Ann. Inst. Statist. Math.
Vol. 52, No. 1, 123–138 (2000)

POSITRON EMISSION TOMOGRAPHY AND RANDOM COEFFICIENTS REGRESSION

ANDREY FEUERVERGER[1]* AND YEHUDA VARDI[2]**

[1]*Department of Statistics, University of Toronto, 100 St. George St., Toronto, Ontario, Canada M5S 3G3*
[2]*Department of Statistics, Rutgers University, 110 Frelinghuysen Rd. Piscataway, NJ 08854-8019, U.S.A.*

(Received December 15, 1997; revised August 3, 1998)

Abstract. We further explore the relation between random coefficients regression (RCR) and computerized tomography. Recently, Beran *et al.* (1996, *Ann. Statist.*, **24**, 2569–2592) explored this connection to derive an estimation method for the nonparametric RCR problem which is closely related to image reconstruction methods in X-ray computerized tomography. In this paper we emphasize the close connection of the RCR problem with positron emission tomography (PET). Specifically, we show that the RCR problem can be viewed as an idealized (continuous) version of a PET experiment, by demonstrating that the nonparametric likelihood of the RCR problem is equivalent to that of a specific PET experiment. Consequently, methods independently developed for either of the two problems can be adapted from one problem to the other. To demonstrate the close relation between the two problems we use the estimation method of Beran, Feuerverger and Hall for image reconstruction in PET.

Key words and phrases: Computerized tomography, fast Fourier transform, nonparametric likelihood, positron emission tomography, projection-slice theorem, Radon transform, random coefficients regression, regularization, smoothed EM algorithm.

1. Introduction and summary

In this paper we explore the connection between random coefficients regression (RCR)—which is of interest in econometrics and other fields—and the image reconstruction problem in positron emission tomography (PET)—a nuclear medicine technique for imaging of, among other things, the brain metabolism. Specifically, we show that the nonparametric likelihood function for the RCR problem is the same as that of a PET reconstruction problem with a suitably specified conditional probability of detection function "$p(b, d)$", in the formulation of Shepp and Vardi (1982), and Vardi *et al.* (1985). This relation automatically enriches both areas of application by making methodologies that were independently developed for one area readily adaptable to the other. In particular, many of the statistical methods for image reconstruction in PET and SPECT (single photon emission tomography) which were developed in recent years, including convolution backprojection, regularized inversions, Bayes methods, penalized likelihood, smoothed EM algorithms, and sieve methods can be modified to suit the RCR

* The work of AF was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. AF also thanks Bell Laboratories, Murray Hill, NJ, for travel assistance and research facilities provided.

** The work of YV was supported in part by grants from the National Security Agency and the National Science Foundation.

problem. (See, for instance, Green (1990*a*, 1990*b*), Silverman *et al.* (1990), Iusem and Svaiter (1994), Eggermont and LaRiccia (1995), Chang and Hsiung (1994), O'Sullivan (1995), Gehman and McClure (1987), Higdon *et al.* (1994) and references therein for some recent methodologies.) On the other hand, methods originally developed for RCR can also be modified to suit PET reconstruction problems. We demonstrate this below by showing that the recently proposed nonparametric method for RCR estimation in Beran *et al.* (1996) can be successfully used for image reconstruction in PET.

Consider the random coefficients regression model where we are given data (X_i, Y_i), $i = 1, \ldots, n$ generated according to the model

$$Y_i = A_i + B_i X_i, \quad i = 1, \ldots, n. \tag{1.1}$$

There is no explicit error term in this model, however the (unobserved) coefficients (A_i, B_i) are assumed to be independent random vectors generated from some unknown distribution having bivariate density $f_{AB}(a, b)$, say. The X's may be chosen by design, or may be sampled randomly from a density $f_X(x)$, say, and independently of the (A_i, B_i)'s. The problem of interest is to estimate f_{AB} from the observed (X_i, Y_i)'s. In the present context we are specifically concerned with the case where f_{AB} is permitted to be an arbitrary member of a nonparametric class of density functions.

Random coefficients linear regression models occur naturally in econometric sampling, in studies of panel data, and in semiconductor materials research, to mention just a few applications. For further background and references see, for example, Longford (1993), Nicholls and Pagan (1985), and Raj and Ullah (1981). These models are useful in contexts where regression coefficients cannot be assumed constant and a less restrictive approach is required. Nevertheless, the RCR model imposes more structure on the data generation mechanism than may at first appear. Observe however that the usual (non-random coefficient) regression model corresponds to the special case of (1.1) when the distribution of B_i is taken to be degenerate. Previous work on RCR has considered only certain special cases, and not the fully non-parametric context. See also Beran (1993), Beran and Hall (1992) and Beran and Millar (1994).

To appreciate the connection between computerized tomography—which is an intensely researched and highly developed field in medical imaging—and the problem of nonparametric random coefficients regression, two relationships are of special importance: Firstly, since $Y = A + BX$, we have that

$$P(Y \le y \mid X = x) = P(A + Bx \le y) = \int_{-\infty}^{\infty} \int_{-\infty}^{y-\beta x} f_{AB}(\alpha, \beta) d\alpha d\beta;$$

therefore the conditional density function of Y, given that $X = x$, is found to be

$$f_{Y|X}(y \mid x) \equiv \int_{-\infty}^{\infty} f_{AB}(y - \beta x, \beta) d\beta. \tag{1.2}$$

Secondly, since the conditional characteristic function $\phi_{Y|X}$ for Y, given that $X = x$, is

$$E(e^{itY} \mid X = x) = E e^{it(A+Bx)}, \tag{1.3}$$

we have that

$$\phi_{Y|X}(t \mid x) \equiv \phi_{AB}(t, tx), \tag{1.4}$$

where ϕ_{AB} is the characteristic function of (A, B). Equation (1.2) in effect says that $f_{Y|X}$ is the Radon transform of f_{AB}, while equation (1.4) in effect is the so-called *projection-slice* theorem of Radon transform theory. (In particular, see the transformation (3.4) and equation (3.5) below for further amplification of these points.) Consequently the determination of the density function f_{AB} from the data (X_i, Y_i) can be viewed, via (1.2), as a problem of inverting a Radon transform, but with the added complication that the conditional density $f_{Y|X}$ (which essentially is the Radon transform) is unknown and can only be estimated from the data. For some background on Radon transforms see, for example, Deans (1983).

