Hotelling tubes, confidence bands and conformal inference

Roger Koenker[1] [ORCID]

Abstract

Stochastic frontier models and methods as pioneered by Peter Schmidt in Aigner et al. (J Econom 6:21–37, 1977), Horrace and Schmidt (J Product Anal 7:257–282, 1996), Amsler et al. (J Econom 190:280–288, 2016) constitute a rare departure from the usual econometric obsession with models for conditional means. They also provided an early stimulus for the development of quantile regression methods. After a brief tutorial on Hotelling tube methods for constructing confidence bands for nonparametric quantile regression, strengthened performance guarantees for such bands are described based on recent developments in conformal inference. These methods may be considered to be a rather idiosyncratic new approach to nonparametric inference for stochastic frontier models.

Keywords Confidence bands · Quantile regression · Conformal inference

1 Introduction

One of my indelible memories of Peter Schmidt was a conversation we had in my kitchen at a party for Midwest Econometrics Group participants in 1993 about the uneasy relationship between statistics and econometrics. "If a statistical tree falls in the forest, but no econometrician sees it," Peter said matter-of-factly, "then it never happened." In 1939 Harold Hotelling, arguably one of the most eminent statisticians and econometricians of the twentieth century witnessed such an event and wrote about in Hotelling (1939). The paper inspired Hermann Weyl to write a highly influential paper, Weyl (1939) generalizing it. Hotelling's idea has attracted a small coterie of admirers in statistics, but it is fair to say that it remains almost unknown in econometrics.

✉ Roger Koenker
rkoenker@illinois.edu

1 University College London, London, UK

My quixotic aim in this paper is to rescue Hotelling's idea from econometric obscurity. I will begin by describing a simple setting in which the idea can be employed to construct a confidence interval for a scalar parameter that enters awkwardly in a standard regression problem. Then, I will describe how it can be used to construct uniform confidence bands for nonparametric regression using penalty methods, and finally I will compare performance with confidence bands constructed with recently developed methods of conformal inference.

2 Hotelling's regression problem

Consider the nonlinear regression model

$$Y_i = x_i^\top \alpha + \lambda_i(\tau)\beta + \varepsilon_i$$

where α, β, τ are unknown parameters, $\lambda_i(\cdot)$ are known functions and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. For the sake of concreteness, we might interpret $\lambda_i(\tau)$ as a Box-Cox transformation of another covariate, say $(z_i^\tau - 1)/\tau$. We would like to test $H_0 : \beta = 0$. Under the null, the Box-Cox parameter τ is not identified, so we need to consider strategies that properly account for this.[1]

By the familiar (Frisch and Waugh 1933) trickery, we can eliminate the α effect.[2] Redefining the notation and assuming for convenience that $\sigma^2 = 1$, we are left with the likelihood ratio statistic

$$L = \inf_\tau \sum (Y_i - \hat{\beta}_\tau \lambda_i(\tau))^2 / \sum Y_i^2$$

Now, denoting the n-vectors, $Y = (Y_i)$, $\lambda = (\lambda_i)$, and the Euclidean norm by $\|\cdot\|$, $\hat{\beta}_\tau = Y^\top \lambda(\tau)/\|\lambda(\tau)\|^2$ so we can rewrite,

$$L = \inf_\tau \|Y\|^{-2}(\|Y\|^2 - 2(Y^\top \lambda)^2/\|\lambda\|^2 + (Y^\top \lambda)^2/\|\lambda\|^2)$$

$$= 1 - \sup_\tau \left(\frac{Y^\top \lambda(\tau)}{\|\lambda(\tau)\|\|Y\|} \right)^2$$

$$\equiv 1 - \sup_\tau (\gamma(\tau)^\top U)^2$$

Now $U = Y/\|Y\|$ is uniformly distributed on the sphere S^{n-1} and $\gamma(\tau) = \lambda(\tau)/\|\lambda(\tau)\|$ is a curve in S^{n-1}. Thus, the test rejects when $W = \sup_\tau \gamma(\tau)^\top U$ exceeds some value $w = \cos\theta$ which is equivalent to

---

[1] There is of course a large literature on such problems, notably: Davies (1977, 1987); Andrews and Ploberger (1994); Hansen (1996), none of whom mention Hotelling. An exception that justifies the qualified "almost unknown" above is Kim et al. (1998). I do not claim that the Hotelling approach is "best" in any sense, only that it is worthy of further consideration. To this end, software to compute the confidence bands described below is available in the R package `quantreg`, Koenker (1999) for a general class of total variation penalized, additive, nonparametric quantile regression models.

