CHAPTER 1

CONFIDENCE NETS FOR CURVES

Tore Schweder

Department of Economics
University of Oslo, Oslo, NORWAY

E-mail: tore.schweder@econ.uio.no

A confidence distribution for a scalar parameter provides confidence intervals by its quantiles. A confidence net represents a family of nested confidence regions indexed by degree of confidence. Confidence nets are obtained by mapping the deviance function into the unit interval. For high-dimensional parameters, product confidence nets, represented as families of simultaneous confidence bands, are obtained from bootstrapping utilizing the abc-method. The method is applied to Norwegian personal income data.

Key words: Abc-method; bootstrapping; confidence cure; likelihood; simultaneous confidence; quantile regression; personal income.

1. Introduction

Confidence intervals, confidence regions and p-values are the prevalent concepts for reporting inferential results in applications, although Bayesian posterior distributions are increasingly used. In 1930, R.A. Fisher challenged the Bayesian paradigm of the time, and proposed fiducial distributions to replace posterior distributions based on flat priors. When pivots are available, fiducial distributions follow, and are usually termed confidence distributions (Efron 1998, Schweder and Hjort 2002). The cdf of a confidence distribution could also be termed a p-value function. Neyman (1941) showed the connection between his confidence intervals and fiducial distributions. Exact confidence distributions are only available in simple models, but approximate confidence distributions might be found through simulation.

The general concept of confidence distribution is difficult in higher dimension. For vector parameters one must therefore settle for a less ambitious construct to capture the inferential uncertainty. I propose to use the *confidence net*. A confidence net is a stochastic function from parameter space to the unit interval with level sets representing simultaneous confidence regions. One important method to construct a confidence net is to map the deviance function into the probability interval such that the distribution of the transformed deviance at the true value is uniformly distributed. That the confidence net evaluated at the true value is uniformly distributed is actually the defining property of confidence nets.

Confidence net for a scalar parameter was introduced by Birnbaum (1961) under the name 'confidence curve'. The method has been been repeatedly proposed under different names (Bender, Berg and Zeeb 2005) but has only found sporadic use in applied work. I will use the term 'confidence net' to keep it apart from the estimated curve.

Curves such as correlation curves (Bjerve and Doksum 1993) are often represented by its ordinate values at a finite number of argument values, and with a method to connect neighboring values by a piece of continuous curve. I regard the set of values defining the curve as a parameter, usually of high but finite dimension. A simultaneous confidence band for the curve is a product confidence region for the vector parameter. It has the shape of a box or a rectangle in parameter space. A nested family of product confidence regions indexed by their coverage probabilities, constitutes the level sets of a confidence net for the curve. Such confidence nets can be constructed from families of point-wise confidence bands by adjusting the nominal levels to be simultaneous coverage probabilities. Beran (1988) developed this construction into a theory of balanced simultaneous confidence sets. My confidence nets for curves are essentially variants of his method.

Confidence nets for curves might be hard to develop analytically except in special cases. Bootstrapping or other simulation techniques are generally more useful. The abc-method of Efron (1987) leads to confidence nets for scalar parameters which are easily combined to a product confidence net for the curve.

In the next section I discuss confidence nets in general, and give some examples. Then I discuss confidence nets for curves, and show how they might be found by bootstrapping and the abc-method. In the final section I study personal income in Norway by quantile regression curves with associated confidence nets. I am particularly interested in the 5% upper quantile of income on capital as a function of wage for given age. I will use the abc-method on a vector of 29 components.

2. Confidence distributions and confidence nets

The setup is the familiar, with data X being distributed over a measurable space according to a parameterized distribution P_θ. The parameter space is Euclidean of finite dimension.

Assume first θ to be scalar, and that a pivot $piv(\theta, X)$, increasing in θ and with continuous cdf F, is available. The probability transformed pivot $C(\theta; X) = F(piv(\theta, X))$ is then the cdf of a confidence distribution for θ. For any observed value of the data X, $C(\theta; X)$ is in fact a cdf in θ representing the confidence distribution inferred from the data, and for any value of θ, $C(\theta; X)$ is uniformly distributed on the unit interval when $X \sim P_\theta$. These two properties are basic for confidence distributions (Schweder and Hjort (2002)).

Let $C^{-1}(p; X)$ be the confidence quantile. Since C is uniformly distributed at the true value, $P_\theta(C(\theta; X) \le p) = P_\theta(\theta \le C^{-1}(p; X)) = p$. The interval $\left(C^{-1}(\alpha; X); \quad C^{-1}(\beta; X)\right)$ is thus a confidence interval of degree $\beta - \alpha$ for all choices of $0 \le \alpha \le \beta \le 1$. The distribution represented by C therefore distributes confidence over interval statements concerning θ. This is the reason for the name *confidence distribution*.

To ease notation, X will often be suppressed. Whether $C(\theta)$ is a stochastic element or a realization, and whether it is a function of θ or a value, should be clear from the context. Similarly, $L(\theta)$, $D(\theta)$, and $N(\theta)$ denotes likelihood function, deviance function, and confidence net (to be defined below) respectively.

