Chapter 13
Nonparametric and Semiparametric Bivariate Modeling of Petrophysical Porosity-Permeability Dependence from Well Log Data

Arturo Erdely and Martin Diaz-Viera

Abstract Assessment of rock formation permeability is a complicated and challenging problem that plays a key role in oil reservoir modeling, production forecast, and the optimal exploitation management. Generally, permeability evaluation is performed using porosity-permeability relationships obtained by integrated analysis of various petrophysical measurements taken from cores and wireline well logs. Dependence relationships between pairs of petrophysical variables, such as permeability and porosity, are usually nonlinear and complex, and therefore those statistical tools that rely on assumptions of linearity and/or normality and/or existence of moments are commonly not suitable in this case. But even expecting a single copula family to be able to model a complex bivariate dependency seems to be still too restrictive, at least for the petrophysical variables under consideration in this work. Therefore, we explore the use of the Bernstein copula, and we also look for an appropriate partition of the data into subsets for which the dependence strucure was simpler to model, and then a conditional gluing copula technique is applied to build the bivariate joint distribution for the whole data set.

13.1 Introduction

Assessment of rock formation permeability is a complex and challenging problem that plays a key role in oil reservoir modeling, production forecast, and the opti-

Arturo Erdely
Programa de Actuaría, División de Matemáticas e Ingeniería
Facultad de Estudios Superiores Acatlán, Universidad Nacional Autónoma de México
e-mail: arturoerdely@yahoo.com.mx

Martín Díaz-Viera
Programa de Investigación en Recuperación de Hidrocarburos, Instituto Mexicano del Petróleo, México D.F., México
e-mail: mdiazv@imp.mx

P. Jaworski et al. (eds.), *Copula Theory and Its Applications*, Lecture Notes in Statistics 198,
DOI 10.1007/978-3-642-12465-5_13,

mal exploitation management. Generally, permeability evaluation is performed using porosity-permeability relationships obtained by an integrated analysis of various petrophysical measurements taken from cores and wireline well logs. In particular, in carbonate double-porosity formations with an heterogeneous structure of pore space this problem becomes more difficult because the permeability usually does not depend on the total porosity, but on classes of porosity, such as vuggular and fracture porosity (secondary porosity). Even more, in such cases permeability is directly related to the connectivity degree of the pore system structure. This fact makes permeability prediction a challenging task.

Dependence relationships between pairs of petrophysical random variables, such as permeability and porosity, are usually nonlinear and complex, and therefore those statistical tools that rely on assumptions of linearity and/or normality and/or existence of moments are commonly not suitable in this case. The use of copulas for modeling petrophysical dependencies is not new [3] and t-copulas have been used for this purpose. But expecting a single copula family to be able to model any kind of bivariate dependency seems to be still too restrictive, at least for the petrophysical variables under consideration in this work. Therefore, we first adopted a non-parametric approach by the use of the Bernstein copula [15, 16] and estimating the quantile function by Bernstein polynomials [12]. Later, we adopted a semiparametric approach by looking for a partition of the data into subsets for which the dependence structure was simpler to model by means of parametric families of copulas, and then a conditional gluing copula technique is applied to build the bivariate joint distribution for the whole data set.

Among many others, two main tasks are usually of interest in petrophysical modeling: to reproduce the underlying dependence structure by data simulation, and to explain a variable of interest in terms of an allegedly predictive variable (regression).

13.2 Methodology

Let $\mathscr{S} := \{(x_1,y_1),\ldots,(x_n,y_n)\}$ be independent and identically distributed bivariate observations of a random vector (X,Y). In this work, the (x_k,y_k) paired values represent porosity-permeability measurements. We may obtain empirical estimates for the marginal distributions of X and Y by means of

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{I}\{x_k \leq x\}\,, \qquad G_n(y) = \frac{1}{n}\sum_{k=1}^{n}\mathbb{I}\{y_k \leq y\}\,, \tag{13.1}$$

where $\mathbb{I}$ stands for an indicator function which takes a value equal to 1 whenever its argument is true, and 0 otherwise. It is well-known [1] that the empirical distribution F_n is a consistent estimator of F, that is, $F_n(t)$ converges almost surely to $F(t)$ as $n \to \infty$, for all t.