[2] Hotelling obviously knew all about how to do this, and one doubts that he learned it from Frisch, but this would probably be hard to establish.

Fig. 1 Angular distance from $\gamma(t)$ to $u = (1, 0)$
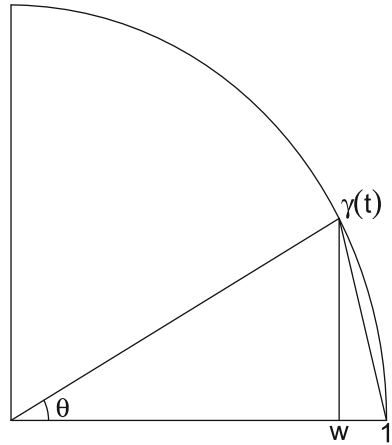
$$U \in \gamma^{\theta} = \{u \in S^{n-1} : \sup_{t} u^{\top}\gamma(t) \geq \cos\theta\}$$
$$= \{u \in S^{n-1} : d(u, \gamma) \leq (2(1-w))^{1/2}\}.$$

Note that the original definition of L is such that we reject for *small* values, so $L < c$, implies we reject for $\sup_{\tau}\gamma(\tau)^{\top}U > w = \cos\theta$ for some critical value of θ. This is illustrated in Fig. 1 of Johansen and Johnstone (1990), reproduced here as Fig. 1. They call this the "angular or geodesic radius θ about γ:"

$$d^2(u, \gamma) = \sin^2(\theta) + (1 - \cos(\theta))^2$$
$$= 1 - 2\cos\theta + \cos^2\theta + \sin^2\theta$$
$$= 2(1 - \cos\theta).$$

So when the distance $d(u, \gamma)$ is small, U falls *inside* tube, and we reject. This may seem a bit counter-intuitive, but is nonetheless correct. There are probably many ways to it sound more intuitive. Here is one possibility. Since it all boils down to a cosine, that is the simple correlation between $\lambda(\tau)$ and Y, we want to reject $H_0 : \beta = 0$ if this correlation/cosine is too large, but Y's that make it too large are the $Y's$ that fall inside the tube.

So how do we compute the critical w or equivalently the critical θ? Since $W > w \equiv \cos\theta$ is equivalent to U being in the tube, we need the volume of the tube. Let $|\gamma|$ denote the length of the arc $\gamma(\tau)$ on the sphere. This can be approximated by the finite difference formula,

$$|\gamma| = \int \|\dot{\gamma}(\tau)\| d\tau \approx \sum_{i=2}^{m} \|(\gamma(\tau_i) - \gamma(\tau_{i-1}))\|,$$

and the τ's are chosen on some relatively fine grid of m points. Note that in the finite difference approximation the $\tau_i - \tau_{i-1}$ that would normally appear in the denominator

Fig. 2 A Hotelling tube on a
2-sphere
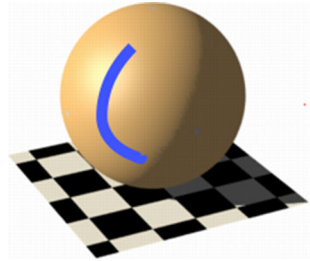
of the difference quotient inside the norm expression cancels with the contribution of
the $d\tau$.