There is another complication relative to analytical Radon inversion, and that is that the projection angles will not be evenly represented in random coefficients regression data because of the randomness of the slopes (in (a, b)-space) of the lines which are determined by the (X_i, Y_i) data points. Indeed the distribution of these angles is determined by f_X; and only if f_X is a standard Cauchy density are the angles uniformly distributed. To understand the key issues here note that the standard Radon transform inversion theory assumes that all line integrals of the original function are given. In particular, for any given direction, all the integrals along parallel lines in that direction are assumed to be given. In applications, where the number of lines that occur is always finite, the theory as it stands is therefore not directly applicable. To close this gap between theory and practice, one needs to estimate the line integrals for all possible lines, and this essentially is done by smoothly averaging over the available line integrals using methods similar in spirit to kernel density estimation. The variability of any estimated line integral then depends on the distribution of the parameters of the lines. For instance, if there is a direction which is relatively empty of lines (this will be determined by the distribution of X) then the estimated values of the line integrals in that direction, will have much higher mean squared error than similar estimates where there is high concentration of lines. Thus if the projection angles are evenly distributed, our approximation of the continuous case from the discrete case will have approximately constant variance, and otherwise it will not. The former case will obviously result in the better overall statistical properties. This fine detail suggests a close similarity of RCR to PET problems, more so than to X-ray CT, as we shall further explore below.

It is useful, at this stage, to focus on the comparative geometries of the RCR and PET problems. Consider firstly the situation in random coefficient regression, depicted in Figs. 1a–1d. The panel (1a) shows a single point (A_i, B_i), equal to $(-1, 2)$ in this instance. The value of X_i was taken to equal 1 here, so that $Y_i = 1$. In practice, the point (A_i, B_i) shown here will not be known, but (X_i, Y_i) will be. The line 'L_i' shown in (1a) consists of the set $\{(a, b) : Y_i = a + bX_i\}$ of values for (A_i, B_i) that are consistent with what has been observed. The panel (1b) shows ten such points and lines. These were generated by taking the A_i and B_i to have independent standard normal distributions, and by taking X_i (independently of the A's and B's) also to have a standard normal distribution. In practice, the situation will be as illustrated in panels (1c) and (1d); these no longer show the (A_i, B_i)'s, but otherwise were generated in the manner just described. Panel (1c) corresponds to a data set of only 50 points, while panel (1d) corresponds to 300 points—which is a more realistic sample size for the nonparametric context. Figures 2a–2d depict the corresponding situation for positron emission tomography (as further described in Section 2 below). In these figures, the circle (chosen to have unit radius here for simplicity) represents the 'detector ring' which surrounds the skull being imaged, while the points shown represent locations where positron emissions have occurred. These points of emission are not observed, however the (nearly) straight lines defined from the resulting flight (at nearly the speed of light) of a

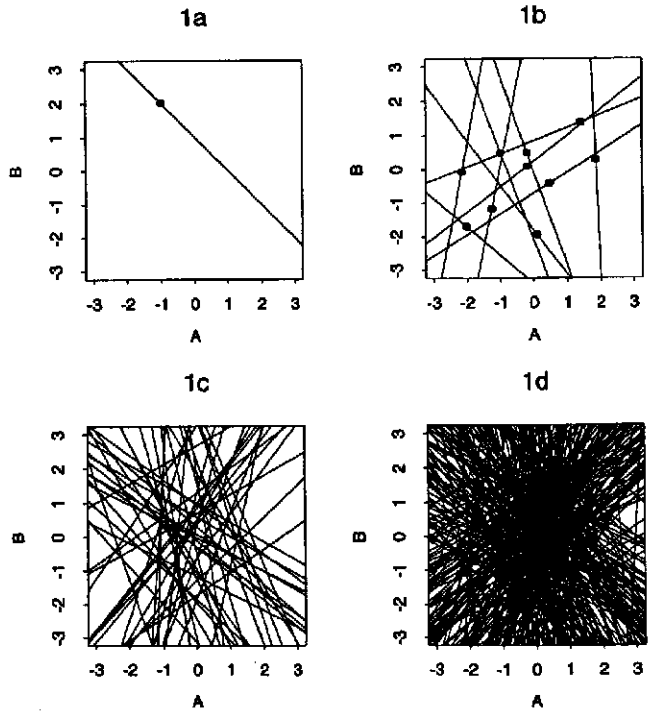

Fig. 1. The random coefficient regression context.

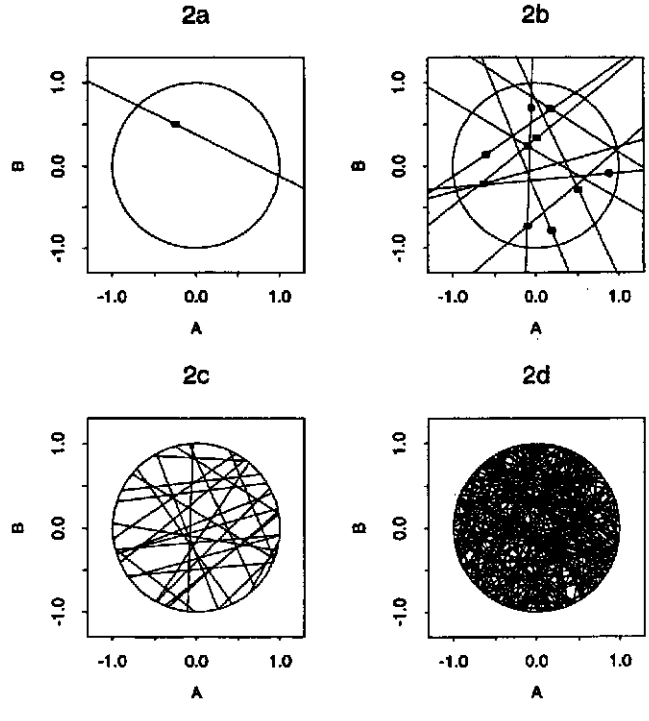

Fig. 2. The positron emission tomography context.

pair of photons in opposite directions are observed; in fact these lines become determined as photon pairs cross the detector ring at (nearly) the same instant. These lines have uniformly random angular orientation. (This will be the case in RCR also if the X_i's are standard Cauchy distributed.) Panel (2a) shows the situation for a single emission point only, while panel (2b) shows ten such points which here were chosen uniformly at random within the sphere, and lines (chords) drawn through these points with uniformly random angular orientations. Panels (2c) and (2d) show only the observed lines, and not the unobserved points, using 25 and 200 lines respectively. In real PET experiments, the number of these lines must be many orders of magnitude larger. The fundamental connexion between the RCR and PET statistical problems is that in each case the data which arises may be regarded as consisting of independently sampled random lines; these occur as randomly oriented lines through randomly chosen points, and it is the nonparametric distribution of the random points which is of primary interest.