We model vuggy porosities as an absolutely continuous random variable X with unknown marginal distribution function F, and permeability as an absolutely continuous random variable Y with unknown marginal distribution function G. From [11] we have bivariate observations from the random vector (X,Y), see Figs. 13.1 (left) and 13.2, and Fig. 13.3 and Table 13.1 for marginal descriptive statistics. For simulation of continuous random variables, the use of the empirical distribution function estimates (13.1) is not appropriate since F_n is a step function, and therefore discontinuous, so a smoothing technique is needed. Since one of our main goals is simulation of porosity-permeability paired variates, it will be better to have a smooth estimation of the marginal quantile function $Q(u) = F^{-1}(u) = \inf\{x : F(x) \geq u\}$, $0 \leq u \leq 1$, which is possible by means of Bernstein-Kantorovic polynomials as in [12]:

$$\widetilde{Q}_n(u) = \sum_{k=0}^{n} \frac{x_{(k)} + x_{(k+1)}}{2} \binom{n}{k} u^k (1-u)^{n-k}, \tag{13.2}$$

where the $x_{(k)}$ are the order statistics, and the analogous case for marginal G in terms of values $y_{(k)}$.

Similarly, we have the *empirical copula* [2], a function C_n with domain $\{\frac{i}{n} : i = 0,1,\ldots,n\}^2$ defined as

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}\{rank(x_k) \leq i\,,\, rank(y_k) \leq j\} \tag{13.3}$$

and its convergence to the true copula C has also been proved [7, 14]. The empirical copula is not a copula, since it is only defined on a finite grid, not in the whole unit square $[0,1]^2$, but by Sklar's Theorem C_n may be extended to a copula. Moreover, a *smooth* extension is possible by means of the Bernstein copula [15, 16]:

$$\widetilde{C}(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{n} C_n\left(\frac{i}{n}, \frac{j}{n}\right) \binom{n}{i} u^i (1-u)^{n-i} \binom{n}{j} v^j (1-v)^{n-j} \tag{13.4}$$

for every (u,v) in the unit square $[0,1]^2$, and where C_n is as defined in (13.3). In order to simulate replications from the random vector (X,Y) with the dependence structure inferred from the observed data $\mathscr{S} := \{(x_1,y_1),\ldots,(x_n,y_n)\}$, accordingly to [13] we have the following:

Algorithm 1

1. *Generate two independent and continuous Uniform* $(0,1)$ *random variates* u,t.
2. *Set* $v = c_u^{-1}(t)$ *where*

$$c_u(v) = \frac{\partial \widetilde{C}(u,v)}{\partial u}, \tag{13.5}$$

   and $\widetilde{C}$ *is obtained by (13.4).*
3. *The desired pair is* $(x,y) = (\widetilde{Q}_n(u), \widetilde{R}_n(v))$, *where* $\widetilde{Q}_n$ *and* $\widetilde{R}_n$ *are the smoothed estimated quantile functions of* X *and* Y, *respectively, accordingly to (13.2).*

For a value x in the range of the random variable X and $0 < \alpha < 1$ let $y = \varphi_\alpha(x)$ denote a solution to the equation $\mathbb{P}(Y \leq y \,|\, X = x) = \alpha$. Then the graph of $y = \varphi_\alpha(x)$

is the *α-quantile regression curve* of Y conditional on $X = x$. Recalling (13.5), we have that

$$\mathbb{P}(Y \leq y \,|\, X = x) \,=\, c_u(v)\Big|_{u=F(x)\,,\,v=G(y)}, \tag{13.6}$$

and this result leads to the following algorithm [13] to obtain the α-quantile regression curve of Y conditional on $X = x$:

Algorithm 2

1. *Set* $c_u(v) = \alpha$.
2. *Solve for the regression curve* $v = g_\alpha(u)$.
3. *Replace* u *by* $\widetilde{Q}_n^{-1}(x)$ *and* v *by* $\widetilde{R}_n^{-1}(y)$.
4. *Solve for the regression curve* $y = \varphi_\alpha(x)$.