Let

$$N(\theta) = 1 - 2\min\{C(\theta), 1 - C(\theta)\} = |1 - 2C(\theta)|. \tag{1}$$

At each level $1-\alpha$, the level set $\left(C^{-1}(\alpha/2); \quad C^{-1}(1-\alpha/2)\right)$ of N is a tail-symmetric confidence interval. I will call N a tail-symmetric *confidence net* for θ. The concept of confidence net is not confined to scalar parameters. Confidence nets share with pivots the property of having a constant distribution at the true value of the parameter:

Definition 1: A stochastic function N from parameter space to the unit interval is a confidence net if for each θ, $N(\theta; X)$ is uniformly distributed on the unit interval when $X \sim P_\theta$.

The level sets $\{\theta : N(\theta; X) \le \alpha\}$ are clearly confidence regions for θ.

N is called a confidence net since its level sets constitute a net in the mathematical sense.

A confidence distribution is only partially characterized by a related confidence net. This is the case in one dimension, and even more so in higher dimensions. There are actually many confidence nets stemming from the same confidence distribution. In the scalar case for example, there are various tail-asymmetric confidence nets such as $N_s(\theta) = 1 - \min\{C(\theta)/s, (1 - C(\theta))\,/\,(1 - s)\}$ for s between zero and one. Here N_s has level sets $\left(C^{-1}(s\,(1-\alpha)); \quad C^{-1}(1-(1-s)\,(1-\alpha))\right)$ which are tail-asymmetric for $s \neq 1/2$.

Note that $C(\theta) = N_0(\theta)$ and $1 - C(\theta) = N_1(\theta)$ both are confidence nets with extreme tail skewness. They are one-sided confidence nets.

A confidence distribution in one dimension might be displayed by its cdf, its density $c(\theta) = C'(\theta)$, or often preferably by a confidence net which usually would be tail-symmetric. Other distributions such as Bayesian posteriors might also be displayed by (confidence) nets rather than by their densities.

Example 1: Let $X > 0$ be exponentially distributed with mean $\theta > 0$. Then $Y = -X/\theta$ is exponentially distributed on the negative half-axis, and is thus a pivot with cdf $exp(y)$, $y < 0$. The probability transformed pivot is the confidence cdf $C(\theta) = \exp(-\theta X)$. Figure 1 shows representations of the confidence distribution. The five realizations of the confidence net in the lower right panel cross the vertical line at the true value $\theta = 1$ at uniformly distributed levels.

With $\widehat{\theta}$ the maximum likelihood estimator,

$$D(\theta) = -2\ln(L\,(\theta)\,/L(\widehat{\theta})) \tag{2}$$

is the deviance function. The deviance gives rise to confidence nets, also when θ is a vector parameter. So do other suitable objective functions, such as the profile deviance. The following proposition is trivial.

Proposition 1: *If the (profile) deviance evaluated at the true value, $D\,(\theta)$, has continuous cumulative distribution function F_θ, then*

$$N\,(\theta) = F_\theta\,(D\,(\theta))$$

is a confidence net.

In regular cases the null distribution F_θ is asymptotically independent of the parameter, and it is approximately the chi-square distribution with degrees of freedom equal to the dimension of θ.

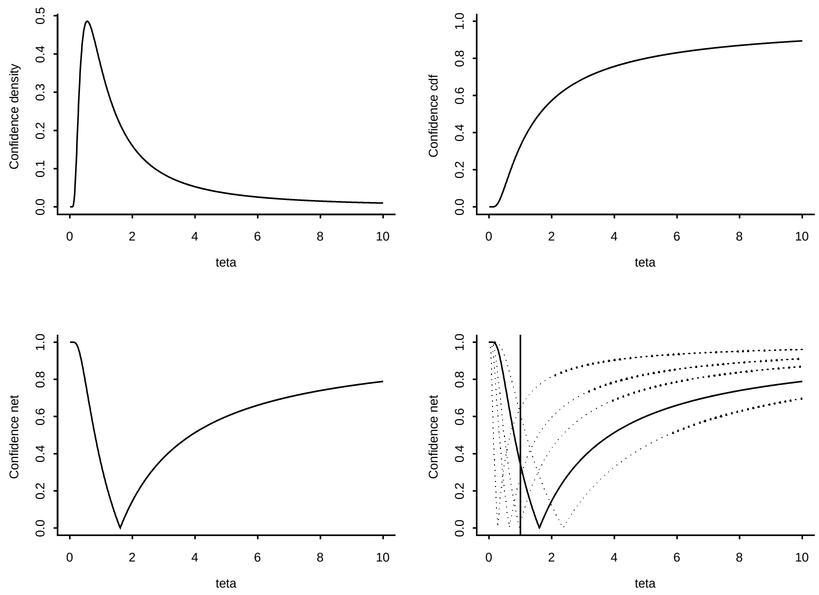

Fig. 1. Confidence density (upper left), confidence cdf (upper right), tail-symmetric confidence net (lover left) for data $X = 1.12$ drawn from the standard exponential distribution, and also four other realizations of N based on data from the same distribution (lower right).

Since the deviance is invariant, the confidence net based on the deviance is invariant. But the deviance might be biased in the sense that the maximum likelihood estimator $\widehat{\theta}$ is biased. I prefer to define bias in terms of the median rather than the mean, to make the notion of no bias invariant to monotonous transformations.