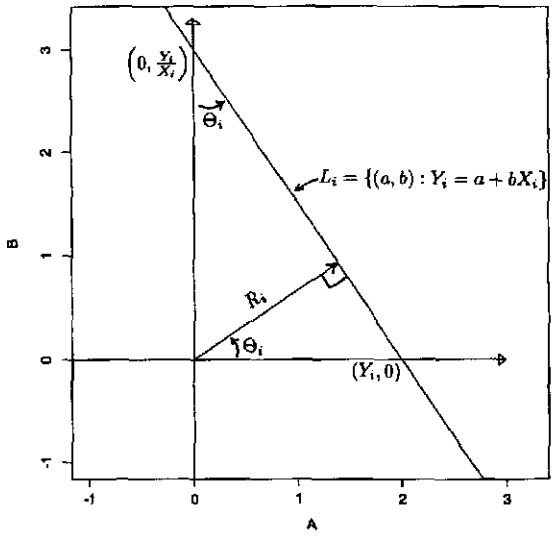

Fig. 3. Parameterization of lines in RCR.

It should also prove helpful, throughout the discussion below, to refer to Figs. 3 and 4 which illustrate the parameterizations to be adopted here for the random lines in the RCR and PET contexts respectively. In the RCR context, the line $L_i \equiv \{(a,b) : Y_i = a + bX_i\}$ will be parameterized via its 'signed distance' R_i from the origin, and the corresponding angle Θ_i as shown; here we take $\Theta_i \in (-\pi/2, \pi/2]$, so that $R_i \in (-\infty, \infty)$. Note that R_i equals both $Y_i \cos\Theta_i$ as well as $(Y_i/X_i)\sin\Theta_i$ so that $X_i = \tan\Theta_i$. Therefore $\Theta_i = \arctan X_i$ and $R_i = Y_i/\sqrt{1 + X_i^2}$. In the PET context (Fig. 4) the line (chord) L_i is determined by the positions of two detectors on the detection ring, located at angles φ_{i1} and φ_{i2}, say, as shown. This line is again parameterized by its signed distance R_i from the origin, and corresponding angle Θ_i shown, with the ranges for R_i and Θ_i being as before. Note that the angle Θ_i here will be the average of the angles φ_{i1} and φ_{i2} modulo π, while R_i will be the inner product of $(\cos\Theta_1, \sin\Theta_1)$ with either $(\cos\varphi_{i1}, \sin\varphi_{i1})$ or $(\cos\varphi_{i2}, \sin\varphi_{i2})$, since our detector ring has unit radius.

The tomographic viewpoint for RCR was first noted in Beran *et al.* (1996). In that paper, a methodology for the nonparametric estimation problem in RCR was developed

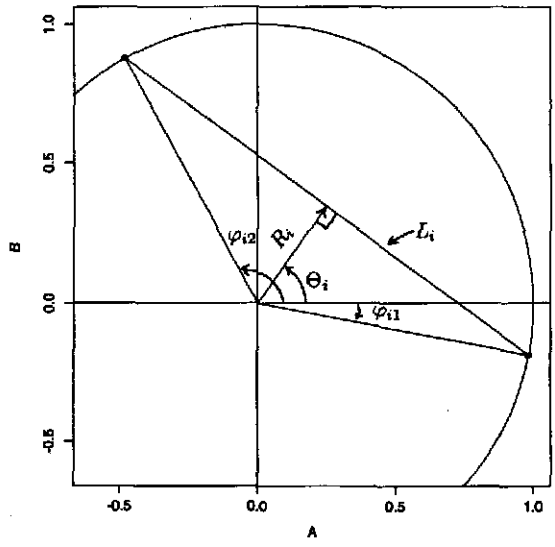

Fig. 4. Parameterization of lines in PET.

based on mathematical methods used in transmission tomography. In particular, that paper implements a direct-type Fourier reconstruction approach based on the projection-slice result (1.4). Our focus in the present paper will be on the connections between RCR and PET. Specifically, in Section 2 we show that the nonparametric likelihood functions for RCR and for a particular (idealized) PET experiment are in fact equivalent. In Section 3 we show explicitly how methodologies developed for either RCR or PET can be carried over from one problem to the other. Finally, as an example of this, in Section 4 we carry out a numerical implementation for PET by means of the RCR algorithm of Beran *et al.* (1996) using the simulated data of Vardi *et al.* (1985). Numerical implementation of RCR based on the EM or other algorithms for PET is, of course, also possible but will not be undertaken here.

2. Nonparametric likelihoods for RCR and PET

In this section we develop the nonparametric likelihood functions for RCR and for a particular (idealized) PET experiment. The two likelihood functions differ only in that the angular orientations for the PET experiment are assumed to be known and uniform, but are otherwise seen to be equivalent.

2.1 *The nonparametric likelihood for RCR*

Suppose that the RCR vectors (A_i, B_i) are sampled from the density $f_{A,B}(a,b)$, and that (independently of the A's and B's) the X_i's are sampled from the density $f_X(x)$. Then the likelihood of the data (X_i, Y_i), $i = 1, \ldots, n$, as a function of the unknown "parameters" $f_{A,B}$ and f_X, is given by

$$\mathcal{L}''(f_{A,B}, f_X \mid x_i, y_i, i = 1, \ldots, n) = \prod_{i=1}^{n} f_X(x_i) \times \prod_{i=1}^{n} f_{Y|X}(y_i \mid x_i).$$

Now from (1.2), the conditional density $f_{Y|X}(y_i \mid x_i)$ of Y_i given X_i, is given by

$$\int_{-\infty}^{\infty} f_{AB}(y_i - \beta x_i, \beta) d\beta;$$

on changing variables of integration, this can be written as

$$\frac{1}{\sqrt{1+x_i^2}}\int_{-\infty}^{\infty} f_{AB}\left(y_i - \frac{x\beta}{\sqrt{1+x_i^2}}, \frac{\beta}{\sqrt{1+x_i^2}}\right) d\beta = \frac{1}{\sqrt{1+x_i^2}}\int_{\ell_i} f_{A,B}(a,b)ds$$

where ℓ_i is the line $a = y_i - x_i b$ in (a,b)-space, and ds is an element of length along that line. (Note that this change of variables is related to the transformation of variables defined below at (3.4).) Since factors of the likelihood function in the x's have no bearing on the inference for $f_{A,B}$ they may be removed. Thus the conditional likelihood of $f_{A,B}$, given $\{X_i = x_i, i = 1, \ldots, n\}$, can be taken to be either

$$\mathcal{L}' = \prod_{i=1}^{n}\int_{-\infty}^{\infty} f_{AB}(y_i - \beta x_i, \beta)d\beta \tag{2.1}$$

or, alternatively,

$$\mathcal{L} = \prod_{i=1}^{n}\int_{\ell_i} f_{A,B}(a,b)ds. \tag{2.2}$$

Note that the terms in these products are in fact values of the Radon transform of $f_{A,B}$.