Theorem 1 *If γ is a non-closed regular curve in S^{d-1} then for w near 1,*

$$P(W \geq w) = \frac{|\gamma|}{2\pi}(1-w^2)^{\frac{d-2}{2}} + \frac{1}{2}P(B\left(\frac{1}{2}, \frac{d-1}{2}\right) \geq w^2) \tag{1}$$

*where $B(1/2, (d-1)/2)$ is a beta random variable. If γ is closed, i.e., forms a closed
loop without end points, then the second "cap" term is omitted.*

We ignore pathological complications involving self-intersections of the curve, γ.
This follows from a result of Hotelling (1939), as does the next theorem.

Theorem 2 *Let γ be a regular closed curve in S^{d-1} with length $|\gamma|$. And*

$$\gamma^\theta = \{u \in S^{d-1} \mid \sup_t u^\top \gamma(t) \geq \theta\}$$
$$= \{u \in S^{d-1} \mid d(u, \gamma) \leq (2(1-w))^{1/2}\}$$

*where $w = \cos\theta$. If θ is sufficiently small, then the volume of the tube $V(\gamma^\theta)$ is given
by*

$$V(\gamma^\theta) = |\gamma|\Omega_{d-2}\sin^{d-2}\theta \tag{2}$$

where $\Omega_{d-2} = \pi^{(d-2)/2}\Gamma(d/2)$ is the volume of the unit ball in R^{d-2}.

Heuristically, the formula is,

$$V(\gamma^\theta) = (\text{length of tube}) \cdot (\text{Volume of unit ball}) \cdot \text{radius}^{d-2}$$

Recall that the volume of the unit ball in dimension d is $V = \pi^{d/2}/\Gamma((d+2)/2)$.
When θ is larger, or γ is twisty, then the tube may intersect itself and the formula would
need some refinement. Figure 2 is a crude attempt to depict tube on the 2-sphere, those
with enhanced geometric imagination may try to visualize a three dimensional tube
on the 3-sphere embedded in 4-space.

When the curve is not closed, then it needs "caps" on each end. These caps are given by

$$w_{d-2} \int_{\cos \theta}^{1} (1 - z^2)^{(d-3)/2} dz$$

where $w_{d-2} = 2\pi^{(d-1)/2}/\Gamma((d-1)/2)$ is the $(d-2)$-volume of S^{d-2}. Note that the volume of the sphere, $V(S^{d-1}) = 2\pi^{d/2}/\Gamma(d/2)$, is not the same as the volume of the ball. Note also that $(1 - z^2)^{1/2}$ is again the radius and integrating out the r^{d-3} yields a $d - 2$ dimensional volume. A useful reference for this sort of geometry is Kendall (1961).

How do we get from (2) to (1)? Recall that U is uniform on the $(d - 1)$ sphere so we need to divide by the volume of that sphere to evaluate the probability of being in the tube, so for closed curves,

$$\frac{V(\text{tube})}{V(\text{sphere})} = \frac{|\gamma| \Omega_{d-2} \sin^{d-2} \theta}{2\pi^{d/2}/\Gamma(d/2)}$$
$$= \frac{|\gamma| (\pi^{(d-2)/2}/\Gamma(d/2)) \sin^{d-2} \theta}{2\pi(\pi^{(d-2)/2}/\Gamma(d/2))}$$
$$= \frac{|\gamma|}{2\pi} (1 - w^2)^{(d-2)/2}.$$

To include caps, we also need to divide by the volume of the sphere. Note that

$$\mathbb{P}(B_{1/2, \frac{d-1}{2}} \geq w^2) = \int_{w^2}^{1} \left[x^{1/2-1} (1 - x)^{\frac{d-1}{2}-1} / B(1/2, \frac{d-1}{2}) \right] dx$$
$$= \int_{w^2}^{1} \left[x^{-1/2} (1 - x)^{\frac{d-3}{2}} / B \right] dx.$$

Changing variables $x \to y^2$, we have

$$= \int_{y_0}^{1} \left[y^{-1} (1 - y^2)^{\frac{d-3}{2}} / B \right] 2y \, dy$$
$$= 2 \int_{y_0}^{1} B^{-1} (1 - y^2)^{\frac{d-3}{2}} dy.$$

It remains to show that $B^{-1} = w_{d-2}/V(\text{sphere})$, which follows after a little simplification and recalling that $\Gamma(1/2) = \sqrt{\pi}$.