Monotonicity and median might be defined in several ways when the dimension is higher than one. For vector parameters representing curves it is natural to define these notions component-wise, as we shall do.

Definition 2: A confidence net N is unbiased when the point estimator $\widetilde{\theta} = \arg\min N(\theta)$ is median unbiased in each component.

Writing m for the vector of component medians, let $b(\theta) = m\left(\widehat{\theta}\right)$ for the maximum likelihood estimator $\widehat{\theta}$. With b invertible $\widetilde{\theta} = b^{-1}\left(\widehat{\theta}\right)$ is median-unbiased and it minimizes $D\left(b(\theta)\right)$. With F_θ the cdf of $D\left(b(\theta)\right)$, $N\left(\theta\right) = F_\theta\left(D\left(b(\theta)\right)\right)$ is a bias-corrected confidence net.

Example 2: The maximum likelihood estimator $\widehat{\sigma}$ of σ is badly biased in the Neyman-Scott example of highly stratified normal data. Let $X_{ij} \sim N\left(\mu_i, \sigma^2\right)$ $i = 1, \cdots, n$ $j = 1, 2$. The maximum likelihood estimator is $\widehat{\sigma}^2 = \frac{1}{4n}\sum_{i=1}^{n}\left(X_{i1} - X_{i2}\right)^2$ and the profile deviance is $D(\sigma) = 2n\left(\widehat{\sigma}^2/\sigma^2 - \ln\left(\widehat{\sigma}^2/\sigma^2\right) - 1\right)$. $\sigma^2/\widehat{\sigma}^2 \sim 2n/\chi^2_n$ is a pivot yielding a confidence distribution and a confidence net. From the pivotal distribution, $b(\sigma) = m\left(\widehat{\sigma}\right) = \sigma\sqrt{m(\chi^2_n)/\left(2n\right)}$. The null distribution of $D(b(\sigma))$ is that of $2n\left(\chi^2_n/m\left(\chi^2_n\right) - \ln\left(\chi^2_n/m\left(\chi^2_n\right)\right) - 1\right)$, which is free of σ and is easily obtained by simulation. The null distribution is actually very nearly that of $2\chi^2_1$ rather than that of χ^2_1. Figure 2 shows both the confidence net based on the bias-corrected profile deviance and the tail-symmetric confidence net from the confidence distribution based on the pivot, for $n = 20$ and $\widehat{\sigma} = 0.854$. The null distribution of the bias-corrected profile deviance was based on a sample from the χ^2_n of size only 500 to enable the dashed line to be seen. The two nets are practically identical when using the exact distribution of $D(b(\sigma))$.

Example 3: Consider a situation with two variance parameters, σ_1^2 and σ_2^2, with independent estimators distributed proportional to chi-square distributions with 9 and 4 degrees of freedom respectively. We are primarily interested in $\psi = \sigma_2/\sigma_1$. The maximum likelihood estimator $\widehat{\psi}$ has median 0.95ψ while $\widehat{\sigma}_1$ has median $0.96\sigma_1$. Applying median bias-correction to each component separately, the deviance function yields the confidence net that is contoured in Figure 3. The maximum likelihood estimates behind this net are $\widehat{\sigma}_1 = 1$ and $\widehat{\psi} = 2$. The distribution of the two-parametric null deviance is independent of the parameter, and is found by simulation. It is nearly chi-square 2. The confidence net based on the bias-corrected profile deviance for ψ, with null distribution nearly that of $1.08\chi^2_1$, is shown in the lower panel. The exact tail-symmetric confidence net based on the F-distribution is also plotted, but is practically identical to that based on the profile deviance.

My experience is that confidence nets in one dimension based on median bias-corrected profile deviances almost perfectly agree with exact confidence nets based on the same statistic when the latter exist.

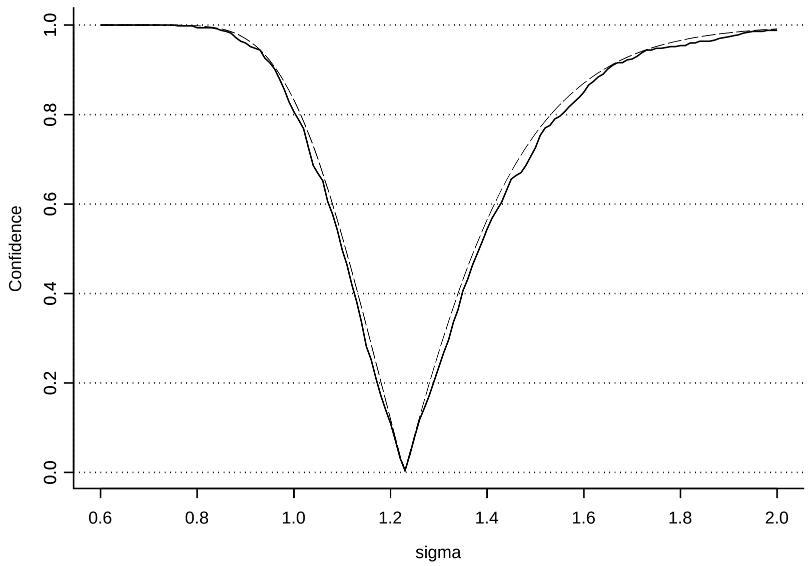

Fig. 2. Exact tail-symmetric confidence net for σ in the Neyman-Scott example (dashed line), and approximate confidence net based on the bias-corrected deviance (solid line, 500 simulation replicates), $n = 20$ and $\widehat{\sigma} = 0.854$.