So far (13.2) and (13.4) constitute a completely nonparametric approach for modeling a jointly continuous random vector (X,Y), without imposing restrictive conditions on the dependence structure and/or marginal behavior. Of course, a "no-free-lunch" principle applies, and a price is paid in terms of a "noisy" regression, as it will be seen later, and the fact that the Bernstein copula is unable to model *tail dependence.*

The nature of the data that will be analyzed in the following section, makes it plausible that certain value ranges for the predictor variable (porosity) are related to different dependence structures (copulas). Under this assumption, we recall the gluing copula technique [17] for the particular case of vertical section gluing and bivariate copulas. For example, given two bivariate copulas C_1 and C_2, and a fixed value $0 < \theta < 1$, we may scale C_1 to $[0,\theta] \times [0,1]$ and C_2 to $[\theta,1] \times [0,1]$ and *glue* them into a single copula:

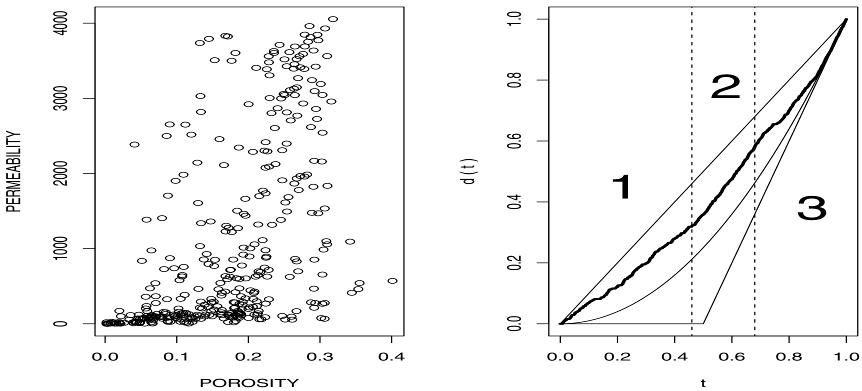
$$C_{1,2,\theta}(u,v) \,=\, \begin{cases} \theta C_1(\frac{u}{\theta}, v)\,, & 0 \leq u \leq \theta, \\ (1-\theta)C_2(\frac{u-\theta}{1-\theta}, v) + \theta v\,, & \theta \leq u \leq 1. \end{cases} \tag{13.7}$$

For a more specific example, in the particular case $C_1(u,v) = \Pi(u,v) = uv$, $C_2(u,v) = W(u,v) = \max\{u+v-1, 0\}$, and $\theta = \frac{1}{2}$, the graph of the diagonal section $\delta_{\Pi,W,1/2}(t) = C_{W,\Pi,1/2}(t,t)$ is shown in Fig. 13.4. This particular kind of copula construction may easily lead to discontinuities in the derivative of the diagonal section at the value $t = \theta$. Therefore, if the empirical diagonal $\delta_n(i/n) = C_n(i/n, i/n)$ exhibits a behavior that suggests a discontinuity of the diagonal derivative at certain points, one may ask if this is possibly due to an abrupt change of the dependence structure at those points, and that being the case, a gluing copula procedure may help to explain the whole dependence structure in terms of two or more simpler dependence models (copulas), in a piecewise manner. It is beyond the scope of this paper to make an exhaustive discussion of different nonparametric techniques to detect possible derivative discontinuity points, but for the data analyzed in this work, using as an heuristic method the Dierckx cubic spline knot detection algorithm [4] has led to a dependence (copula) decomposition that was easier to model by means of known parametric families of copulas, and that is compatible with a petrophysical interpretation. Under this scheme, Algorithms 1 and 2 are used piecewise, with (estimated) parametric copulas instead of $\widetilde{C}$ (the nonparametric Bernstein copula).

