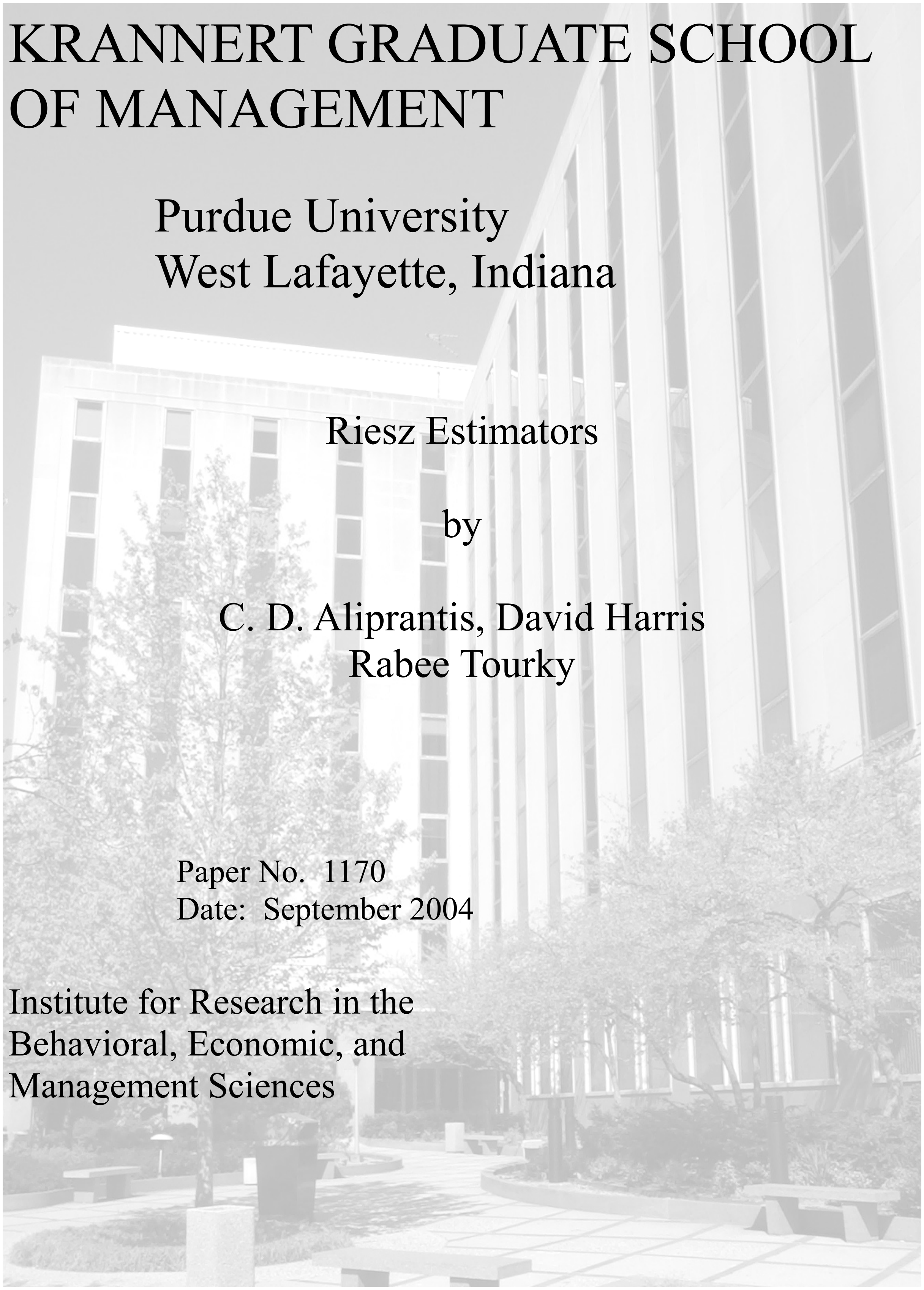


KRANNERT GRADUATE SCHOOL OF MANAGEMENT

Purdue University
West Lafayette, Indiana

Riesz Estimators

by

C. D. Aliprantis, David Harris
Rabee Tourky

Paper No. 1170
Date: September 2004

Institute for Research in the
Behavioral, Economic, and
Management Sciences

RIESZ ESTIMATORS

CHARALAMBOS D. ALIPRANTIS, DAVID HARRIS, AND RABEE TOURKY

ABSTRACT. We consider properties of estimators that can be written as vector lattice (Riesz space) operations. Using techniques widely used in economic theory, we study the approximation properties of these estimators. We also provide two algorithms `RIESZVAR(i-ii)` for the consistent parametric estimation of continuous multivariate piecewise linear functions.