3. Confidence nets for curves

For non-parametric curves, represented by a vector of ordinates θ, simultaneous confidence bands are product confidence regions for θ. Product confidence nets are therefore desirable for capturing the inferential uncertainty in curve estimates.

For parametric curves, the confidence bands might be obtained from simultaneous confidence regions for the basic parameter. In the linear normal model for example, the method of Scheffé (1959) for multiple comparison is based on elliptical confidence nets. A curve constructed as an indexed set of linear parameters, is obtained from the elliptical confidence net dating back to Working and Hotelling (1929). If the curve represents all the linear functions that can be constructed from an r-dimensional linear parameter, the product confidence net for the curve is equivalent to the r-dimensional elliptical confidence net for the parameter.

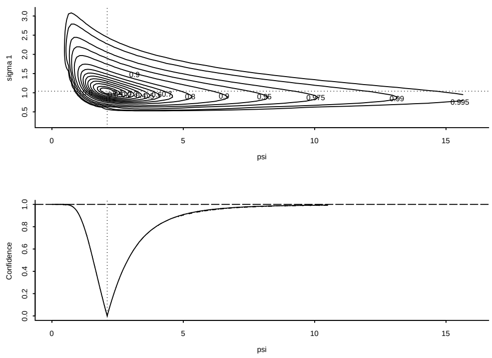

Fig. 3. Confidence net for $(\psi = \sigma_2/\sigma_1, \sigma_1)$ (upper panel) for two normal samples. In lower panel, confidence net for ψ based on the bias-corrected profile deviance. The median unbiased estimates are indicated by dotted lines. Null distributions found by simulation with 100,000 replicates.

Tukey's method of simultaneous inference for pair-wise differences (Scheffé 1959) might be regarded as a problem of developing a product confidence net for the vector of pair-wise differences, and could be solved by simulation as outlined below when the Studentized range distribution is of questionable validity.

Nair (1984) constructed simultaneous confidence bands for survival functions by suitably expanding the set of point-wise intervals. Beran (1988) went a step further by constructing simultaneous product confidence sets from what he called a root. The root is essentially a collection of confidence nets for the components of the vector parameter. For given nominal confidence coefficient for all the point-wise intervals, their product set is a box-shaped confidence set of a degree that depends on the dependence structure and other particulars. This simultaneous confidence degree is often found by simulation. Beran's method is then simply to chose nominal degree to obtain the desired simultaneous degree. It leads to balance in the sense that the simultaneous confidence set has the same marginal coverage probability for each component of the parameter seen in isolation.

My product confidence net for a vector parameter is simply representing the collection of Beran's balanced product sets. Let the curve be represented by the vector parameter θ of dimension T and with generic component θ_t. Let $N_t(\theta_t)$ be the one dimensional confidence net for θ_t. This point-wise confidence net is typically found from a confidence distribution, or from a marginal, conditional or profile deviance function. N_t is uniformly distributed at the true value of the parameter. The confidence net for θ I am looking for is the product net $(K(N_t(\theta_t)) \quad t = 1, \cdots, T)$, written $(K(N_t(\theta_t)))$, for a suitable transformation $K : (0,1) \to (0,1)$ called the *adjustment function.*

Definition 3: An increasing transform K turns a set of point-wise confidence nets $N_t(\theta_t)$ into a balanced product confidence net for $\theta = (\theta_1, \cdots, \theta_T)$,

$$N(\theta) = (K(N_t(\theta_t))),$$

if

$$P_\theta(K(N_t(\theta_t)) \leq \alpha \ \text{ for } t = 1, \cdots, T) = P_\theta\left(\max_{t=1\cdots T} N_t(\theta_t) \leq K^{-1}(\alpha)\right) = \alpha$$

for all θ and $0 < \alpha < 1$.

Analogous to Tukey's problem, which is solved by the Studentized range distribution, my problem is to find the distribution with cdf K of the maximum null net $\max_{t=1\cdots T} N_t(\theta_t)$. Only in rare cases is it possible to solve this problem analytically.

The net $N(\theta)$ is balanced in the sense of Beran (1988) since each of its confidence regions is the product of intervals with the same nominal degree of confidence across the coordinates.

Example 4: Continue the previous example. With $\psi = \sigma_2/\sigma_1$, $\theta = (\psi, \sigma_1)$ is not much of a curve, but is used to illustrate the Chinese box structure of product confidence nets. A product confidence net has rectangular level sets shown in Figure 4. The root for this product net consists of the two confidence nets $N(\sigma_1) = |1 - 2F_{\nu_1}\left(\nu_1\widehat{\sigma}_1^2/\sigma^2\right)|$ and $N(\psi) = |1 - 2F_{\nu_2\nu_1}\left(\left(\widehat{\psi}/\psi\right)^2\right)|$. Here, F_{ν_1} and $F_{\nu_2\nu_1}$ are the cdfs of the appropriate chi-square distribution and the F-distribution respectively. The adjustment function K is now the cdf of $\max\{|1 - 2F_{\nu_1}(X)|, |1 - 2F_{\nu_2\nu_1}(Y/X)|\}$ where X and Y are independent chi square distributed with ν_1 and ν_2 degrees of freedom respectively. Figure 4 shows also the adjustment function K together with the approximate adjustment function determined by the simple Bonferroni method.