To check how the Hotelling tube procedure performs in moderate sample sizes, Table 1 reports results of a small simulation experiment. Data are generated with iid x_i standard log-normal and

$$y_i = \beta_n (x_i^\tau - 1)/\tau + \epsilon_i, \quad u \sim \mathcal{N}(0, 1).$$

Table 1 Rejection frequencies for the Hotelling likelihood ratio test for a simple Box-Cox example

| | $\beta_0 = 0$ | | | $\beta_0 = 1$ | | | $\beta_0 = 2$ | | |
	$\tau = -0.5$	$\tau = 0$	$\tau = 0.5$	$\tau = -0.5$	$\tau = 0$	$\tau = 0.5$	$\tau = -0.5$	$\tau = 0$	$\tau = 0.5$
n = 20	0.056	0.058	0.049	0.313	0.193	0.182	0.781	0.459	0.380
n = 50	0.049	0.051	0.057	0.275	0.225	0.342	0.639	0.577	0.782
n = 100	0.063	0.048	0.056	0.350	0.261	0.281	0.840	0.637	0.704
n = 500	0.048	0.052	0.055	0.298	0.243	0.288	0.747	0.612	0.735
n = 1000	0.063	0.046	0.047	0.299	0.218	0.250	0.724	0.549	0.667

Tests are nominal level $\alpha = 0.05$. Local alternatives are employed of the form: $\beta_n = \beta_0/\sqrt{n}$

Three values of τ are considered $\tau \in \{-0.5, 0, 0.5\}$. Local alternatives, $\beta_n = \beta_0/\sqrt{n}$, are considered with $\beta_0 \in \{0, 1, 2\}$. The nominal level of the Hotelling test is taken to be 0.05. and 1000 replications of the experiment are made for each parametric setting. When $\beta = 0$ so the null is true, the test delivers quite accurate size for all of the sample sizes considered, and power is respectable when β deviates from zero.

3 Uniform confidence bands for nonparametric regression

Consider the series expansion model

$$Y_i = \sum_{j=1}^{d} \beta_j a_j(t_i) + \varepsilon_i$$

with $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ as before and $t \in I \subset \mathbb{R}$. Our objective is to find a positive c such that

$$P_{\beta,\sigma,\Sigma}(|\beta^\top a(t) - \hat{\beta}^\top a(t)| \leq c\sigma (a(t)^\top \Sigma a(t))^{1/2} \ \forall \ t \in I) \approx 1 - \alpha,$$

uniformly in β, σ. Johansen and Johnstone (1990) write this as, $P_{\beta,\sigma,\Sigma}(T < c\sigma)$ where

$$T = \sup_{a \in C} \frac{a^\top(\hat{\beta} - \beta)}{\sqrt{a^\top \Sigma a}}$$

Now consider $X \sim \mathcal{N}(\xi, \Sigma)$, so X plays the role of $\hat{\beta}$ and ξ of β. We'd like to make a confidence statement about $\{a^\top \xi | a \in C\}$ and C is some sort of "curve." So now we write,

$$T = T(X, \xi) = \sup_{a \in C} \frac{a^\top(X - \xi)}{\sqrt{a^\top \Sigma a}}.$$

We want the distribution of T, so we can obtain the confidence set

$$R_x = \{\{a^\top \xi\}_{a \in C} | T(X, \xi) < c_{1-\varepsilon}\}$$

where $P_{\xi, \Sigma}(T < c_{1-\varepsilon}) = 1 - \varepsilon$. Write $T = RW$ where,

$$R^2 = (X - \xi)^\top \Sigma^{-1}(X - \xi) \sim \chi_d^2,$$

and

$$W = \sup_{a \in C} \frac{a^\top (X - \xi)}{\sqrt{a^\top \Sigma a}\sqrt{(X - \xi)\Sigma^{-1}(X - \xi)}} = \sup_{a \in C} \frac{(\Sigma^{1/2}a)^\top \Sigma^{-1/2}(X - \xi)}{|\Sigma^{1/2}a||\Sigma^{-1/2}(X - \xi)|}.$$