We shall next relate the conditional likelihood $\mathcal{L}$ to that of a particular PET experiment. We start with a brief review of the mathematical model for a PET reconstruction problem. More details (including medical applications) can be found for instance in Vardi *et al.* (1985).

2.2 *The mathematical model for PET (idealized)*

Consider a (nonhomogeneous) spatial Poisson process with a two-dimensional intensity function $\lambda(x,y)$, where $(x,y) \in \mathcal{B}$. The set $\mathcal{B}$, called the "brain space", is a closed bounded region in R^2 containing the planar head section that is being imaged. The Poisson events, occurring in the set $\mathcal{B}$, are points of emission of positrons from a radioactive labeled metabolite which is distributed in brain space according to λ. When the positrons are emitted they find a nearby electron to annihilate with, and as a result a pair of photons, originating at the point of annihilation, fly off in (nearly) opposite direction along an (approximately) straight line having a uniformly random orientation. Ignoring second order effects, such as the distance traveled by the positron from point of emission to point of annihilation, and the slight deviation from linearity, we see that each emission gives rise to a line in random orientation defined by the flight of the photons. The observed data are these random lines. The points of emission themselves are not observed. The problem is to estimate the intensity function $\lambda(x,y)$ on the basis of the data consisting of these lines.

We note that in real PET experiments these lines would be grouped according to cylindrical regions ("detector tubes") defined by matching pairs of detectors that the lines fall into. These detectors—which are arranged on a ring surrounding the planar head section—have light-sensitive crystals and photo-multipliers, and are wired for time-coincident detection of the photons. We shall use the term *idealized* PET experiment here for the case where the actual lines (not grouped) are considered to be observed, together with the other simplifying approximations indicated above.

2.3 *Nonparametric likelihood of the (idealized) PET experiment*

The contribution to the conditional likelihood of a specific line ℓ, in the idealized PET experiment, given that a line is observed, is

$$\frac{1}{\pi(1+\beta_\ell^2)} \int_\ell \frac{\lambda(x,y)}{\Lambda(\mathcal{B})} ds$$

where ds is an infinitesimal increment of length along the line ℓ, and β_ℓ is the observed slope of ℓ relative to a fixed axis, and

$$\Lambda(\mathcal{B}) = \int_{\mathcal{B}} \lambda(x,y) dx dy.$$

This may be seen as follows. Given an observed line ℓ, the pdf for the emission to have occurred at a point (x,y) on ℓ is $\lambda(x,y)/\Lambda(\mathcal{B})$ and the pdf for the slope to be β_ℓ is $1/\pi(1+\beta_\ell^2)$ (same for all (x,y)). This is then integrated over all points (x,y) on the observed line ℓ. Note, incidentally, that if instead of observing the slope of the line, we observe (equivalently) its angular orientation instead, then the factor $1/\pi(1+\beta_\ell^2)$ in the likelihood would be replaced by $1/2\pi$. The two resulting likelihoods—which correspond to a one to one transformation on the data—will of course be equivalent.

Now, in a PET experiment $\Lambda(\mathcal{B})$ is a function of the original dosage of the radioactively labeled metabolite and the duration of the scan and is of no particular interest (it can be easily estimated by the total Poisson event count) while the quantity of interest, of course, is the *relative* intensity $\lambda(x,y)/\Lambda(\mathcal{B})$. Thus, without loss in generality, we can simply take $\Lambda(\mathcal{B}) = 1$, and the conditional likelihood based on the observed lines $\ell_1, \ldots, \ell_n$ is then proportional to

$$\prod_{i=1}^{n} \int_{\ell_i} \lambda(x,y) ds. \tag{2.3}$$

Note that this is of the same form as $\mathcal{L}$ at (2.2) with $\lambda/\Lambda(\mathcal{B})$ now playing the role of $f_{A,B}$.

The reason why this conditional likelihood is different from the one commonly appearing in most PET literature lies in the 'idealization' of the experiment. In particular, there are two levels of quantization (discretization) that distinguish the situation here from the 'standard' PET model: (a) the grouping of the lines (into 'detector tubes') and (b) the pixelization of the set $\mathcal{B}$. The grouping (a) is typically introduced into the PET model to conform to the data collection mechanism, which registers lines as they cross a pair of scintillation detectors in the scanner ring that surrounds the region $\mathcal{B}$. The pixelization (b) is usually introduced into the model for a variety of mathematical and statistical justifications, such as numerical computability and statistical consistency. (See, for example, the discussion in Chapter 7 of Bickel *et al.* (1993) regarding inconsistency of the nonparametric MLE and regularization, and Chang and Hsiung (1994) regarding consistency of the nonparametric MLE in the PET model when pixelization is viewed as sieving.)

3. Estimation

In this section we discuss the implementation and transfer of methodologies from one discipline to the other. We discuss first how maximum likelihood, based for example on the EM algorithm for PET, can be adopted to RCR. Our focus will be on how to fit the *idealized* version of PET (which is equivalent to RCR) to the standard PET methodology, thereby making the numerous PET reconstruction approaches relevant to RCR. We then continue with a discussion of transferring RCR methodology to PET. Here we will focus on a recent methodology, developed in Beran *et al.* (1996) and show how it can be applied to the PET reconstruction problem. As an example, we shall demonstrate the latter method with an actual RCR-based PET reconstruction in Section 4. Computer implementation of RCR via the EM algorithm for PET is left for future research. For further background on the EM algorithm, see, for example, Dempster *et al.* (1997).