Balance is not always desirable for product confidence nets. Some components of the parameter might be of more interest than others, and, for

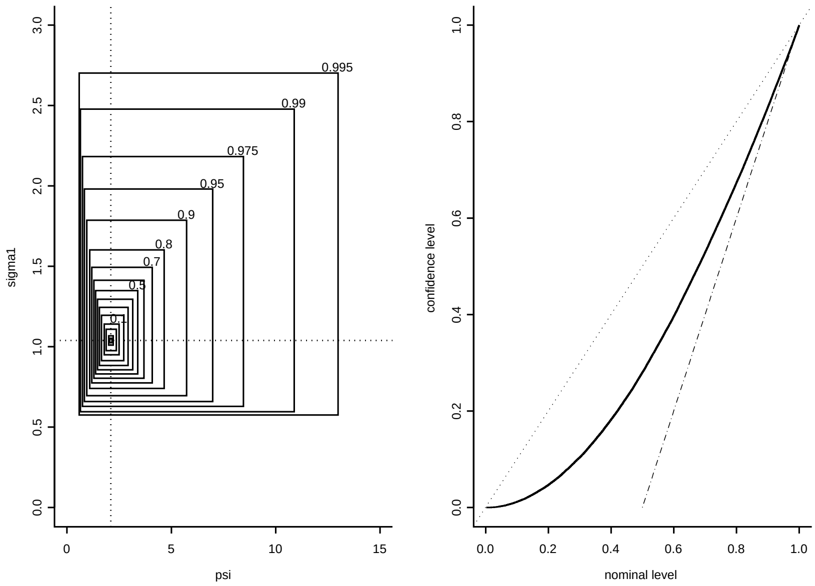

Fig. 4. Level sets of a product confidence net for $\psi = \sigma_2/\sigma_1$ and σ_1 (left panel), and K (solid line, right panel) with the diagonal and also the simple Bonferroni adjustment function (dashed line). $\nu_1 = 9$ and $\nu_2 = 4$. 100,000 simulations.

these, narrower projected nets are required on the expense of wider projected nets for less interesting components. A practical weighting scheme is provided by component-specific transformations of the form

$$K_t\left(c\right) = K\left(c^{1/w_t}\right).$$

Here, w_t is a weight of interest in component θ_t. With K being the cdf of $\max_{t=1\cdots T} N_t\left(\theta_t\right)^{1/w_t}$, $N(\theta) = \left(K_t\left(N_t\left(\theta_t\right)\right)\right)$ is indeed a confidence net for θ.

I return to a balanced product confidence net based on a root of pointwise confidence nets. When the latter is obtained by way of simulation, the adjustment function K can be estimated from the same set of simulated data sets. Let N_t^* be the one dimensional confidence net for θ_t based on a random set of data simulated with $\theta = \theta^0$. When the simulation is done by non-parametric bootstrapping, the estimate $\widehat{\theta}$ based on the observed data serves as reference value, $\theta^0 = \widehat{\theta}$. Maximizing the value of the net at the reference value, across components, leaves us with $\max_{t=1\cdots T} N_t^*\left(\theta_t^0\right)$. The adjustment function K is then simply the cdf of this random variable.

Since the point-wise confidence nets are transformed to a common scale, e.g. to have uniform null distribution, the distribution of the maximum null net will be independent of the true value of θ^0 to a good approximation when the correlation structure in the null nets varies little with the parameter.

Exact confidence nets are available for the individual components only if pivots are available. Unfortunately, models with exact pivots are rare. Under regularity conditions, approximate pivots are however available for large data. Beran (1988) suggests to use bootstrapping to obtain such approximate pivots as input to his construction, and he develops first order asymptotic results which apply to the direct simulation approach and the bootstrapping approach based on Efron's adjusted percentile method for constructing confidence intervals.

Efron (1987) introduced acceleration and bias corrected percentile intervals for one dimensional parameters, see also Schweder and Hjort (2002) who term them abc intervals. The idea is that on some transformed but monotonous scale Γ, $\widehat{\gamma} = \Gamma\left(\widehat{\theta}\right)$ is normally distributed with mean $\gamma - b\left(1 + a\gamma\right)$ and variance $\left(1 + a\gamma\right)^2$, and with $\gamma = \Gamma(\theta)$. With a value for the acceleration constant a, which might take some effort to find, and for the bias constant $b = \Phi^{-1}\left(H(\widehat{\theta})\right)$, the tail-symmetric abc-net is

$$\begin{aligned} N_{abc}(\theta;\widehat{\theta}) &= |1 - 2C_{abc}(\theta)|, \\ C_{abc}(\theta;\widehat{\theta}) &= \Phi\left(\frac{\Phi^{-1}\left(H(\theta)\right) - b}{1 + a\left(\Phi^{-1}\left(H(\theta)\right) - b\right)} - b\right), \end{aligned} \tag{3}$$

where H is the cdf of the bootstrap distribution for the estimator $\widehat{\theta}$ assumed here to be based on $B = \infty$ replicates. The scale transformation is related to H as

$$\Gamma(s) = \left(1 + a\widehat{\gamma}\right)\left\{\Phi^{-1}\left(H(s)\right) - b\right\} + \widehat{\gamma}. \tag{4}$$

The confidence adjustment of the product of these point-wise abc-nets is easily obtained, as explained in the following proposition.

