0.95, 0.99, 0.995,$ 0.999. On the first row of Table 5 we provide the empirical values of the $q_\alpha(S)$ calculated with the 518 observations in the sample of the aggregate loss $S = X_1 + X_2$ for different confidence levels α; below we show the same $q_\alpha(S)$ that have been estimated by the Monte Carlo simulation method for the five copulas considered here. We note the importance of using an extreme value copula and extreme value marginal distributions when the data indicate this behaviour.

Table 5: *Quantiles of total loss.*

α	0.95	0.99	0.995	0.999
Empirical	7905.6000	24821.1400	28420.8700	92112.9300
		Log-normal		
Normal	6635.427	15628.804	20762.765	39733.894
t-Student	6547.524	16638.175	22521.175	39547.101
Gumbel	6432.017	15464.969	22011.382	40001.210
Galambos	6429.160	15471.400	22066.000	39925.670
Hüsler-Reiss	6421.028	15465.126	22122.110	39841.559
		Champernowne		
Normal	7237.591	25504.175	38682.444	110082.261
t-Student	7302.165	25740.933	42223.504	117447.015
Gumbel	7264.831	23944.798	41461.743	119401.409
Galambos	7253.166	24056.946	41409.717	118982.012
Hüsler-Reiss	7241.504	24103.038	41107.537	118539.744

In Table 5 we show that by using log-normal marginal distributions, the estimated quantile is below the empirical quantile for the five copulas considered here. Therefore, the risk is underestimated. We also note that the selected copula does not have much influence on the risk estimation. However, if we use Champernowne marginal distributions, which has a heavier right tail than log-normal distribution, the influence of the selected copula is not significant at lower confidence levels (0.95 and 0.99) but it is sig-

nificant for extreme confidence levels (0.995 and 0.999). As indicated by the goodness of fit measures for our data, the best selection is the Gumbel copula with Champernowne marginal distributions.

5. Conclusions

The test we have introduced for the adequacy of extreme value copulas allows us to determine the suitable copula, especially when the data have extreme values.

In our empirical application, the K-Plot identified a positive and increasing dependence between variables related to automobile insurance claims, and the new test we presented for extreme value copulas confirms that, in our case, we should use an extreme value copula.

In the selection of the marginal distribution we have considered a modified Champernowne distribution. It provides interesting results, due to its similarity to the log-normal distribution for low values of the variable and, additionally, due to its convergence to a Pareto distribution in the right tail.

When the marginal distributions have heavy right tail, as is the case with the Champernowne distribution and if the aim is to estimate extreme quantiles, the results show the importance of testing the adequacy of an extreme value copula to the data.

References

Alemany, R., Bolancé, C. and Guillén, M. (2013). A nonparametric approach to calculating value-at-risk. *Insurance: Mathematics and Economics*, 52, 255–262.

Bolancé, C. (2010). Optimal inverse beta(3,3) transformation in kernel density estimation. *SORT-Statistics and Operations Research Transactions*, 34, 223–237.

Bolancé, C., Guillén, M. and Nielsen, J. (2008a). Inverse beta transformation in kernel density estimation. *Statistics & Probability Letters*, 78, 1757–1764.

Bolancé, C., Guillén, M., Pelican, E. and Vernic, R. (2008b). Skewed bivariate models and nonparametric estimation for cte risk measure. *Insurance:Mathematics and Economics*, 43, 386–393.

Buch-Larsen, T., Guillén, M., Nielsen, J. and Bolancé, C. (2005). Kernel density estimation for heavy-tailed distributions using the Champernowne transformation. *Statistics*, 39, 503–518.

Cossette, H., Gaillardetz, P., Marceau, E. and Rioux, J. (2002). On two dependent individual risk models. *Biometrika*, 30, 153–166.

Denuit, M., Genest, C. and Marceau, E. (1999). Stochastic bounds on sums of dependent risks. *Insurance: Mathematics and Economics*, 25, 85–104.

Fermanian, J., Radulović, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10, 847–860.

Fisher, N. (2000). *Encyclopedy of Statistical Science*. Kotz, S., Johnson, N. L. and Read, C. B. eds., New York.