3.1 *Implementation of PET methodology to RCR*

We assume that the support of the unknown RCR density $f_{A,B}(a,b)$ is bounded and contained in some known region $\mathcal{B}$. In practice this assumption generally is not very restrictive as the unobserved (A_i, B_i)'s will necessarily be contained in a bounded region. Without loss of generality we assume that $\mathcal{B}$ is a square; in fact, with the appropriate scaling, we can take $\mathcal{B} = \{(a,b); |a| \leq 1, |b| \leq 1\}$. Consider now a circle surrounding $\mathcal{B}$, which is partitioned into D arcs of equal length. This is the analogue of our 'detector ring'. A line through $\mathcal{B}$ must cross two of the 'detectors' on the ring and hence lie in the corresponding 'tube'. (The probability of falling on a boundary between detectors is zero for continuously distributed line angles.) We can ignore the possibility of a line crossing the same detector twice, as we assume that the detector arclengths are small and that the region $\mathcal{B}$ is strictly interior to the disc defined by the 'detector ring'. Let $n^*(d)$ be the number of lines crossing the detector pair $d = (d_1, d_2)$. The random coefficient pair (a,b) will be detected at the tube $d = (d_1, d_2)$ only if the intercept y and slope $-x$ of the line $a = y - xb$ are such that the line crosses both d_1 and d_2. Thus the probability that a pair of random coefficients (a,b) is 'detected' at $d = (d_1, d_2)$ is just the probability that the line $a = y - xb$ crosses the detectors d_1 and d_2. As the partition of the ring becomes finer ($D \uparrow \infty$), this probability would infinitesimally tend to $f_X(-\text{slope}(d_1, d_2))dx$ if the infinitesimally narrow strip (detector tube) defined by $d = (d_1, d_2)$ covers (a,b), and zero otherwise, where dx corresponds to the range of the 'angle of view' into the tube. (In practice, this 'angle of view' is usually approximated by computing it geometrically from the center of the pixel—see, e.g. Vardi *et al.* (1985).)

For consistency of the estimators we want to assume that $f_{A,B}$ is piecewise constant on the pixels formed by a suitable grid. We shall denote by j the index which ranges over these pixels, and—slightly abusing our notation—by $f_{A,B}(j)$ the constant value of $f_{A,B}$ on the j-th pixel.

With this notation, and with a derivation similar to that in Vardi *et al.* ((1985), Section 1), the RCR data obtained via the detector ring, $\{n^*(d)\}$ say, may be viewed as being independent Poisson random variables having means

$$EN^*(d) = \sum_j f_{A,B}(j) P_X(j,d) \tag{3.1}$$

where

$$P_X(j,d) = \frac{1}{\text{Area}(\text{Pixel}(j))} \int_{(a,b)\in\text{Pixel}(j)} \int_x f_X(-x) I[\ell(a,b;x) \in d] dx da db \tag{3.2}$$

and $\ell(a,b;x)$ is a line through (a,b) with slope $-x$. Note that $P_X(j,d)$ is the probability of detection at d of a random coefficients pair (A,B) originating within pixel j. With this interpretation, the RCR problem takes on the structure of an incomplete Poisson data problem, where $f_{A,B}(\cdot)$ is the unknown intensity, and $P_X(j',d)$ describes the random mapping from the (unobserved) complete RCR data (i.e. the (A,B) pairs) to the (observed) incomplete data (i.e. $\{n*(d)\}$), as it specifies the chance of an (A,B) pair *originating in the vicinity of pixel j' to land at the arclength-pair* (i.e. detector-pair) labeled "d". The EM algorithm as discussed, for instance in Vardi *et al.* (1985), or Lange and Carson (1984), is now applicable to this maximum likelihood problem and leads to the following basic EM iterative algorithm for RCR:

$$\hat{f}_{A,B}(j) \leftarrow \hat{f}_{AB}(j) \sum_d \frac{n^*(d) P_X(j,d)}{\sum_{j'} \hat{f}_{A,B}(j') P_X(j',d)}. \tag{3.3}$$

This algorithm is iterated until numerical convergence is attained.

We note that the discretization of $\mathcal{B}$ into pixels and of the 'detector ring' into tubes can be viewed as a form of sieving for a nonparametric estimator (Chang and Hsiung (1994)). In many RCR applications one may want to assume a smooth density for (A,B). In such cases, some recent reconstruction techniques which are strongly geared toward smoothed reconstruction could be very useful. See Green (1990*a*, 1990*b*), Silverman *et al.* (1990), Eggermont and LaRiccia (1995) and Iusem and Svaiter (1994) for some recent modifications of the EM procedure to achieve smooth estimates. We also note that in typical RCR applications where f_X will be unknown, it must be replaced by an estimate which can be derived from the x_i's sample. Thus the data will enter the estimation procedure in two different ways here: once in determining the $n^*(d)$, and once in the calculation of the $P_X(j,d)$ based on the estimated f_X.

3.2 *Implementation of RCR methodology to PET*

Consider now the RCR context, and in particular the meaning of equations (1.3) and (1.4). These equations tell us that if we have an observation $X_j = x$, say, and the corresponding Y_j, then $\exp(itY_j)$—which we will regard here as being a function of t—*is an unbiased estimator of* $\phi_{AB}(t,tx)$. *In other words,* $\exp(itY_j)$ may be regarded as an unbiased estimator—based on a single degree of freedom—for the function $\phi_{AB}(t,u)$ along the line in the (t,u)-Fourier-space that is traced out by (t,tx). This line through the origin, shown in Fig. 5a, has slope x and angle $\arctan(x)$ relative to the t-axis. The situation in practice is as depicted in Fig. 5b, where we have one-degree-of-freedom estimators for $\phi_{AB}(t,u)$ along as many lines in (t,u)-space as there are data points. In order to estimate ϕ_{AB} at a particular value of (t,u), we somehow need to average the estimates from suitably selected points on nearby lines. It is clear from this discussion that methods related to nonparametric regression (in the Fourier space) will play a useful role in the estimation procedure for the nonparametric RCR problem, and furthermore that the distribution of the X_j's must have support on the whole real line in order that it be possible to consistently estimate $\phi_{AB}(t,u)$ for *any* (t,u), *regardless of the* 'angle' at which it is located.

We can now describe our algorithm for nonparametric random coefficient regression. In the RCR model $Y = A + BX$, define $\Theta = \arctan X$ and $R = Y/\sqrt{(1+X^2)} \equiv Y\cos\Theta$. Then with this transformation of variables

$$R = \frac{A+BX}{\sqrt{1+X^2}} = (A,B)\cdot(\cos\Theta, \sin\Theta), \tag{3.4}$$

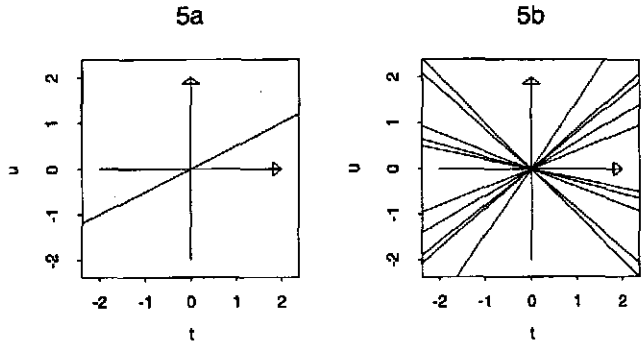

Fig. 5. View of the (t, u)-Fourier space for the characteristic function $\phi_{AB}(t, u)$.