13.3 Data Analysis

In order to find knot candidates (or *gluing points*) we applied the Dierckx cubic spline knot detection algorithm [4], which is available as an R contributed package [5], to the empirical diagonal δ_n, obtaining the partition shown in Fig. 13.1 (Right), suggesting to decompose the total sample into three subsamples. Since we are willing to explain permeability (Y) in terms of porosity (X), without loss of information about the joint distribution of the random vector (X,Y) we will consider the total sample $\mathscr{S} := \{(x_1,y_1),\ldots,(x_n,y_n)\}$ to be ordered in the x_k values, that is $x_1 < x_2 < \cdots < x_n$, and so the partition will be induced in the observed x_k values total range $[\min x_k, \max x_k]$.

Fig. 13.1 *Left:* Scatterplot of porosity-permeability data. *Right:* Empirical diagonal (*thick line style*) with suggested gluing points $\theta_1 = 0.46$ and $\theta_2 = 0.68$ (*dashed vertical lines*), Fréchet-Hoeffding diagonal bounds and independence diagonal (*thin line style*).



In Table 13.2 we present a summary of empirical values for the *concordance measures* known as Spearman's rho (ρ_S) and Kendall's tau (τ_K), and an empirical version of the *dependence measure* Φ_H which is just the square root of Hoeffding's *dependence index* [10]. We notice that the concordance and dependence values of the total sample are basically due to subsample 1, since subsamples 2 and 3 clearly exhibit significantly lower values. Also, it should be noticed that the values of ρ_S and Φ_H are quite similar under subsample 1, in contrast with subsamples 2 and 3.

Fig. 13.2 Scatterplot of porosity-permeability data ranks, rescaled to $[0,1]^2$.

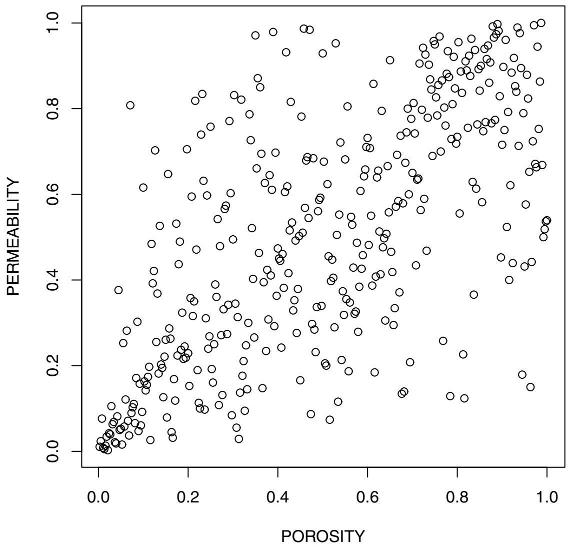


Fig. 13.3 Frequency histograms of porosity and permeability data.

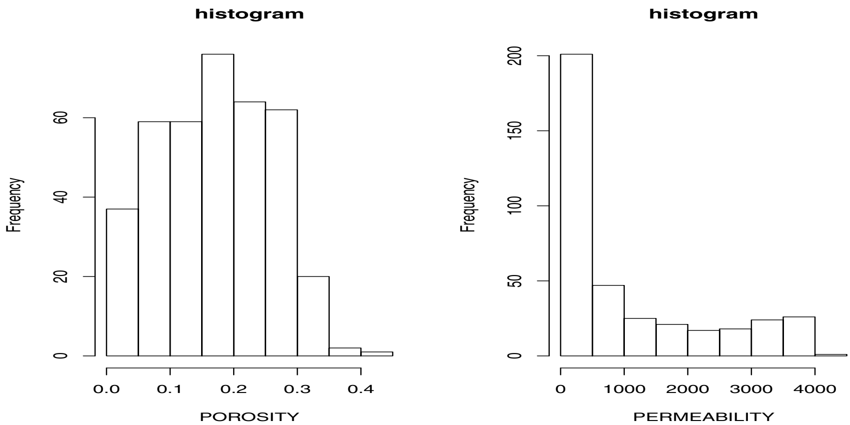


Fig. 13.4 Diagonal section (*thick line style*) of gluing copulas $\Pi(u,v) = uv$ and $W(u,v) = \max\{u+v-1,0\}$ with $\theta = 1/2$, along with Fréchet-Hoeffding diagonal bounds (*thin lines*).