Galambos, J. (1975). Order statistics of samples from multivariate distributions. *Journal of the American Statistical Association*, 70, 674–680.

Genest, C. and Boies, J. (2003). Detecting dependence with kendall plots. *Journal of the American Statistical Association*, 57, 275–284.

Genest, C., Kojadinovic, I., Neslehová, J. and Yan, J. (2011). A goodness-of-fit test for bivariate extreme-value copulas. *Bernoulli*, 17, 253–275.

Genest, C. and Rivest, L. P. (1993). Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association*, 88, 1034–1043.

Ghorbal, B. Genest, C. and Neslehová, J. (2009). arge sample tests of extreme value dependence for multivariate copulas. *The Canadian Journal of Statistics*, 39, 703–720.

Gronneberg, S. and Hjort, N. (2014). The copula information criteria. *Scandinavian Journal of Statistics*, forthcoming.

Gumbel, E. (1958). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55, 698–707.

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2013). Multivariate dependence with copulas. *R package*.

Hüsler, J. and Reiss, R. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics and Probability Letters*, 7, 283–286.

Kaas, R. Dhaene, J. and Goovaert, M. (2000). Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, 27, 151–168.

Kojadinovic, I., Segers, J. and Yan, J. (2011). Large sample tests of extreme value dependence for multivariate copulas. *The Canadian Journal of Statistics*, 39, 703–720.

Nelsen, R. (2006). *An Introduction to Copulas*, 2nd Ed. Springer, Portland, OR, USA.

Quessy, J., Genest, C. and Neslehova, J. (2012). Test of symmetry for bivariate copulas. *Annals of Institute of Statistical Mathematics*, 64, 811–834.

Ren, J. (1995). Hadamard differentiability on $d[0,1]^p$. *Journal of Multivariate Analysis*, 55, 14–28.

Sarabia, J. and Gómez-Déniz, E. (2008). Construction of multivariate distributions: a review of some recent results. *SORT-Statistics and Operations Research Transactions*, 32, 3–36.

Scaillet, O. (2005). A kolmogorov smirnov type test for positive quadrant dependence. *The Canadian Journal of Statistics*, 33, 415–427.

Segers, J. (2012). Max-stable models for multivariate extremes. *REVSTAT-Statistical Journal*, 10, 61–82.

Sklar, A. (1959). Fonctions de répartition á n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'université de Paris*, 8, 229–231.

Van der Vaart, A. (2000). *Weak Convergence and Empirical Processes*, 2nd Ed. Springer.

Information for authors and subscribers

Author Guidelines

SORT accepts for publication only original articles that have not been submitted simultaneously to any other journal in the areas of statistics, operations research, official statistics or biometrics.

Articles should be preferably applied and may include computational or educational elements. Publication will be exclusively in English. All articles will be forwarded for systematic peer review by independent specialists and/or members of the Editorial Board, except for those articles specifically invited by the journal or reprinted with permission. Reviewer comments will be sent to the first corresponding author if changes are requested in form or content.

To submit an article, the author must upload it in **PDF format**.

The article should be prepared in double-spaced **format**, using a 12-point typeface.

The **title page** must contain the following items: title, name of the author(s), professional affiliation and complete address, and an abstract (75100 words) followed by the keywords and MSC2010 classification of the American Mathematical Society.

Before submitting an article, the author(s) would be well advised to ensure that the text uses **correct English**.

Bibliographic references within the text must follow this format: author surname followed by the year of publication in parentheses [i.e., Mahalanobis (1936), Rao (1982b)]. The complete reference citations should be listed alphabetically at the end of the article, with multiple publications by a single author listed chronologically. Examples of reference formats are as follows:

- Article: Casella, G. and Robert, C. (1998). Post-processing accept-reject samples: recycling and rescaling. Journal of Computational and Graphical Statistics, 7, 139–157.
- Book: Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2003). Bayesian Data Analysis, 2nd Ed. Chapman & Hall / CRC, New York.
- Chapter in book: Engelmann, B. (2006). Measures of a rating's discriminative power-applications and limitations. In: Engelmann, B. and Rauhmeier, R. (eds), The Basel Risk Parameters: Estimation, Validation, and Stress Testing. Springer, New York.
- Online article (put issue or page numbers and last accessed date): Marek, M. and Lesaffre, E. (2011). Hierarchical Generalized Linear Models: The R Package HGLMMM. Journal of Statistical Software, 39 (13). http://www.jstatsoft.org/v39/i13. Last accessed 28 March 2011.