Now to put things back into the earlier framework of γ and U we set,

$$\gamma(a) = \frac{\Sigma^{1/2}a}{|\Sigma^{1/2}a|}$$
$$U = \Sigma^{-1/2}(X - \xi)/|\Sigma^{-1/2}(X - \xi)|.$$

So as before, $\gamma = \gamma(C) \subset S^{d-1}$, and U is uniform on S^{d-1}. R and W do not depend on ξ, Σ or they do, but only via γ. R^2 is independent of W and $R^2 \sim \chi_d^2$ so,

$$\mathbb{P}(T > c) = \int_c^\infty \mathbb{P}(W > c/r)\mathbb{P}(R \in dr).$$

The random variable W has the same form as in the simple example so,

$$\mathbb{P}(W > w) = \frac{|\gamma|}{2\pi}(1 - w^2)^{(d-2)/2} + \frac{1}{2}\mathcal{P}(B \geq w^2) \equiv b_\gamma(w).$$

(Naiman 1986) bounds this probability by,

$$\mathbb{P}(T > c) \leq \int_c^\infty \min\{b_\gamma(c/r), 1\}\mathcal{P}(R \in dr),$$

and Knowles (1987) suggests ignoring the $b_\gamma < 1$ constraint and then integrates the bound to obtain,

$$\mathbb{P}(T > c) \leq \frac{|\gamma|}{2\pi}e^{-c^2/2} + 1 - \Phi(c).$$

This integration may appear somewhat miraculous, but does actually work out provided that one carefully observes the $\mathcal{P}(R \in dr)$ term. Since $R^2 \sim \chi_d^2$, letting F denote the distribution function of χ_d^2, we have,

$$\mathbb{P}(R \leq r) = \mathcal{P}(R^2 \leq r^2) = F(r^2)$$

so the corresponding density of R is

$$f_R(r) = 2r F'(r^2) = 2r f_{R^2}(r^2).$$

Once one has this bound then various other things fall into place. For example,

$$\mathcal{P}(|T| > c) \leq 2\mathbb{P}(T > c).$$

(Johansen and Johnstone 1990) give further details on the accuracy of the bounds and applications.

4 Additive models for total variation penalized nonparametric quantile regression

In Koenker (2011), I have described a general approach to estimation and inference for additive nonparametric quantile regression models of the form,

$$Q_{Y_i|x_i,z_i}(\tau|x_i, z_i) = x_i^\top \theta_0 + \sum_{j=1}^{J} g_j(z_{ij}).$$

The components $g = (g_1, \cdots, g_J)$ can be univariate or bivariate. Their smoothness can be controlled by penalizing total variation of the functions themselves or their gradients. Estimation is carried out by solving the linear program,

$$\min_{(\theta_0, g)} \sum \rho_\tau(y_i - x_i^\top \theta_0 - \sum g_j(z_{ij})) + \lambda_0 \|\theta_0\|_1 + \sum_{j=1}^{J} \lambda_j \bigvee(\nabla g_j) \qquad (2)$$

where $\rho_\tau(u) = u(\tau - \mathbb{1}(u < 0))$ is the usual quantile objective function, $\|\theta_0\|_1 = \sum_{k=1}^{K} |\theta_{0k}|$ and $\bigvee(\nabla g_j)$ denotes the total variation of the derivative or gradient of the function g. Recall that for g with absolutely continuous derivative g' we can express the total variation of $g' : \mathcal{R} \to \mathcal{R}$ as

$$\bigvee(g'(z)) = \int |g''(z)| dz$$

while for $g : \mathbb{R}^2 \to \mathbb{R}$ with absolutely continuous gradient,

$$\bigvee(\nabla g) = \int \|\nabla^2 g(z)\| dz$$

where $\nabla^2 g(z)$ denotes the Hessian of g and $\| \cdot \|$ denotes the Hilbert–Schmidt norm for matrices. In contrast, total variation penalization of the component functions themselves yields piecewise constant solutions.