so that

$$\phi_{R|\Theta}(r \mid \theta) = \phi_{A,B}(r\cos\theta, r\sin\theta), \tag{3.5}$$

for $-\infty < r < \infty$ and $-\pi/2 < \theta \leq \pi/2$. Here $\phi_{R|\Theta}$ is the characteristic function of R conditional on $\Theta = \theta$, and $\phi_{A,B}(\cdot,\cdot)$ is the joint characteristic function of A and B. Observe from (3.4) that the conditional density of R given Θ is precisely the Radon transform of the density, $f_{A,B}$, of the coefficients A and B, and also that (3.5) is precisely the *projection-slice* theorem. (This theorem states that the one-dimensional Fourier transform (in the radial variable) of the Radon transform of a two-dimensional function is identical to the two-dimensional Fourier transform of that function.) Note that these assertions are in fact equivalent to (1.2) and (1.4).

Our procedure is based on first obtaining an estimate of $\phi_{R|\Theta}$, then of $\phi_{A,B}$ using (3.5), and finally on numerical Fourier inversion to obtain an estimate of the density $f_{A,B}$. We have available the data (X_j, Y_j), or equivalently, the corresponding (R_j, Θ_j) for $j = 1, 2, \ldots, n$. The conditional characteristic function $\phi_{R|\Theta}(r \mid \theta)$, which is just $E(\exp\{irR\} \mid \Theta = \theta)$, may be estimated from this data by essentially averaging those values of $\exp\{irR_j\}$ that correspond to values of Θ_j which are "close" (in the appropriate sense) to θ. In essence, this is a problem in nonparametric regression, and as such may be approached in various ways. There is, however, the added feature that the regressor variable, Θ, has (except for conjugation which occurs at the boundary) an essentially circular structure.

Our approach is based on a straightforward adaptation to the circle of the locally linear regression algorithm described, for example, in Fan (1992). Thus let $K_\kappa(\cdot)$ be a kernel for local linear regression on the circle, where $\kappa = \kappa(\theta)$ is an (angular) bandwidth parameter. This function is taken to be π-periodic, even on $(-\pi/2, \pi/2]$, have its mode at $\theta = 0$, and to become more concentrated around 0 as $\kappa \to 0$. Write $\theta_1 \ominus \theta_2$ for $\theta_1 - \theta_2$ if this quantity lies in the interval $(-\pi/2, \pi/2]$, and for $\theta_1 - \theta_2 + \pi$ (respectively, $\theta_1 - \theta_2 - \pi$) if $\theta_1 - \theta_2 \leq -\pi/2$ (respectively, $> \pi/2$). The quantities $\theta_1 \ominus \theta_2$ appearing below will thus always fall in the interval $(-\pi/2, \pi/2]$. Then following Fan (1992), we define first

$$s_k(\theta) = \sum_{j=1}^{n} K_\kappa(\theta - \Theta_j)(\theta \ominus \Theta_j)^k, \tag{3.6}$$

for $k = 1, 2$, and then

$$u_j(\theta) = K_\kappa(\theta - \Theta_j)\{s_2(\theta) - (\theta \ominus \Theta_j)s_1(\theta)\}. \tag{3.7}$$

Finally, the proposed 'locally linear regression' estimator of $\phi_{R|\Theta}$ is

$$\hat{\phi}_{R|\Theta}(r \mid \theta) = \sum_{j=1}^{n} u_j(\theta) \exp\{irR_jS_j(\theta)\} \Big/ \sum_{j=1}^{n} u_j(\theta) \tag{3.8}$$

where $S_j(\theta) = \pm 1$ according as $\theta \ominus \Theta_j$ equals $\theta - \Theta_j$ or not, respectively. The presence of the S_j-factors, which are relevant primarily at the circle boundaries, may be understood most easily by appreciating how an observation (X_j, Y_j) corresponding to a large value of X_j, and hence to a value of Θ_j near to $\pi/2$, may be used to help estimate $\phi_{R|\Theta}$ at a value of θ close to $-\pi/2$, by means of negating the value of r.

The estimator (3.8) is based on smoothing in the angular variable of the polar coordinates in the Fourier domain. From this estimator and equation (3.5) we then obtain our estimator of $\phi_{A,B}$:

$$\hat{\phi}_{A,B}(t,u) = \hat{\phi}_{A,B}(r\cos\theta, r\sin\theta) = \hat{\phi}_{R|\Theta}(r \mid \theta) \tag{3.9}$$

where (r, θ) here and below are the polar coordinates corresponding to the cartesian variates (t, u), except that θ is restricted to lie in $(-\pi/2, \pi/2]$, so that r ranges over $(-\infty, \infty)$.

Next, a tapering function $w_\theta(r)$ is applied to the estimated $\phi_{A,B}$. This taper must, in effect, be 0 outside the bounded region in Fourier space where the estimator (3.9) is computed, and must have a Fourier transform with reasonable properties for its role as a smoothing (or convolution) kernel in density space. After tapering, basic two-dimensional FFT-based numerical inversion methods are then used to compute an estimate for $f_{A,B}$:

$$\hat{f}_{A,B}(a,b) = \frac{1}{(2\pi)^2} \iint \hat{\phi}_{A,B}(t,u) w_\theta(r) e^{-iat-ibu} du dv. \tag{3.10}$$

Numerical inversion involves, of course, both truncation and discretization. It is worth noting here that truncation is accounted for, in the expression above, by the presence of the taper function w_θ, and this, of course, results in some convolution-smoothing in the density domain. The discretization, on the other hand, ordinarily would result in aliasing of the image; however in the PET context the image is supported over a bounded area. It will therefore be easy to arrange things so that aliasing of PET images will not occur.