Explanatory footnotes should be numbered sequentially and placed at the bottom of the corresponding page. **Tables and figures** should also be numbered sequentially.

Once the article has been accepted, the journal editorial office will **contact the author** with instructions about this final version, which should be submitted using **LaTeX**. The journal secretary will provide authors with LaTeX templates and appropriate references to the MSC2010 classification of the American Mathematical Society. All the submissions should be handled by RACO (Revistes Catalanes en Accs Obert) website.

New Authors: please register at: http://www.raco.cat/index.php/SORT/user/register. Upon successful registration you will be sent an e-mail with instructions to verify your registration.

Submission Preparation Checklist

As part of the submission process, authors are required to check off their submission's compliance with all of the following items, and submissions may be returned to authors that do not adhere to these guidelines.

1. The submitted manuscript follows the guidelines to authors published by SORT
2. Published articles are under a Creative Commons License BY-NC-ND
3. Font size is 12 point
4. Text is double-spaced
5. Title page includes title, name(s) of author(s), professional affiliation(s), complete address of corresponding author
6. Abstract is 75100 words and contains no notation, no references and no abbreviations
7. Keywords and MSC2010 classification have been provided
8. Bibliographic references are according to SORT's prescribed format
9. English spelling and grammar have been checked
10. Manuscript is submitted in PDF format

Copyright notice and author opinions

The articles are licensed under a Creative Commons Atribution-NonCommercial-NoDerivs 3.0 Spain License.

You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work), you may not use the work for commercial purposes and you may not alter, transform, or build upon the work.

Published articles represent the author's opinions; the journal SORT-Statistics and Operations Research Transactions does not necessarily agree with the opinions expressed in the published articles.

SORT Statistics and Operations Research Transactions
Institut d'Estadstica de Catalunya (Idescat)
Via Laietana, 58 08003 Barcelona. SPAIN
Tel. +34-93.557.30.76 – Fax +34-93.557.30.01
sort@idescat.cat

How to cite articles published in SORT

Commenges, D. (2003). Likelihood for interval-censored observations from multi-state models. *SORT*, 27 (1), 1-12.

Subscription form
SORT *(Statistics and Operations Research Transactions)*

Name _____

_____

Organisation _____

_____

Street Address _____

_____

Zip/Postal code _____ City _____

State/Country _____ Tel. _____

Fax _____ NIF/VAT Registration Number _____

E-mail _____

Date _____

Signature

I wish to subscribe to **SORT *(Statistics and Operations Research Transactions)*** for the year 2014 (volume 38)

Annual subscription rates:

– Spain: €40 (4% VAT included)
– Other countries: €44 (4% VAT included)

Price for individual issues (current and back issues):

– Spain: €15/issue (4% VAT included)
– Other countries: €17/issue (4% VAT included)

Method of payment:

☐ Bank transfer to account number 2100 5000 50 0200051156
IBAN NUM. ES61 2100 5000 5002 0005 1156

☐ Check made payable to the Institut d´Estadística de Catalunya

Please send this subscription form (or a photocopy) to:

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-557 30 01

Bank copy

Authorisation for automatic bank withdrawal in payment for
SORT *(Statistics and Operations Research Transactions)*

The undersigned _____

authorises Bank/Financial institution _____

located at (Street Address) _____

Zip/postal code _____ City _____

Country _____

to draft the subscription to **SORT *(Statistics and Operations Research Transactions)*** from my account

number ☐☐☐☐ ☐☐☐☐ ☐☐ ☐☐☐☐☐☐☐☐☐☐

Date _____

Signature

SORT *(Statistics and Operations Research Transactions)*
Institut d´Estadística de Catalunya (Idescat)
Via Laietana, 58
08003 Barcelona
SPAIN
Fax: +34-93-412 31 45