Adapting the Hotelling tube idea to construct uniform confidence bands for these components is also described in Koenker (2011), as is selection of the smoothing parameters λ_j, $j = 0, 1, \cdots J$. It should be stressed that all of this machinery relies on the validity of Gaussian approximations for the fitted parameters and estimated functions and is conditional on selected tuning parameters. This is in accord with a large strand of earlier literature including (Wahba 1983; Nychka 1983), and Krivobokova et al. (2010); however, there are inevitable questions that can be raised about both aspects. To explore this, we consider some recent proposals for strengthening coverage guarantees based on conformal inference in the next section.

5 Conformal quantile regression

Conformal prediction, and conformal inference more generally, has grown out of work by Vladimir Volk and colleagues, see, e.g., Shafer and Vovk (2008) for an overview. It has emerged as an essential tool in uncertainty quantification throughout statistics and machine learning. An essential feature of the conformal inference approach in regression is a sample splitting device that allows one to adjust a confidence band constructed with training data based on its performance on a validation sample. Strong finite sample performance guarantees can be proven based on seemingly rather weak exchangeability assumptions. In regression settings, early work presumed conventional iid error structure when constructing the initial bands from the training data; however, (Romano et al. 2019) noted that in more heterogeneous settings narrower bands could be constructed using quantile regression methods. This approach has been further developed in Lei and Candès (2022). In high-dimensional regression, this typically would involve some form of random forest or neural network model for the initial bands, but the same methods can be used in simpler models like the additive models described above.

Construction of conformal prediction bands for additive quantile regression models can be described briefly as follows:

Algorithm 1 Split Conformal Quantile Regression (CQR)

1: **procedure** CQR(x, y, τ)
2: Split the sample into training, T, and validation, V, samples.
3: Compute initial lower, $\hat{q}_0(x) = Q_{Y|X}(\tau_0|x)$ and upper, $\hat{q}_1(x) = Q_{Y|X}(\tau_1|x)$ confidence band limits on training sample.
4: Predict response quantiles y_{i0} and y_{i1} at τ_0 and τ_1 for each of the validation sample observations.
5: Compute conformal scores, $E_i = \max\{\hat{y}_{i0} - y_i, y_i - \hat{y}_{i1}\}$.
6: Compute Q, the $\tau_1 - \tau_0$ quantile of the conformal scores.
7: Return the augmented prediction band $C(x) = [\hat{q}_0(x) - Q, \hat{q}_1(x) + Q]$.

Note that the conformal adjustment of the initial band can make it wider or narrower. When $Q < 0$, it indicates that the validation sample fell well inside the initial band indicating that it is safe to shrink the width of the initial band.

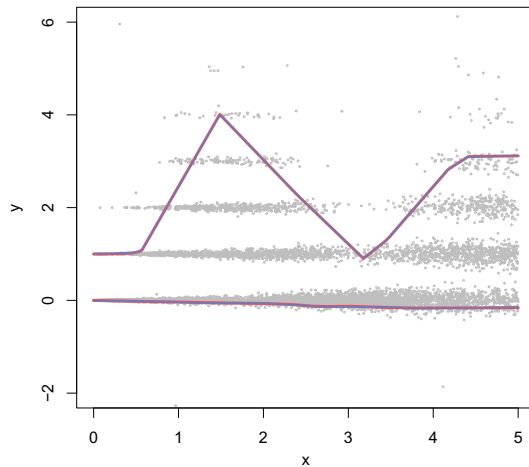There are several potential difficulties with the foregoing recipe.