Proposition 2: *The balanced product confidence net for $\theta = (\theta_1, \cdots, \theta_T)$ obtained from point-wise abc-nets $N_{abc}^t(\theta_t;\widehat{\theta}_t)$ given by (3) from non parametric bootstrapping of the data, is*

$$N_{abc}(\theta) = \left(K\left(N_{abc}^t\left(\theta_t\right)\right)\right)$$

where K is the cdf of

$$V = \max_t |1 - 2H_t(\theta_t^*)|, \tag{5}$$

and H_t is the cdf of the bootstrapped component estimates θ_t^*.

Proof: The basis for non parametric bootstrapping is that the bootstrapped curve estimate θ^* has nearly distribution $P_{\widehat{\theta}}$ when $\widehat{\theta}$ has distribution P_θ. The max null-net distribution of the unadjusted product net is thus found from the joint distribution of $C_{abc}(\widehat{\theta}_t; \theta_t^*)$. Consider a given component t. Using (4) for Γ_t and utilizing the invariance property of confidence distributions,

$$C_{abc}(\widehat{\theta}_t; \theta_t^*) = C_{abc}(\widehat{\gamma}_t; \gamma_t^*) = \Phi\left(\frac{\widehat{\gamma}_t - \Gamma_t(\theta_t^*)}{1 + a_t\widehat{\gamma}} - b_t\right) = 1 - H_t\left(\theta_t^*\right).$$

This proves that the adjustment function K indeed is given by (5). □

As mentioned, the same set of bootstrapped curve estimates can be used to calculate the point-wise abc-nets and the adjustment function K. This is correct to first order when the curve is of finite dimension and the assumptions behind the abc method holds for each component.

For non-parametric curves, or parameters of infinite dimension, the transformation K is not directly available. Beran (1988) showed however that sampling from the infinite index set solves the problem, at least asymptotically. For regression curves over compact support, Claeskens and Van Keilegom (2003) found asymptotically correct simultaneous confidence sets based on bootstrapping.

4. Application to Norwegian income data

Statistics Norway surveyed income and wealth for a random sample of individuals in 2002. I consider males 18 years and above, and only look at yearly income on capital Y by yearly income from all other sources X, controlled for age A. Income is measured in Norwegian crowns, and all figures are before tax. Sample size is 22496. I am interested in the upper quantiles in the conditional distribution of Y given X and A, particularly in how the 95% quantile varies with X when controlling for the effect of age. I assume an additive quantile regression function of the form

$$Q_p(Y|X, A) = h(X) + g(A) + error.$$

Q_p is the p-quantile function, here acting on the conditional distribution of Y given X and A. The smooth curve h is represented by the vector

$(h(x_1), \cdots, h(x_{29}))$ for x_t the $t/30$ quantile in the marginal distribution of X. Similarly, g is represented by a 29-dimensional vector.

The model is fitted by a simplified version of the backfitting algorithm of Yu and Lu (2004) as follows. Let $bin(x_t)$ be the bin for X containing the value x_t. Similarly for $bin(a_t)$. The iteration alternates between updating h and g by taking quantiles of adjusted values of Y for values in appropriate bins, and starts with $g = g_0 = 0$. In the i-th iteration, first $h_i(x_t) = Q_p(Y - g_{i-1}(A)|X \in bin(x_t))$ $t = 1, \cdots, 29$. Then $g_i^1(a_t) = Q_p(Y - h_i(X)|A \in bin(a_t))$, which is median shifted to $g_i(a_t) = g_i^1(a_t) - Q_{.5}\left(g_i^1\right)$, also for $t = 1, \cdots, 29$. Since the sample size is large, I used no smoothing across bins in this simplified algorithm.

Using $p = .95$ the backfitting algorithm converges quickly, see Figure 5. I thus settle for 6 iterations in the estimation. The curve estimator to be bootstrapped is $h = s(h_6)$ where s is the smoothing spline with 10 degrees of freedom found by gam in Splus. This degree of smoothing was chosen to allow the rather sharp bend in h come through (Figure 5).

A non-parametric bootstrap experiment with 1000 replicates was carried out. The bootstrap results were first studied for each component separately to see whether there seems to be acceleration in the standard deviation on the transformed scale on which the bootstrap values are normally distributed. If, on the other hand, the assumption $a = 0$ can be made, the level sets of the abc-confidence nets are particularly simple to compute. After some trial and error I found that $\sqrt[10]{\widehat{h}(x_t)}$ is reasonably normally distributed for each component. The question is then whether the standard deviation in bootstrap estimates transformed to this scale is close to constant when plotted against the transformed estimate across the components.