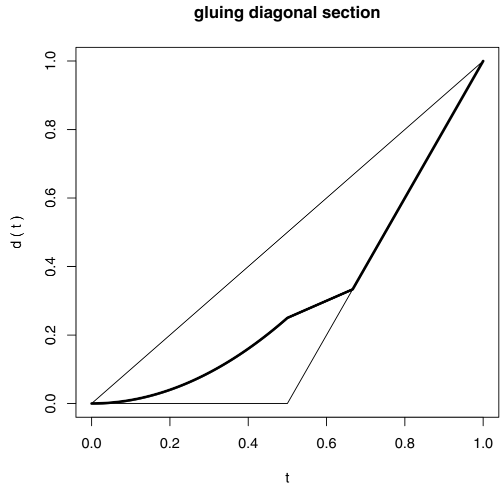


Table 13.1 Summary statistics of porosity and permeability data.

Variable	Min.	1st quartile	Median	Mean	3rd quartile	Max.
Porosity	0.001855	0.098670	0.176700	0.169700	0.236100	0.401300
Permeability	0.05584	0.01006	412.3	1,050	1,826	4,054

This last observation may have as an explanation either the effect of randomness due to smaller subsample sizes or that there could be some kind of dependence that neither Spearman's rho nor Kendall's tau are being able to detect (recall that if a concordance measure is equal to zero this does not imply independence). Therefore, a powerful nonparametric test of independence *à la Deheuvels* based on a Cramér–von Mises statistic [8] has been applied, see Table 13.2 for p-values, clearly rejecting independence for both the total sample and subsample 1, not rejecting independence for subsample 2, but with some doubts about rejecting independence in case of subsample 3, since there is no rule to decide if 0.1274 is a sufficiently low p-value to reject the null hypothesis. Fortunately, a nonparametric symmetry test [6] has been helpful for taking a final decision on subsample 3: by rejecting symmetry we have to reject independence.

In terms of choosing a parametric copula, strong evidence against symmetry is challenging since there is not such a huge catalog of parametric asymmetric copulas as there is indeed for the symmetric case. For the particular data under consideration, it has been possible to transform the data in order to "remove" the asymmetry, by means of Theorem 2.4.4 (1) in [13] and the particular case

$$C_{X,Y}(u,v) = u - C_{X,-Y}(u,1-v)\,, \tag{13.8}$$

Table 13.2 Empirical concordance and dependence values, and p-values for nonparametric tests of independence and symmetry.

Subsample	Size	ρ_S	τ_K	Φ_H	Independence test p-value	Symmetry test p-value
Total	380	$+0.6579$	$+0.4843$	0.6294	0.0000	0.9249
1	174	$+0.6072$	$+0.4379$	0.5819	0.0000	0.8893
2	85	$+0.0538$	$+0.0487$	0.1617	0.4661	0.3194
3	121	$+0.0004$	$+0.0143$	0.1795	0.1274	0.0014

with which, fortunately in this case, it was not possible to reject symmetry for the observations of the random vector $(X,-Y)$ from the transformed subsample 3 (denoted as 3T), see Table 13.3, and Fig. 13.5 for level curves of the empirical copulas for subsamples 3 and 3T. But still we mantain our rejection of independence for the transformed subsample 3T since the independence copula $\Pi(u,v) = uv$ is invariant under transformation (13.8).

Table 13.3 Empirical concordance and dependence values, and p-values for nonparametric tests of independence and symmetry for transformed subsample 3T.

Subsample	Size	ρ_S	τ_K	Φ_H	Independence test p-value	Symmetry test p-value
3T	121	-0.0004	-0.0143	0.1795	0.1274	0.4040
3Ta	75	-0.2668	-0.1870	0.2795	0.0358	0.8323
3Tb	46	$+0.2771$	$+0.1903$	0.3149	0.0803	0.7869

The close to zero values for the concordance measures ρ_S and τ_K under subsample 3 or 3T, but the not so close to zero value for Φ_H along with the independence rejection for subsample 3 and 3T, might have as an explanation that under non-monotone dependence, positive and negative quadrant dependencies "cancel out" under concordance measures (for a theoretical example see Example 5.18 in [13]), but not under a dependence measure such as Φ_H. Therefore, we searched for a partition of subsample 3T to see if it was possible to detect a *gluing point* for positive and negative dependence, again with the aid of the Dierckx cubic spline knot detection algorithm [4], which suggested $\theta_{3T} = 0.615$, see Fig. 13.6, and with such partition into subsamples 3Ta and 3Tb it was possible to decompose subsample 3T into negative and positive dependence, with better p-values to reject independence, and with p-values far away from willing to reject symmetry, see Table 13.3.