1. INTRODUCTION

Envisage a situation where we seek to estimate a random variable Y based on some observed random vector $\mathbf{X} = (X_1, X_2, \ldots, X_m)$. This paper studies estimators of the conjunctive Boolean form:

$$\widehat{Y} = \bigvee_{j\in J} \bigwedge_{i\in E_j} \left(r_i^0 + r_i^1 X_1 + r_i^2 X_2 + \cdots + r_i^m X_m\right), \qquad (\mathcal{R})$$

where $\{E_j\}_{j\in J}$ is a finite family of finite sets and $\vee$ and $\wedge$ are the vector lattice operations *almost sure supremum* and *almost sure infimum*, respectively. We dub these estimators ***Riesz estimators***.

We are motivated in this paper by the desire to estimate multivariate continuous piecewise linear regressions, by the close relation between Riesz estimators and multivariate threshold models and spline regressions, and because the methods and techniques used to study these estimators have over the past two decades become important tools in economic theory. Moreover, we were drawn to consider Riesz estimators by their close relation to financial derivatives, and our thinking in this paper was initially influenced by the work of Brown, Huijsmans, and de Pagter (1991) on the span of call options with a single underlying security. Recall that if X is the random payoff of a security, then its call option with strike price k is $(X-k)\vee 0$ and its put option is $(k-X)\vee 0$. Therefore, the Riesz estimators in $(\mathcal{R})$ can be thought of as generalized options (recursively, they are options on options—and so they are special cases of derivatives). This identification of Riesz estimators indicates that they could be useful as parametric estimators in some financial models. Further, the relation between Riesz estimators and generalized options also points to the usefulness of the estimators in a non-parametric setting. Indeed, an important question in theoretical finance is the type of options that one needs to add to complete an incomplete market; see Ross (1976) and Brown and Ross (1991). The basic result is that if the available securities are $(X_1, X_2, \ldots, X_m)$ and this set of securities resolves fully revealing information, then adding all the derivatives of the form $(\mathcal{R})$ approximately completes the market. That is, each contingent claim is uniformly approximated by one of these derivatives.

First version: March 14, 2004. This version June 23, 2004.

The research of C. D. Aliprantis is supported by the NSF Grants EIA-0075506 and SES-0128039. The research of R. Tourky is funded by the Australian Research Council Grant A00103450.

The main application of Riesz estimators in this paper is to the parametric estimation of multivariate continuous piecewise linear functions. Many models in economic theory give rise to continuous piecewise linear functions. Examples include better response functions in finite games, reaction functions in Cournot and Bertrand models, kinked demand, and specialized models of multi-market price equilibrium arising from the work of Samuelson (1952). Furthermore, piecewise linear functions arise in a multitude of settings in computational economics, for instance in the computation of competitive equilibrium (Wilson, 1978), the approximation of equilibrium prices (Scarf, 1967), the computation of the equilibria of finite games (Lemke and Howson, 1964), quadratic programming problems in economics, as well as the optimal invariant capital stock problem (Dantzig and Manne, 1974).

We consider the parametric setting in which we know that $\mathbb{E}(Y|\mathbf{X}) = f \circ \mathbf{X}$, where $f\colon \mathbb{R}^m \to \mathbb{R}$ is a continuous piecewise linear function. How can we estimate such a function? In the univariate case, where $m = 1$, this is relatively easy and is a (continuous) threshold regression model where the threshold variable is X. The situation is much more complicated in the multivariate case $m > 1$. A basic problem in the multivariate case is finding an appropriate parameterization of continuous piecewise linear functions. This problem arises in two ways. First, in the multivariate setting the regions of a piecewise linear function can be very complicated. Second, even if we have a parametric representation of these regions, the continuity of the function places additional restrictions on the affine components of the piecewise linear function. Therefore, listing the regions of the function and associating each region with an affine function does not usually result in a continuous piecewise linear function. The continuity of the function provides useful additional structure to piecewise linear models which in turn complicates the parameterization of the regression. We shall see in this paper that this problem is solved by considering the Riesz estimators in ($\mathcal{R}$).

In computation theory the basic method has been to parameterize continuous piecewise linear functions using simplicial methods in algebraic topology. The idea is to define a simplicial subdivision of the domain of the function and associate each vertex of this subdivision with a point in the codomain of the function. Any point in the domain that is not a vertex is a unique convex combination of the vertices and its value is calculated accordingly. This approach was employed for instance by Eaves and Scarf (1976) and Eaves and Lemke (1981, 1983) to study solutions to piecewise linear equations—a study motivated by the economic problems of approximating equilibrium price systems and computing the equilibria of finite games.