We now show how this algorithm can be adapted for reconstruction with PET data. Thus suppose that $j = 0, 1, \ldots, (D-1)$ now ranges over the detectors on our detector ring. We shall take $D = 128$ here, as this will correspond to the experiment whose data we will analyze below. This gives rise to $T \equiv D(D-1)/2 = 8128$ detector tubes which we index by $m = 1, 2, \ldots, T$. Each of the n_m coincidences detected in the m-th tube will be considered to correspond to the line defined by the center of that tube. It is necessary now only to correspond each of these PET-based (center) lines with an RCR-based value for Y and X, or rather, for R and Θ. Now the centres of the 128 detectors may be considered to be located at the angular positions $2\pi j/128$ radians on the unit circle, for $j = 0, \ldots, 127$. The tube center line corresponding to the detector pair j_1 and j_2 may thus be shown to correspond to that angle Θ which is the average of the two angles $2\pi j_1/128$ and $2\pi j_2/128$, but adjusted, modulo π, so as to lie within the interval $(-\pi/2, \pi/2]$. (See Fig. 4.) The corresponding value of R may then be obtained

by dotting a unit vector in the Θ direction with a unit vector in either the $2\pi j_1/128$ or the $2\pi j_2/128$ direction. For example:

$$R = \cos(\Theta)\cos(2\pi j_1/128) + \sin(\Theta)\sin(2\pi j_1/128). \tag{3.11}$$

The summations in the formulae for RCR given above can now be done over the $T = 8128$ tubes (instead of over the $n = 10^7$ lines which will be used in the example provided below) by introducing the tube counts data (i.e. the n_m factors) into the sums, in the appropriate way. Specifically then, (3.6) will change to

$$s_k(\theta) = \sum_{m=1}^{T} n_m K_\kappa(\theta - \Theta_m)(\theta \ominus \Theta_m)^k, \tag{3.12}$$

while $u_m(\theta)$ remains basically as in (3.7). The estimator $\hat{\phi}_{R|\Theta}$ at (3.8) changes to

$$\hat{\phi}_{R|\Theta}(r \mid \theta) = \sum_{m=1}^{T} n_m u_m(\theta) \exp\{irR_m S_m(\theta)\} \Big/ \sum_{m=1}^{T} n_m u_m(\theta) \tag{3.13}$$

while $\hat{\phi}_{A,B}$ and $\hat{f}_{A,B}$ remain as in (3.9) and (3.10).

As a final observation here, it is constructive to note that both the RCR and the PET problems are inherently 'ill-posed'. Much of the recent statistical literature attempting to regularize the maximum likelihood reconstruction in PET is focused on adding a penalty function to the likelihood, or on adopting a Bayesian approach by defining a *Markov random field structure for the data*. Some examples of this are given in Gehman and McClure (1987), Green (1990*a*, 1990*b*), Higdon *et al.* (1994), and Silverman *et al.* (1990). The approach we describe here regularizes the problem by first smoothing the raw data. In the PET context this corresponds to smoothing the random lines data by means of weighted averaging (or by binning them into detectors) with weights becoming smaller as the lines move away from the center of the detector, and then applying a tapered Fourier inversion (in accordance with the projection slice theorem) to the smoothed data. This is close in spirit, but different in details, to the convolution back-projection method that is often employed in tomography (especially in X-ray CT, but also in emission tomography). Relevant discussions can be found, for example, in Deans (1983), O'Sullivan (1995), and Shepp and Kruskal (1978). The other direction, namely applying penalized likelihood or Bayesian reconstructions developed for PET to RCR problems, would entail assuming a Bayesian prior on the distribution of the coefficients, or introducing a penalty component to the (nonparametric) RCR likelihood. Through some clever methods of numerical analysis, for example as in Green (1990*a*, 1990*b*), one can solve (approximately) a complex Bayesian estimation problem in RCR using PET 'technology', such as Green's 'one-step-late EM algorithm'.

4. Numerical implementation for PET

As one example of how RCR and PET methods may be adapted and transfered from either problem to the other, in this section we provide the results of a numerical implementation of the nonparametric RCR estimation algorithm as applied to PET data in the manner outlined in Subsection 3.2. The specific example of PET data used here is, for the sake of comparability, the same as in the simulation described by Vardi *et al.* (1985) and in Kaufman (1993). This is based on the rate function (emission density)

corresponding to the phantom used by Vardi *et al.* (1985) (and also Kaufman (1993)) and shown here in Fig. 6, and on a simulated count of $n = 10^7$ positron emissions from this density. Through each of the simulated positron emissions, a line was drawn with uniformly distributed angular orientation, and for each line, the pair of detectors (of the 128 on the detector ring) through which it passes was determined. These data (i.e. the 10^7 lines) were then grouped into the 8128 possible detector tubes. For purposes of direct comparability, the actual data realization used here (i.e. the tube counts) is identical to the one used in Vardi *et al.* (1985). It turns out that only 3277 of the 8128 tubes actually contain nonzero counts for this data set.

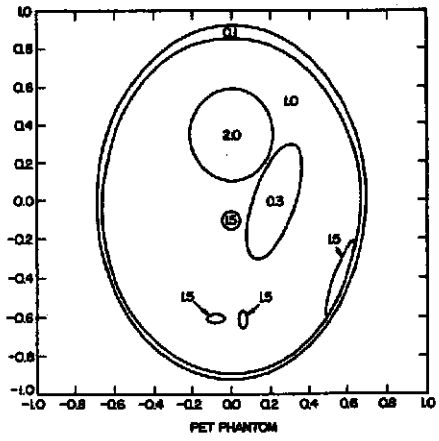

Fig. 6. The phantom for the PET simulation experiment.

For each tube with a nonzero count, we computed R_m and Θ_m, corresponding to the tube center line, as indicated at, and just prior to, (3.11). Next we performed the computations (3.12), (3.7), (3.13) and (3.10) respectively. This last computation was performed by means of a two-dimensional FFT algorithm applied on a grid of dimension 512×512 (this being the largest FFT grid which our implementation of the algorithm in the S-Plus statistical package could comfortably handle) extending over a range of ± 800 in each of the (cartesian) Fourier coordinates t and u. On inversion, this yields a grid of 512×512 pixels, each having dimension $\pi/800 \times \pi/800$ or $.00393 \times .00393$ squared units. These choices provide satisfactory pixel dimensions for a data set of this size, and also (since $.00393 \times 512 = 2.01$ exceeds each dimension of the image) allow the entire image to be produced, without aliasing effects.

We used the radially symmetric function tapering function

$$w(r) = \exp\left\{c\left[1 - \frac{1}{1-(r/r_1)^2}\right]\right\} = \exp\left\{-c\sum_{j=1}^{\infty}\left(\frac{r}{r_1}\right)^{2j}\right\}, \tag{4.1}$$

for $|r| < r_1$, and zero otherwise, with the values $r_1 = 180$ and $c = 4$ being used in our reconstruction. For convenience, we used the same functional form for the kernel K_κ of the locally linear regression on the circle; because the layout of tubes and detectors appears essentially the same when viewed from any angle, we set the angular bandwidth parameter $\kappa = \kappa_\theta$ to be constant in θ. The tube center-lines have a range of 128

(uniformly spaced) angular orientations here, corresponding to a spacing of $\pi/128 = .0245$ radians, so that the angular bandwidth must be larger than this; we used the value $\kappa = 0.125$ radians (same in all directions).