So far, it has been possible to decompose the total sample into subsamples 1,2,3Ta, and 3Tb, which do not exhibit strong evidence against symmetry, but with strong evidence against the hypothesis of independence for all cases except subsample 2, and therefore for subsample 2 the final decision is to be modeled by independence. For the remaining subsamples we looked for an appropriate fit among the

Fig. 13.5 Level curves of empirical copulas for subsamples 3 and 3T.

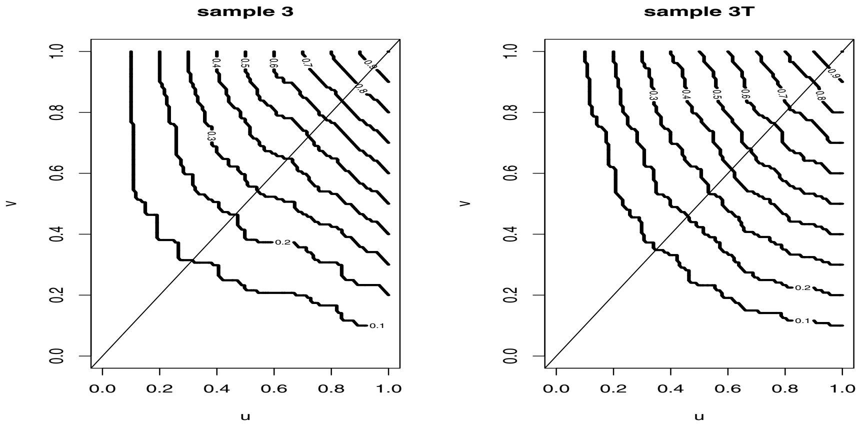
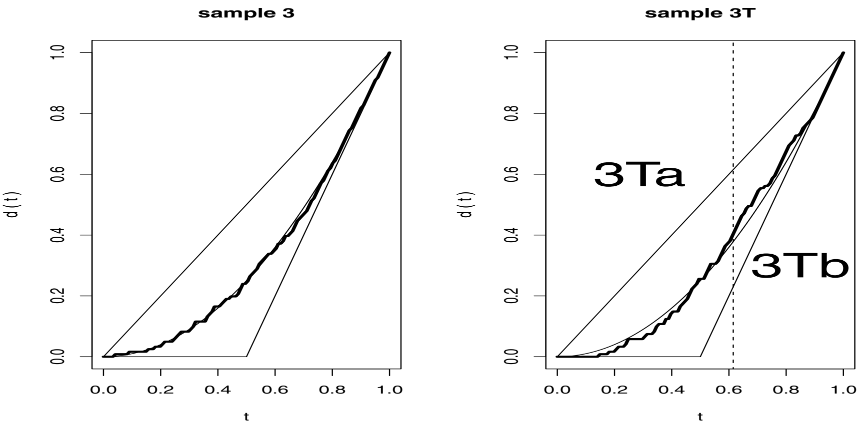


Fig. 13.6 In *thin line style*, Fréchet-Hoeffding diagonal bounds and independence diagonal section. In *thick line style*: empirical diagonals for subsample 3 (*Left*) and subsample 3T (*Right*).