The lower right panel of Figure 6 shows the scatter over the 29 values of x of marginal standard deviation in bootstrapped samples versus curve estimate, both at the 10th root scale. On this scale, standard deviation increases only slightly with curve estimate. From simple regression the slope is estimated to be 0.06. This small value allows the acceleration to be neglected, and $a = 0$ is assumed in (3). Figure 6 also shows that little bias correction is needed (upper left panel), and that the variance on the transformed scale indeed is relatively stable (lower left panel). The normal probability plot of the bootstrapped $\sqrt[10]{\widehat{h}(332000)}$ in the upper right panel shows that this bootstrap distribution is close to normal. This is the case also at the other values of x.

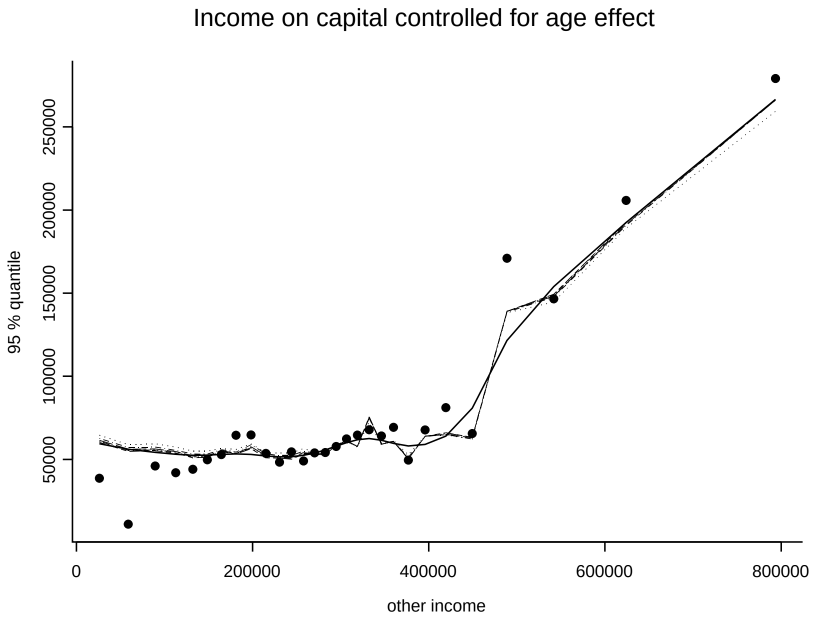

Fig. 5. Points are 95% quantiles of capital income in bins by other income, regardless of age. The iterates of the algorithm of Yu and Lu (2004) are shown by different line types. They converge quickly to the rugged curve appearing to be solid. The smooth solid curve is the smooth of the 6th iteration, $df = 10$.

Table 1. Nominal pointwise confidence needed to obtain the abc net and the simple Bonferroni net at given simultaneous confidence. The default smoothing by gam in Splus is denoted by default df.

	.50	.75	.90	.95	.99
$abc : K^{-1}(\alpha)$ $df = 10$	.908	.964	.988	.994	.998
$abc : K^{-1}(\alpha)$ default df	.816	.926	.976	.990	.998
Bonferroni : $1 - (1 - \alpha)/29$	.983	.991	.997	.998	1.00

The max null net (5) has distribution with cdf K found from the same set of 1000 bootstrap replicates as that used for the point-wise abc-nets.

Bjerve and Doksum (1993) used the Bonferroni method to construct a simultaneous confidence band. This construction leads to a competing confidence net. Figure 7 compares the Bonferroni adjustment with the abc product net adjustment. The latter method is more conservative, as is well known, and by a considerable degree. Table 1 spells this out numerically. It

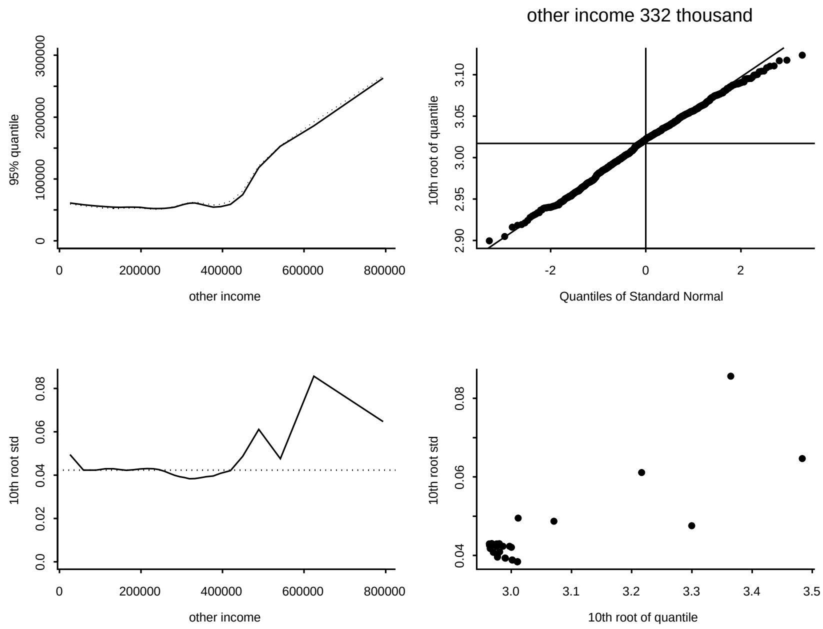

Fig. 6. Curve estimate, bias corrected by the abc method (solid line) and uncorrected estimate (dotted line) in upper left panel. Standard deviation of 10th root of bootstrapped estimates in lower left panel. Upper right panel: normal probability plots of tenth root of curve estimator at a typical level of other income. Lower right panel shows the scatter of marginal standard deviation in the bootstrap samples versus curve estimate, both at the 10th root scale.

also shows adjustment results under more smoothing of the non-parametric quantile regression curve. The smoothed curve in Figure 5 has 10 degrees of freedom. With heavier smoothing there is more autocorrelation in the curve estimator, and the adjustment curve is lifted up.