This homotopy approach was recently used for estimating continuous piecewise linear functions from data in Groff, Khargonekar, and Koditschek (2003) and the Ph.D. dissertation of Groff (2003). In these works, the authors introduce a novel non-linear parametric algorithm called **Minvar** for estimating piecewise linear functions from data and prove a convergence result for this algorithm. It appears that **Minvar** is the only specialized algorithm available in the literature for the parametric estimation of piecewise linear functions.

The main idea in **Minvar** is to estimate the best fitting simplicial subdivision of the data. For a fixed sample size, **Minvar** is divided into two steps. The first one partitions the data using a simplicial subdivision of the support of the independent variable and obtains a discontinuous estimator of the continuous piecewise linear function f. The second stage of the algorithm adjusts the subdivision defining a continuous piecewise linear function that is close to the discontinuous approximation of the first stage. This is done by minimization of a piecewise quadratic function. The two stages are then iterated until the continuous estimator converges.

In this paper we provide two alternative algorithms based on Riesz estimators, `RIESZVAR(i)` and `RIESZVAR(ii)`, for estimating continuous piecewise linear functions. Following the recent work of Ovchinnikov (2002), we observe that the Riesz estimators in $(\mathcal{R})$ give parsimonious Boolean representations of continuous piecewise linear regressions. The underlying idea in our paper is that if $F = \{f_1, f_2, \ldots, f_p\}$ is a finite set of affine functions, then there is only a finite number of ways that these functions can be "glued" together to make a continuous piecewise linear function. For instance, the three affine functions shown in Figure 1 generate thirteen continuous piecewise linear functions. Now each one of these functions is a "level" in the hyperplane arrangement induced by the graphs of the functions in F and these levels are precisely the vector lattice operations on the components in F.

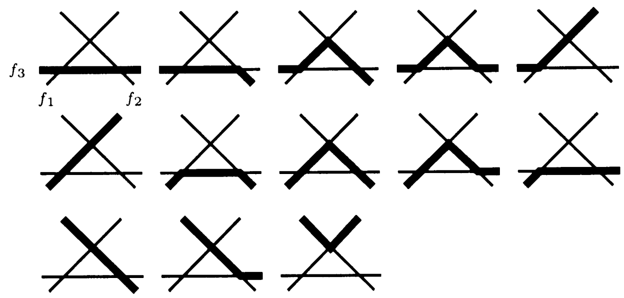


FIGURE 1. The graphs of the functions in f_1, f_2, f_3 and the continuous piecewise linear functions they generate.

Like **Minvar**, the first algorithm, `RIESZVAR(i)`, comprises two steps:

(a) Ordinary least squares on appropriate divisions of the data gives a set of affine functions which contains a subset of functions that estimate the affine components of f.
(b) A finite minimization problem chooses the appropriate Boolean sup-inf representation.

In the second algorithm, `RIESZVAR(ii)`, we construct a piecewise quadratic programming problem whose solution gives both the affine estimates of the components of f as well as the Boolean sup-inf representation. We establish that both algorithms consistently estimate continuous piecewise linear functions, and provide some examples of the implementation of the algorithms.

We can see from the discussion above that the Riesz estimator is a generalized continuous threshold regression model. These generalized thresholds are not taken to be known, which is a generalization from the usual case where the threshold variable is a known element of the set of regressors. Although, for instance, Hansen (1996) discusses the possibility of selecting a threshold variable from a finite set. Threshold autoregressive models are well known in the univariate time series literature having been popularized by Tong (1983) and subsequently widely analyzed. Threshold models also have considerable current interest in econometrics, often where there is a single threshold. For example Hansen and Seo (2002) discuss testing for threshold cointegration and provide a range of references to recent applications, and Hansen (1999) considers thresholds in static panel data models. A continuous version of the multiple threshold model considered by Gonzalo and Pitarakis (2002) is also included in $(\mathcal{R})$ provided the threshold variable is an element of the set of regressors.

In this paper, we also study the space of Riesz estimators and their relation to other estimators in the literature. We observe that the space of Riesz estimators is a linear space and is precisely the lattice hull of the space of affine estimators. In particular, Riesz estimators give rise to series estimators of the form:

$$\widehat{Y} = \alpha + \sum_{i=1}^{m} \beta_i X_i + \sum_{j=1}^{p} \gamma_j \Big[\bigvee_{k=1}^{u_j} \bigwedge_{\ell=1}^{v_j} \Big(r^0_{jk\ell} + r^1_{jk\ell} X_1 + r^2_{jk\ell} X_2 + \cdots + r^m_{jk\ell} X_m \Big) \Big] , \qquad (\mathcal{SR})$$

where α, β_i, and γ_j are the parameters we seek to estimate and $r^i_{jk\ell}$ are constants. It is shown that the Riesz series estimators approximate all estimators based on $\mathbf{X}$. The series estimator ($\mathcal{SR}$) is closely related to the usual multivariate regression models using splines (Friedman, 1991). Basically, a spline is a piecewise polynomial continuous function. The splines defined by the Riesz estimators have affine components. An important analytical question in the spline literature is the characterization of the space of possibly non-affine splines (Schumaker, 1984). The characterization of the space of Riesz estimators as the lattice hull of affine estimators shows that there is a dramatic change when one moves from the affine splines to the non-affine splines. One surprise is that non-affine splines cannot be represented as lattice operations of their components, even in the case of univariate quadratic splines.