Fig. 3 Example 1 of Romano, Patterson and Candès: As described in the text the response is concentrated in bands determined by a Poisson component with some quite extreme outliers that are (mostly) invisible in this plot. The Poisson rate is periodic accounting for the obvious heteroscedasticity. The red curves depict the predicted 0.05 and 0.95 conditional quantile estimates based on the training data, using penalization of the derivative of the fitted function, while the blue curves depict the conformally modified estimates. In this example, the conformity scores E_i are quite small and the conformal modification is almost negligible

- Predictions based on the training sample typically are not equipped to extrapolate beyond the empirical support of the training data, so if the validation data, or new data requiring a conformal interval, lie outside that support some accommodation must be made.
- Performance guarantees are based on marginal coverage of the band, so it may happen that in certain regions of design space there may be failures of coverage that are compensated by satisfactory coverage elsewhere. As shown by Foygel Barber et al. (2020), conditional coverage is not achievable in any generality.
- All of the familiar challenges of penalty methods for regression smoothing persist, so choice of smoothing parameters, in particular, can cause headaches, even though poor λ selection can in principle be ameliorated by the conformal adjustment.

We conclude this section by illustrating the use of the conformal method in an artificial data example taken from Romano et al. (2019). Simulated data are generated as,

$$X \sim U[0, 5]$$
$$Y|X \sim \text{Pois}(\sin^2(X) + 0.10) + V$$
$$V \sim 0.03 X Z_0 + 25\mathbb{1}(U < 0.01)Z_1$$
$$U \sim U[0, 1]$$
$$(Z_0 \perp\!\!\!\perp Z_1) \sim \mathcal{N}(0, 1)$$

There are 7000 observations plotted in grey. The Poisson contribution to the response produces a banded structure to the scatterplot with pronounced heteroscedasticity.

Fig. 4 Example 1 of Romano, Patterson and Candès: In contrast to the earlier piecewise linear fit obtained by total variation penalization of the first derivative of g, in this figure total variation of the fitted function itself is penalized resulting in a piecewise constant fit. Clearly this penalty is better suited to the example and mimics quite well the fit depicted in that paper. Again, the conformal adjustment is only barely visible
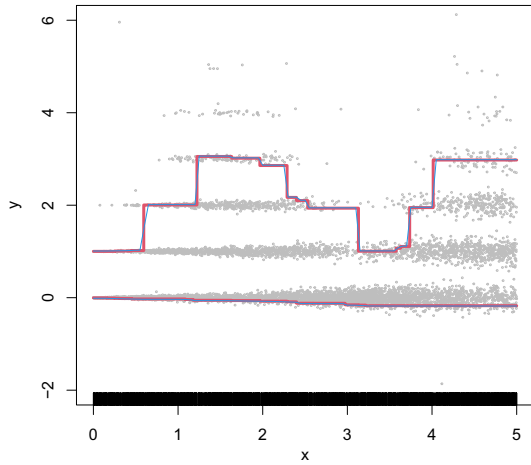
Fig. 5 Pointwise and uniform confidence bands for RPC Example: In contrast to the conformal prediction band, pointwise and uniform bands for the 0.05 and 0.95 conditional quantile functions are considerably wider. The uniform band is based on the Hotelling tube construction described in Koenker (2011) and is depicted as the light grey shaded band enclosing the darker grey pointwise band
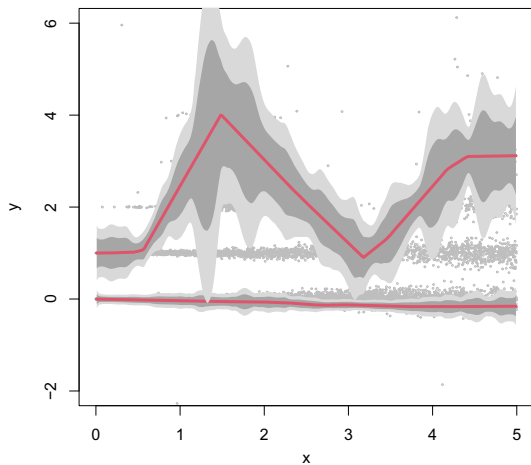
There are a small number of extreme outliers many of which lie outside the frame of the figure; such outliers are harmless since we are estimating conditional quantile functions. Penalizing total variation of g' yields a piecewise linear fit that does not fit the scatter as well as the piecewise constant estimate obtained by penalizing the total variation of g itself. It is striking here that the conformal adjustment in both figures is almost imperceptible. Thus, if interest focuses on prediction intervals for the response, the initial estimates provided by the penalized quantile regression estimates are fine, even though they are based on only half the original sample.