The resulting confidence net is shown in Figure 8. One conclusion is that the 95% quantile regression of capital income on other income, controlling for the effect of age, is nearly flat up to other income 400000 crowns (slightly less than a professor's wage), and is then nearly linearly increasing. I cannot explain this strong signal in the data, and will pass the question to the economists.

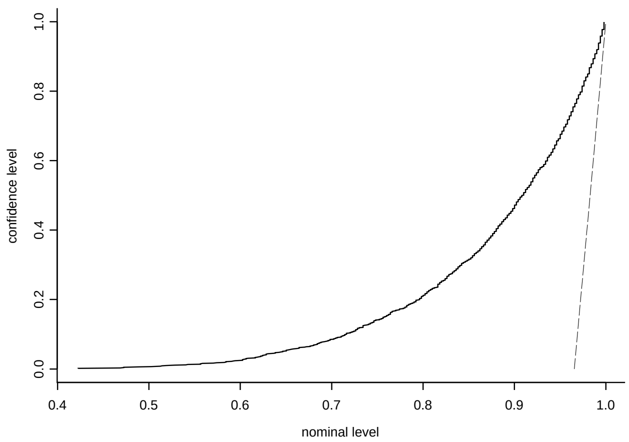

Fig. 7. Adjustment function K for the abc product net (solid curve) and the simple Bonferroni adjustment (dashed line).

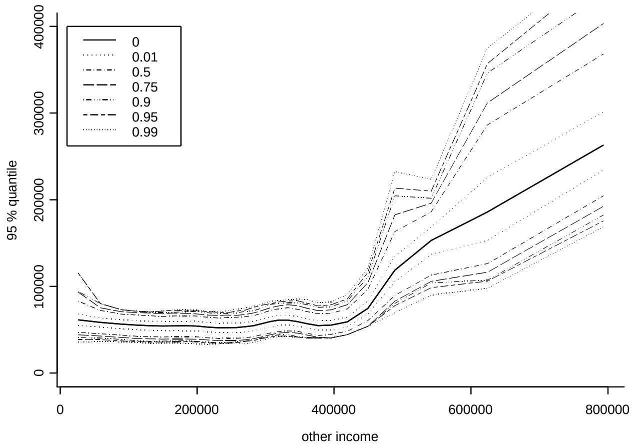

Fig. 8. Level sets for the abc product confidence net for the quantile regression curve of income on capital by other income, controlling for age. Levels given in the legend.

Acknoledgments:

Data from Statistics Norway's Survey of Income and Wealth were made available to me in anonymized form by Norwegian social science data services.

References

1. Beran, R. 1988. Balanced simultaneous confidence sets. *J. Amer. Statis. Assoc.* **83**: 679-686.
2. Bender R., Berg G., Zeeb H. 2005. Tutorial: Using confidence curves in medical research. *Biometrical Journal.* **47**: 237-247
3. Birnbaum, A. 1961. Confidence curves: an omnibus technique for estimation and testing statistical hypotheses. *J. Amer. Statis. Assoc.* **56**: 246-249.
4. Bjerve, S. and Doksum, K. 1993. Correlation curves: measures of association as functions of covariate values. *The Annals of Statistics.* **21**:890-902.
5. Claeskens, G. and Van Keilegom, I. 2003. Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics.* **31**: 1852-1884.
6. Efron, B. 1987. Better bootstrap confidence intervals (with discussion). *J. Amer. Statis. Assoc.* **82**, 171-200.
7. Efron, B. 1998. R.A. Fisher in the 21st century (with discussion). *Statistical Science*, **13**:95-122.
8. Fisher, R.A. 1930. Inverse probability. *Proc. Cambridge Phil. Society.* **26**: 528-535.
9. Nair, V.N. 1984. Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**: 265-275.
10. Neyman, J. 1941. Fiducial argument and the theory of confidence intervals. *Biometrika* **32**: 128-150.
11. Scheffé, H. 1959. *The Analysis of Variance.* Wiley
12. Schweder, T. and Hjort, N.L. 2002. Confidence and likelihood. *Scandinavian Journal of Statistics.* **29**: 309-332.
13. Working, H. and Hotelling, H. 1929. Application of the theory of error to the interpretation of trends. *J. Amer. Stat. Assoc.*, Mar. Suppl.: 73-85.
14. Yu, K. and Lu, Z. 2004. Local linear additive quantile regression. *Scandinavian Journal of Statistics.* **31**: 333-346.