It turns out that Riesz estimators have a familiar form in generalized additive models. In these models there exists a known "link function" $g\colon \mathbb{R} \to \mathbb{R}$ so that the conditional expectation has the form

$$g \circ \mathbb{E}(Y|\mathbf{X}) = \alpha + g_1 \circ X_1 + g_2 \circ X_2 + \cdots + g_m \circ X_m ,$$

where each $g_i\colon \mathbb{R} \to \mathbb{R}$ is an unknown non-parametric function. Such models include the general estimation problem when there is a single observed random variable X, the additive regression where g is the identity, and multiplicative regressions when g is the logarithm. We show that in additive models (where g is known) the relevant Riesz estimators are of the piecewise linear form:

$$\widehat{Z} = \alpha + \sum_{i=1}^{m} \beta_i X_i + \sum_{i=1}^{m} \sum_{j=1}^{p} \gamma_{ij} (X_i - r_{ij})^+ , \qquad (\mathcal{AR})$$

where $\widehat{Z}$ is an estimator of $g \circ \mathbb{E}(Y|\mathbf{X})$.

We close the paper by showing that each constant r_{ij} in ($\mathcal{AR}$) is a possible structural break or threshold in the otherwise linear relationship between the observed variable X_i and the estimator $\widehat{Z}$. Importantly, as we progressively add all the constants r_{ij} (and $p \to \infty$) these additive Riesz estimators approximate all additive estimators based on $\mathbf{X}$. There is nascent economic intuition for this. As mentioned earlier, the terms $(X_i - r_{ij})^+$ can be thought of as call options in which the underlying security pays the contingent claim X_i and the strike price is r_{ij}. So the idea is that by adding these put options $(X_i - r_{ij})^+$ gives us all the needed information in additive models. Moreover, in the statistics literature the form ($\mathcal{AR}$) with $m = 1$ is used as a non-parametric estimator using linear regression splines. Such estimators are the subject of a large literature, see Agarwal and Studden (1980) and Zhou, Shen, and Wolfe (1998) for some asymptotic properties covering linear regression splines.

Much of the analysis in this paper applies from well known, as well as recent, results in the theory of Riesz spaces (or vector lattices) and their applications to economic theory. Riesz spaces are ordered vector spaces whose vector orderings are lattice orderings. They are abstractions of the order properties of spaces such as $C[0,1]$ (continuous functions on $[0,1]$), the classical Lebesgue L_p-spaces, and the space ca$[0,1]$ of all σ-additive signed measures on $[0,1]$ of finite total variation. The theory of Riesz spaces has become central to modern general equilibrium theory where the lattice structures on the space of price systems are used to solve, for instance, separating hyperplane problems arising from the study of the fundamental welfare theorems.[1] Riesz spaces have also been used in theoretical finance and have been employed to solve some special forms of the portfolio insurance problem; see for example Aliprantis, Brown, Polyrakis, and Werner (1998), Aliprantis, Brown, and Werner (2000), Aliprantis, Polyrakis, and Tourky (2002b). The theory of vector lattices also underlines much of the work on linear optimization and more generally the Linear Complementarity Problem with its applications to computational economics and game theory; see for instance Cottle, Pang, and Stone (1992). Of course, there is a very long tradition of using vector lattice theory in probability and statistics. The best known example is the beautiful work of Le Cam (1986); see also Bomze (1990).

2. Riesz spaces, Banach lattices, and estimators

The objective of this section is to present a brief discussion of the mathematical background needed to study Riesz estimators. The mathematics behind the theory of Riesz estimators are those of Riesz spaces and Banach lattices. We recall here some basic properties of Riesz spaces and for details and terminology we refer to the monographs Abramovich and Aliprantis (2002a), Aliprantis and Border (1999), Aliprantis and Burkinshaw (2003), Schaefer (1974), Luxemburg and Zaanen (1971).

An ***ordered vector space*** is a vector space L equipped with an order relation $\geq$ that is compatible with the algebraic structure of L in the sense that if $x \geq y$, then:

(a) $x + z \geq y + z$ for each $z \in L$, and
(b) $\alpha x \geq \alpha y$ for all $\alpha \geq 0$.