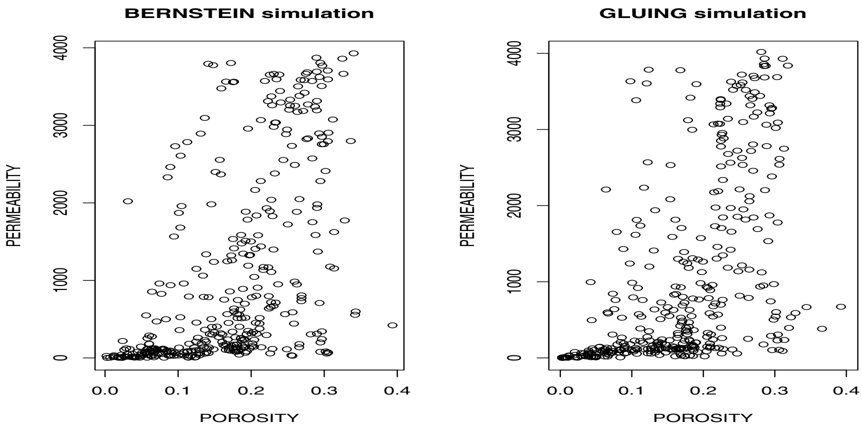
known catalog of symmetric parametric copulas (see for example [13]) by means of a goodness-of-fit test as in [9], with the results summarized in Table 13.4.

Table 13.4 Parameter estimation and goodness-of-fit p-values for the selected copulas.

Subsample	Copula	Parameter	GoF p-value
1	Clayton	$+1.5581$	0.7880
3Ta	Clayton	-0.3151	0.9102
3Tb	A-M-H	$+0.6869$	0.7725

With the above results, it is now possible to perform simulations and regression in a semiparametric fashion (parametric copula and nonparametric Bernstein marginals as described in Sect. 13.2). We also show analog results under a totally nonparametric fashion, as described in Sect. 13.2, see Figs. 13.7 and 13.8.

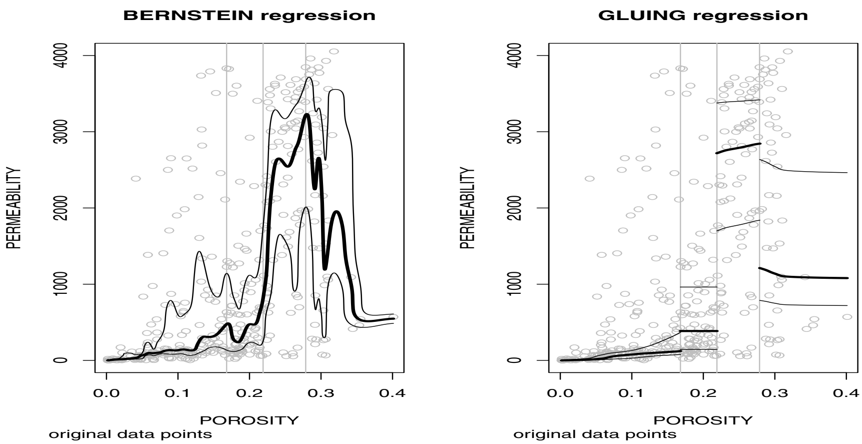
Fig. 13.7 Simulated porosity-permeability paired values, sample size $n = 380$. *Left:* Under Bernstein copula and Bernstein marginals. *Right:* Under a Gluing copula (Clayton + Independence + asymmetric Clayton + asymmetric Ali-Mikhail-Haq) and Bernstein marginals.



13.4 Final Remarks

For the data under consideration (see Figs. 13.1 and 13.2), eventhough there was not strong evidence against the symmetry of the underlying copula [6], it was not possible to find a single known parametric family of copulas that could avoid being

Fig. 13.8 Median, first and third quartile regression curves. *Left:* Bernstein copula and Bernstein marginals. *Right:* Gluing copula (Clayton + Independence + asymmetric Clayton + asymmetric Ali-Mikhail-Haq) and Bernstein marginals, with *vertical lines* indicating *porosity gluing points* 0.168, 0.219, and 0.2786.



rejected by a goodness-of-fit-test (see Table 13.5), and an analogous situation for the marginal distributions. One way of tackling such situation is by a totally nonparametric approach, using Bernstein copula [15, 16] and Bernstein marginals [12], but paying the price of a noisy regression (see Fig. 13.8, Left) and being unable to model *upper and lower tail dependence,* as an immediate consequence of its definition.