Prediction bands for Y are fine as far as they go, but what if we wanted confidence bands for the conditional quantile functions? Some might argue, e.g., Geiser (1993); Clarke and Clarke (2018), that it is pointless to predict quantities that can never be observed, but I subscribe to the principle: every decent estimate deserves a standard error. Figure 5 illustrates confidence bands for the lower, $\tau = 0.05$ and upper, $\tau = 0.95$ conditional quantile functions as estimated using penalization of g'. The dark grey

bands are the pointwise bands, while the lighter grey bands are those based on the Hotelling tube approach. Note that the bands for the 0.05 estimate are extremely narrow since the data are very concentrated in this region, so the $\tau = 0.05$ conditional quantile is very precisely estimated.

6 Discussion

The large literature in econometrics about stochastic frontier models is mostly concerned with parametric models of the tail behavior of the response "near the production frontier." Nonparametric quantile regression offers yet another perspective on estimating such models. It would be extremely foolish to make any claims for alternative methodology described here on the basis of the flimsy evidence offered, let me conclude simply by saying that it might be worthy of further consideration.

Declarations

References

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic Frontier production function models. J Econom 6:21–37

Amsler C, Prokhovov A, Schmidt P (2016) Endogeneity in stochastic Frontier models. J Econom 190:280–288

Andrews D, Ploberger W (1994) Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62:1383–1414

Clarke BS, Clarke JL (2018) Predictive statistics: analysis and inference beyond models. Cambridge University press, Cambridge

Davies R (1977) Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64:247–254

Davies R (1987) Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74:33–43

Foygel Barber R, Candès EJ, Ramdas A, Tibshirani RJ (2020) The limits of distribution-free conditional predictive inference. Inf Inference A J IMA 10:455–482

Frisch R, Waugh FV (1933) Partial time regressions as compared with individual trends. Econometrica 1:387–401

Geiser S (1993) Predictive inference: an introduction. Chapman and Hall, London

Hansen B (1996) Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64:413–430

Horrace W, Schmidt P (1996) Confidence statements for efficiency estimates from stochastic Frontier models. J Product Anal 7:257–282

Hotelling H (1939) Tubes and Spheres in n-spaces, and a class of statistical problems. Am J Math 61:440–460

Johansen S, Johnstone IM (1990) Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. Ann Stat 18:652–684

Kendall M (1961) Course in the geometry of n dimensions. Hafner, Ohio

Kim P, Naiman D, Li Q, Stengos T (1998) On Hotelling's approach to hypothesis testing when a nuisance parameter is present only under the alternative. J Econom Theory Econom 4:105–130

Knowles M (1987) Simultaneous confidence bands for random functions. Ph.D. thesis, Stanford University

Koenker R (1999) Quantreg: an R package for quantile regression methods. https://cran.r-project.org/web/packages/quantreg/index.html

Koenker R (2011) Additive models for quantile regression: model selection and confidence bandaids. Braz J Probab Stat 25:239–262

Krivobokova T, Kneib T, Claeskens G (2010) Simultaneous confidence bands for penalized spline estimators. J Am Stat Assoc 105:852–863

Lei L, Candès E (2022) Conformal inference of counterfactuals and individual treatment effects. J Royal Stat Assoc (B) Forthcom 34:1660

Naiman D (1986) Conservative confidence bands in curvilinear regression. Ann Stat 14:896–906

Nychka D (1983) Bayesian confidence intervals for smoothing splines. J Am Stat Assoc 83:1134–43

Romano Y, Patterson E, Candès E J (2019) Conformalized quantile regression. In: 33rd Conference on neural information processing systems Vancouver, Canada

Shafer G, Vovk V (2008) A tutorial on conformal prediction. J Mach Learn Res 9:371–421

Wahba G (1983) Bayesian confidence intervals for the cross-validated smoothing spline. J Royal Stat Assoc (B) 45:133–50

Weyl H (1939) On the volume of tubes. Am J Math 61:461–472