An ordered vector space L is said to be a **Riesz space** (or a **vector lattice**) if L is also a lattice in the sense that every nonempty finite subset of L has a supremum (least upper bound) and an infimum (greatest lower bound). Following the standard terminology from lattice theory, we shall denote the supremum and infimum of a set $\{x_1, \ldots, x_n\}$ by

$$\bigvee_{i=1}^{n} x_i \quad \text{and} \quad \bigwedge_{i=1}^{n} x_i\,,$$

respectively. In particular, the supremum and infimum of any pair of vectors x and y are denoted by $x \vee y$ and $x \wedge y$, respectively. The simplest example of a Riesz space is $\mathbb{R}$ with the usual order. Here $x \vee y$ and $x \wedge y$ are the largest and smallest numbers of the set $\{x, y\}$; for instance, $2 \vee 3 = 3$, $1 \wedge 0 = 0$, and $3 \wedge 3 = 3$.

For an element x of a Riesz space L the **positive part** of x is defined by $x^+ = x \vee 0$, the **negative part** by $x^- = (-x) \vee 0$, and the **absolute value** by $|x| = x \vee (-x)$.

The following is a simple but very useful result.

Lemma 2.1. *An ordered vector space is a Riesz space if and only if x^+ exists for each vector x.*

For an illustration of the above notions let $L = C[0,1]$, the vector space of all continuous real valued functions defined on $[0,1]$. With the pointwise ordering and algebraic operations $C[0,1]$ is a Riesz space such that for each $x \in L$ and each $t \in [0,1]$ we have

$$x^+(t) = \max\{x(t), 0\}\,, \quad x^-(t) = \max\{-x(t), 0\}\,, \quad \text{and} \quad |x|(t) = \max\{x(t), -x(t)\} = |x(t)|\,.$$

[1] For a survey of the recent literature see Aliprantis, Cornet, and Tourky (2002a). See also the monograph Aliprantis, Brown, and Burkinshaw (1990) and Mas-Colell and Zame (1991).

Similarly, if $x \in L$ and $r \in \mathbb{R}$, then for each $t \in [0,1]$ we have

$$[(x-r)^+](t) = \begin{cases} x(t) - r & \text{if } x(t) \geq r, \\ 0 & \text{if } x(t) < r \end{cases} \quad \text{and} \quad [(x-r)^-](t) = \begin{cases} r - x(t) & \text{if } x(t) \leq r, \\ 0 & \text{if } x(t) > r. \end{cases}$$

Also, notice that if $\{x_1, \ldots, x_n\} \in C[0,1]$, then for each $t \in [0,1]$ we have

$$\Big[\bigvee_{i=1}^{n} x_i\Big](t) = \max\{x_1(t), \ldots, x_n(t)\} \quad \text{and} \quad \Big[\bigwedge_{i=1}^{n} x_i\Big](t) = \min\{x_1(t), \ldots, x_n(t)\}.$$

Since $C(\mathbb{R}^n)$ with the pointwise ordering is a Riesz space, the above formulas are also true for functions of $C(\mathbb{R}^n)$.

Our interest here is in the structure of the Riesz subspaces of a Riesz space. A vector subspace M of a Riesz space L is said to be a **Riesz subspace** (or a **vector sublattice**) if $x, y \in M$ imply that both $x \vee y$ and $x \wedge y$ belong to M. If we consider the product vector space $\mathbb{R}^\Omega$ (where Ω is any nonempty set) and order it pointwise, then (with the above lattice operations) $\mathbb{R}^\Omega$ is a Riesz space. Moreover, if Ω is a topological space, then $C(\Omega)$ (the vector space of all continuous real-valued functions on Ω) and $C_b(\Omega)$ (the vector space of all uniformly bounded continuous real-valued functions on Ω) are both Riesz subspaces of $\mathbb{R}^\Omega$.

It should be clear that arbitrary intersections of Riesz subspaces are Riesz subspaces. This implies that every nonempty subset A of a Riesz space L is included in a smallest Riesz subspace, called the **Riesz subspace** (or the **vector sublattice**) **generated** by A and denoted $\mathcal{R}(A)$.

Next, we shall briefly describe the Riesz subspace $\mathcal{R}(A)$, an important subspace for our work. For every nonempty subset A of a Riesz space L, the symbol $A^\wedge$ will denote the collection of all vectors that can be written as infima of finite subsets of A. That is, a vector $a \in L$ belongs to $A^\wedge$ if there exist vectors $a_1, a_2, \ldots, a_k \in A$ such that $a = \bigwedge_{i=1}^k a_i$. Similarly, $A^\vee$ is the set consisting of all suprema of finite subsets of A. We write $A^{\vee\wedge}$ for $(A^\vee)^\wedge$ and $A^{\wedge\vee}$ for $(A^\wedge)^\vee$. So, a vector a belongs to $A^{\vee\wedge}$ if and only if there exists a finite family $\{E_j\}_{j\in J}$ of non-empty finite subsets of L such that $a = \bigvee_{j\in J} \bigwedge E_j$. It turns out that $A^{\vee\wedge} = A^{\wedge\vee}$ is always true.