Table 13.5 Goodness-of-fit p-values for several families of copulas for the whole sample, calculated with R package "copula" [18].

Copula	GoF p-value
Normal	0.00535
Plackett	0.00535
Frank	0.00055
Clayton	0.00015
Husler-Reiss	0.00015
t-Copula	< 0.00005
Galambos	< 0.00005
Gumbel	< 0.00005

An alternative found was a *gluing semiparametric approach:* decomposing the total sample into subsamples whose dependence structures (copulas) were simpler to model by symmetric and parametric families of copulas, maintaining a nonparametric estimation of the marginals. The result was specially satisfactory in terms of regression, since the discontinuous curve therein obtained essentially matches what it was expected by petrophysical experts [11], see Fig. 13.8 (Right): a moderate and close to linear increase of permeability for porosity values under a *percolation threshold* (porosity around 0.2), a stable level of permeability around the percolation threshold, and an explosive increase in permeability after such threshold. In addition, under this approach it was possible to model a significant lower tail dependence (subsample 1: Clayton copula) of $\lambda_U = 0.641$ (on a $[0,1]$ scale).

References

1. Billingsley, P.: Probability and Measure, 3rd edn. Wiley, New York, NY (1995)
2. Deheuvels, P.: La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance. Acad. Roy. Belg. Bull. Cl. Sci. **65**(5), 274–292 (1979)
3. Díaz-Viera, M., Casar-González, R.: Stochastic simulation of complex dependency patterns of petrophysical properties using t-copulas. Proc. IAMG'05 GIS Spat. Anal. **2**, 749–755 (2005)
4. Dierckx, P.: Curve and Surface Fitting with Splines. Oxford University Press, New York, NY (1995)
5. Dorai-Raj, S.: DierckxSpline: R companion to "Curve and Surface Fitting with Splines". R package version 1.0–9 (2008)
6. Erdely, A., González-Barrios, J.M.: A nonparametric symmetry test for absolutely continuous bivariate copulas. IIMAS UNAM Preimpreso No. **151**, 1–26 (2009)
7. Fermanian, J-D., Radulovíc, D., Wegcamp, M.: Weak convergence of empirical copula processes. Bernoulli **10**, 547–860 (2004)
8. Genest, C., Quessy, J.-F, Rémillard, B.: Local efficiency of a Cramér–von Mises test of independence. J. Multivar. Anal. **97**, 274–294 (2006)
9. Genest, C., Rémillard, B., Beaudoin, D.: Goodness-of-fit tests for copulas: a review and a power study. Insur. Math. Econ. **44**, 199–213 (2009)
10. Hoeffding, W.: Scale-invariant correlation theory. In: Fisher, N.I., Sen, P.K. (eds.) The Collected Works of Wassily Hoeffding, pp. 57–107. Springer, New York, NY (1940)
11. Kazatchenko, E., Markov, M., Mousatov, A., Parra, J.: Carbonate microstructure determination by inversion of acoustic and electrical data: application to a South Florida Aquifer. J. Appl. Geophys. **59**, 1–15 (2006)
12. Muñoz-Pérez, J., Fernández-Palacín, A.: Estimating the quantile function by Bernstein polynomials. Comput. Stat. Data Anal. **5**, 391–397 (1987)
13. Nelsen, R.B.: An Introduction to Copulas, 2nd edn. Springer, New York, NY (2006)
14. Rüschendorf, L.: Asymptotic distributions of multivariate rank order statistics. Ann. Statist. **4**, 912–923 (1976)
15. Sancetta, A., Satchell, S.: The Bernstein copula and its applications to modeling and approximations of multivariate distributions. Econom. Theory **20**, 535–562 (2004)
16. Sancetta, A.: Nonparametric estimation of distributions with given marginals via Bernstein-Kantorovic polynomials: $L1$ and pointwise convergence theory. J. Multivar. Anal. **98**, 1376–1390 (2007)
17. Siburg, K.F., Stoimenov, P.A.: Gluing copulas. Commun. Stat. Theory Methods **37**, 3124–3134 (2008)
18. Yan, J., Kojadinovic, I.: R package 'copula'. Version 0.8–12 (2009)