We followed the protocol of Vardi *et al.* (1985) by linearly mapping the reconstruction to a grey scale from 0 (white) to 127 (black), then remapping the values 0–30 into 0, and 70–127 into 127. The resulting reconstruction is shown here in Fig. 7. (This image was produced with a Hewlett Packard HP4Si printer using 8-shades of grey.) This figure may be compared to Fig. 3 of Vardi *et al.* (1985), and with Pictures 2–5 in Kaufman (1993).

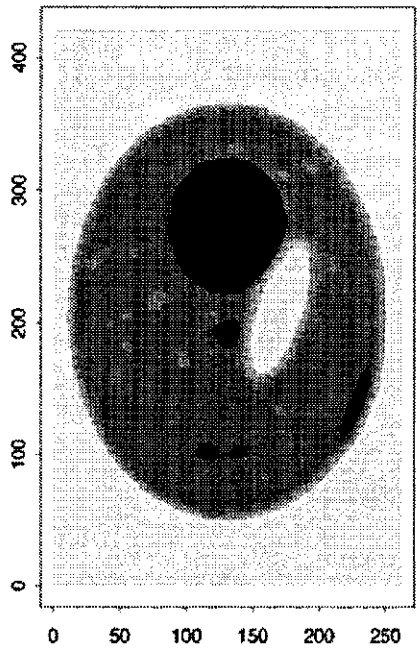

Fig. 7. PET reconstruction by means of the RCR-based algorithm.

Strictly speaking, our display parameters have to some extent been tuned for this particular set of data, however we note that the resulting image actually is quite stable over reasonably broad ranges of the two bandwidths which are the main tuning parameters here. If these same parameter values were applied to a new simulation from this image, we expect the overall result to be only slightly degraded. In practice, of course, the values of these parameters would have to be tuned for the type of images and sample sizes occurring in practice.

Finally, we remark that the computations here were carried out using the S-Plus statistical package (version 3.3) (see Becker *et al.* (1988)) on an SGI Challenge computer. Our implementation, which involves a 512×512 computational loop in S Plus, and therefore is not highly efficient, required approximately 5 minutes of CPU time to produce the image in Fig. 7.

Acknowledgements

The authors wish to thank Rudy Beran and an anonymous referee for their careful reading of the manuscript.

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *The New S Language*, Wadsworth & Brooks/Cole, Pacific Grove, California.

Beran, R. (1993). Semiparametric random coefficient regression models, *Ann. Inst. Statist. Math.*, **45**, 639–653.

Beran, R. and Hall, P. (1992). Estimating coefficient distributions in random coefficient regressions, *Ann. Statist.*, **20**, 1970–1984.

Beran, R. and Millar, P. W. (1994). Minimum distance estimation in random coefficient regression models, *Ann. Statist.*, **22**, 1976–1992.

Beran, R., Feuerverger, A. and Hall, P. (1996). On nonparametric estimation of intercept and slope distributions in random coefficient regression, *Ann. Statist.*, **24**, 2569–2592.

Bickel, P. J., Klaasen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation in Semiparametric Models*, Johns Hopkins University Press, Baltimore.

Chang, I.-S. and Hsiung, C. A. (1994). Asymptotic consistency of the maximum likelihood estimate in positron emission tomography and applications, *Ann. Statist.*, **22**, 1871–1883.

Deans, S. R. (1983). *The Radon Transform and Some of Its Applications*, Wiley, New York.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion), *J. Roy. Statist. Soc. Ser. B*, **39**, 1–38.

Eggermont, P. P. B. and LaRiccia, V. N. (1995). Maximum smoothed likelihood density estimation for inverse problems, *Ann. Statist.*, **23**, 199–220.

Fan, J. (1992). Design-adaptive nonparametric regression, *J. Amer. Statist. Assoc.*, **87**, 998–1004.

Gehman, S. and McClure, D. E. (1987). Statistical methods for tomographic image reconstruction, *Bulletin of the International Statistical Institute*, **52**, 5–21.

Green, P. J. (1990*a*). Bayesian reconstructions from emission tomography data using a modified EM algorithm, *IEEE Transactions Medical Imaging*, **9**, 84–93.

Green, P. J. (1990*b*). On use of the EM algorithm for penalized likelihood estimation, *J. Roy. Statist. Soc. Ser. B*, **52**, 443–452.

Higdon, D. M., Bowsher, J. E., Johnson, V. E., Turkington, T. G., Gilland, D. R. and Jaszczak, R. J. (1997). Fully Bayesian estimation of Gibbs Hyperparameters for emission computed tomography data, *IEEE Transactions on Medical Imaging*, **16**(5), 516–526.

Iusem, A. N. and Svaiter, B. F. (1994). A new smoothing-regularization approach for a maximum-likelihood estimation problem, *Appl. Math. Optim.*, **29**, 225–241.

Kaufman, L. (1993). Maximum likelihood, least squares, and penalized least squares for PET, *IEEE Transactions on Medical Imaging*, **12**, 200–214.

Lange, K. and Carson, R. (1984). EM reconstruction algorithms for emission and transmission tomography, *Journal of Computer Assisted Tomography*, **8**, 306–316.

Longford, N. T. (1993). *Random Coefficient Models*, Oxford University Press.

Nicholls, D. F. and Pagan, A. R. (1985). Varying coefficient regression, *Handbook of Statistics*, Vol. 5 (eds. E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 413–449, North-Holland, Amsterdam.

O'Sullivan, F. (1995). A study of least squares and maximum likelihood for image reconstruction in positron emission tomography, *Ann. Statist.*, **23**, 1267–1300.

Raj, B. and Ullah, A. (1981). *Econometrics, A Varying Coeffients Approach*, Croom-Helm, London.

Shepp, L. A. and Kruskal, J. B. (1978). Computerized tomography: The new medical X-ray technology, *Amer. Math. Monthly*, **85**, 420–439.

Shepp, L. A. and Vardi, Y. (1982). Maximum likelihood reconstruction in positron emission tomography, *IEEE Transactions on Medical Imaging*, **1**, 113–122.

Silverman, B. W., Jones, M. C., Wilson, J. D. and Nychka, D. W. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography (with discussion), *J. Roy. Statist. Soc. Ser. B*, **52**, 271–324.

Vardi, Y., Shepp, L. A. and Kaufman, L. (1985). A statistical model for positron emission tomography (with discussion), *J. Amer. Statist. Assoc.*, **80**, 8–37.