Now we can describe the Riesz subspace generated by a set as follows. For proofs and more discussion see Sections 5 of Abramovich and Aliprantis (2002a,b).

Lemma 2.2. *The Riesz subspace $\mathcal{R}(A)$ generated by a vector subspace A of a Riesz space coincides with $A^{\wedge\vee}$ and also with $A^{\vee\wedge}$. That is, $\mathcal{R}(A) = A^{\wedge\vee} = A^{\vee\wedge}$.*

Corollary 2.3. *The Riesz subspace generated by a nonempty subset A of a vector lattice is precisely the vector space $\mathcal{R}(A) = [A]^{\wedge\vee}$, where $[A]$ denotes the linear span of A.*

When a Riesz space L is equipped with a norm that is compatible with the order structure of the space in the sense that $|x| \leq |y|$ implies $\|x\| \leq \|y\|$, then L is called a normed Riesz space.[2] A **Banach lattice** is a Riesz space that is a Banach lattice under a lattice norm. It is not difficult to see that in a Banach lattice the closure of a Riesz subspace is likewise a Riesz subspace.

The two classical examples of Banach lattices are the $C(\mathcal{X})$-spaces, where $\mathcal{X}$ is a compact topological space and the norm is the sup norm $\|\cdot\|_\infty$, i.e.,

$$\|f\|_\infty = \sup_{x\in\mathcal{X}} |f(x)|,$$

[2]Any norm on a Riesz space such that $|x| \leq |y|$ implies $\|x\| \leq \|y\|$ is called a **lattice** (or a **Riesz**) **norm**.

and the $L_p(\mu)$-spaces, where $1 \le p \le \infty$, and the norm is given by

$$\|f\|_p = \left[\int |f|^p \, d\mu\right]^{\frac{1}{p}}, \ \text{if } 1 \le p < \infty \text{ and } \|f\|_\infty = \text{ess sup} f\,, \ \text{if } p = \infty\,.$$

These are the two Banach lattices that will appear repeatedly in this work.

Finally, we shall close the section with some terminology regarding the Hilbert space L_2. In this paper, $(\Omega, \mathcal{F}, \pi)$ will denote a fixed probability space. The functions (i.e., the equivalence classes) in the Hilbert space $L_2(\pi) \equiv L_2(\Omega, \mathcal{F}, \pi)$ will be referred to as **random variables**. The constant random variable one on Ω will be denoted by $\mathbf{1}$, i.e., $\mathbf{1}(\omega) = 1$ for each $\omega \in \Omega$. The norm of $L_2(\pi)$ will be denoted by $\|\cdot\|$, that is, $\|f\| = \left(\int_\Omega f^2 \, d\pi\right)^{\frac{1}{2}}$. For any sub-$\sigma$-algebra $\mathcal{A}$ of $\mathcal{F}$ we shall denote by $L_2(\mathcal{A})$ the closed vector subspace of $L_2(\pi)$ consisting of all $\mathcal{A}$-measurable square integrable (equivalence classes) random variables.

The following basic result characterizes the closed vector sublattices that contain the constant function $\mathbf{1}$. For the proof of the next result see for example Aliprantis and Border (1999, Thm 12.11, p. 433)

Lemma 2.4. *Every closed Riesz subspace of $L_2(\pi)$ containing the constant function $\mathbf{1}$ is of the form $L_2(\mathcal{A})$, where $\mathcal{A}$ is a sub-σ-algebra of $\mathcal{F}$.*

We shall use the boldface notation $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ to designate **random vectors** in $L_2(\pi)^m$. For any random vector $\mathbf{X}$, we shall denote by $\sigma(\mathbf{X})$ the sub-σ-algebra generated by $\mathbf{X}$, i.e., $\sigma(\mathbf{X})$ is the smallest (with respect to inclusion) sub-σ-algebra of $\mathcal{F}$ for which each X_i is measurable.

Definition 2.5. *A random variable $\widehat{Y}$ is said to be an* **estimator** *based on a random vector $\mathbf{X}$, if we can write $\widehat{Y}$ as a function of realizations of $\mathbf{X}$ in the sense that there exists a Borel measurable function $g\colon \mathbb{R}^m \to \mathbb{R}$ satisfying $\widehat{Y} = g \circ \mathbf{X}$. The function g is called the* **estimating** *function of the random vector $\mathbf{X}$.*

Clearly, the collection of all estimators based on a random vector $\mathbf{X}$ is a closed vector subspace of $L_2(\pi)$. Moreover, it is a Riesz subspace that contains $\mathbf{1}$. The following result is well known and is a simple consequence of Lemma 2.4. For a complete proof see for instance Aliprantis and Border (1999, Thm 4.40, p. 145).

Lemma 2.6. *The vector subspace of all estimators based on a random vector $\mathbf{X}$ coincides with the Riesz subspace $L_2(\sigma(\mathbf{X}))$ of $L_2(\pi)$.*

To continue our discussion, we need the following basic result in Hilbert spaces that guarantees the existence of a closest point from a given point to any non-empty closed convex subset. For completeness, we present a short proof of this result.

Lemma 2.7. *Let C be a non-empty convex closed subset of a Hilbert space. If $x \notin C$, then there exists a unique $c_0 \in C$ such that $\|c_0 - x\| = \inf_{c \in C} \|c - x\|$.*

Moreover, if a sequence $\{c_n\}$ of C satisfies $\|c_n - x\| \to \inf_{c \in C} \|c - x\|$, then $c_n \to c_0$.

Proof. By translating, we can assume that $x = 0 \notin C$. Let $d = \inf_{c \in C} \|c\|$ and then pick a sequence $\{c_n\} \subseteq C$ such that $\|c_n\| \to d$. From the Parallelogram Law, we see that

$$\|c_n - c_m\|^2 = 2\|c_n\|^2 + 2\|c_m\|^2 - 4\left\|\tfrac{x_n + x_m}{2}\right\|^2 \le 2\|c_n\|^2 + 2\|c_m\|^2 - 4d^2 \xrightarrow[n,m\to\infty]{} 2d^2 + 2d^2 - 4d^2 = 0\,.$$

This shows that $\{c_n\}$ is a norm Cauchy sequence. If $c_n \to c_0$, then $d = \|c_0\|$.

Now assume that another sequence $\{c_n^*\} \subseteq C$ satisfies $\|c_n^*\| \to d$. Define the sequence $\{c_n^{**}\}$ of C by letting $c_{2n}^{**} = c_n$ and $c_{2n-1}^{**} = c_n^*$. Clearly, $\|c_n^{**}\| \to d$. By the preceding case, there exists some $c^* \in C$ satisfying $c_n^{**} \to c^*$ and $\|c^*\| = d$. Now note that $c^* = \lim_{n\to\infty} c_{2n}^{**} = \lim_{n\to\infty} c_n = c_0$. This implies $\lim_{n\to\infty} c_n^* = \lim_{n\to\infty} c_{2n-1}^{**} = c_0$, and the proof is finished. ■

If Y is a random variable and V is a closed vector subspace of $L_2(\pi)$, then the ***best*** (or ***minimum***) ***variance estimator of Y based on*** V is the unique (according to Lemma 2.7) random variable $\widehat{Y} \in V$ that satisfies $\|\widehat{Y} - Y\| \le \|\widehat{Z} - Y\|$ for all $\widehat{Z} \in V$, that is, $\widehat{Y}$ is the unique solution of the following minimization problem:

$$\begin{aligned} &\operatorname{Min} \|\widehat{Z} - Y\| \\ &\text{s.t.}: \widehat{Z} \in V \end{aligned}$$

If Y is a random variable, then the ***best variance estimator of Y based on a random vector*** $\mathbf{X}$ is simply the best variance estimator of Y based on $L_2(\sigma(\mathbf{X}))$.

Regarding best variance estimators, the following important result is well-known—we sketch a proof below based on the Riesz space theory of positive operators.

Lemma 2.8. *If Y is a random variable, then the best variance estimator $\widehat{Y}$ of Y based on a random vector $\mathbf{X}$ is $\mathbb{E}(Y|\mathbf{X})$, i.e., $\widehat{Y} = \mathbb{E}(Y|\mathbf{X})$, the conditional expectation of Y based on $\mathbf{X}$.*

Proof. The conditional expectation operator $Y \mapsto \mathbb{E}(Y|\mathbf{X})$ is a contractive projection on $L_1(\pi)$. Moreover, $Y \mapsto \mathbb{E}(Y|\mathbf{X})$ leaves invariant each $L_p(\pi)$ $(1 \le p \le \infty)$, is a contractive projection on each $L_p(\pi)$ and has range $L_p(\sigma(\mathbf{X}))$; see (Abramovich and Aliprantis, 2002a, Thms 5.37 and 5.38). To complete the proof notice that on the Hilbert space $L_2(\pi)$ the contractive projections are precisely the orthogonal projections. (See for instance Abramovich and Aliprantis (2002b, Problem 5.3.14).) ■

The next lemma is also very useful. Its proof follows immediately from Lemma 2.7.

Lemma 2.9. *Assume that $\mathbf{X}$ is a random vector and Y is a random variable. If a sequence $\{Z_n\}$ of random variables in $L_2(\sigma(\mathbf{X}))$ satisfies $\|Y - Z_n\| \to \|Y - \mathbb{E}(Y|\mathbf{X})\|$, then $\|Z_n - \mathbb{E}(Y|\mathbf{X})\| \to 0$.*

3. One-dimensional piecewise linear functions

We present here a few properties and formulas dealing with piecewise linear functions defined on $\mathbb{R}$ or on a closed interval of $\mathbb{R}$.

Definition 3.1. *A function $f\colon \mathbb{R} \to \mathbb{R}$ is called **piecewise linear** (**affine**) if there exist real numbers $-\infty < a_0 < a_1 < \cdots < a_k < \infty$ and pairs of real numbers (m_i, b_i), $i = 0, 1, \ldots, k, k+1$, such that*

$$f(t) = \begin{cases} m_i t + b_i & \text{if } a_{i-1} \le t \le a_i \text{ for some } 1 \le i \le k, \\ m_0 t + b_0 & \text{if } t \le a_0, \\ m_{k+1} t + b_{k+1} & \text{if } t \ge a_{k+1}\,. \end{cases}$$

The parameters $\{a_0, a_1, \ldots, a_k\}$ *and the pairs* (m_i, b_i), $i = 0, 1, \ldots, k, k+1$, *are referred to as a* ***representation*** *of* f *and the functions* $f_i(t) = m_i t + b_i$ *as the* ***components*** *of the representation.*

Similarly, a function $f\colon [a,b] \to \mathbb{R}$, *where* $[a,b]$ *is a closed interval of* $\mathbb{R}$, *is* ***piecewise linear*** *if there exist a partition* $a = a_0 < a_1 < \cdots < a_k = b$ *of the interval* $[a,b]$ *and pairs of real numbers* (m_i, b_i), $i = 1, \ldots, k$, *such that* $f(t) = m_i t + b_i$ *for all* $a_{i-1} \le t \le a_i$.

Notice that, according to these definitions, piecewise linear functions are automatically continuous. The following result should be obvious.

Lemma 3.2. *If* $f\colon \mathbb{R} \to \mathbb{R}$ *is piecewise linear, then its restriction to any closed interval of* $\mathbb{R}$ *is likewise piecewise linear. Moreover, if* $[a,b]$ *is any closed subinterval of* $\mathbb{R}$, *then the components of the piecewise linear function* $f\colon [a,b] \to \mathbb{R}$ *are among the components of* $f\colon \mathbb{R} \to \mathbb{R}$.

In addition, every piecewise linear function on a closed interval of $\mathbb{R}$ *can be extended to a piecewise linear function to all of* $\mathbb{R}$.

The piecewise linear functions on a closed interval are characterized as follows. The idea is depicted in Figure 2.

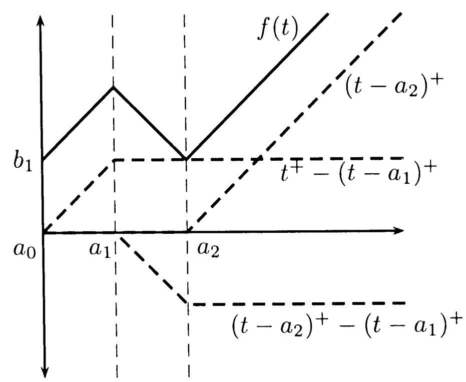


FIGURE 2. Notice that $f(t) = b_1 + t^+ - 2(t-a_1)^+ + 2(t-a_2)^+$.

Lemma 3.3. *Let* $f\colon [a,b] \to \mathbb{R}$ *be a piecewise linear function. If* $\{a_0, a_1, \ldots, a_k\}$ *and* (m_i, b_i), $i = 1, \ldots, k$, *is any representation of* f, *then for each* $t \in \mathbb{R}$ *we have*

$$f(t) = b_1 + m_1 t + \sum_{i=1}^{k-1} (m_{i+1} - m_i)(t - a_i)^+ .$$

In particular, a function $f\colon [a,b] \to \mathbb{R}$ *is piecewise linear if and only if there exist a partition* $a = a_0 < a_1 < \cdots < a_k = b$ *of* $[a,b]$ *and constants* $c, c_0, c_1, \ldots, c_k$ *such that for each* $t \in [a,b]$ *we have* $f(t) = c + \sum_{i=0}^{k} c_i (t - a_i)^+$.

Proof. Let $a \le t \le b$. If $a_0 \le t \le a_1$, then note that

$$b_1 + m_1 t + \sum_{i=1}^{k-1} (m_{i+1} - m_i)(t - a_i)^+ = b_1 + m_1 t = f(t)\,.$$

So, we can assume that $a_{j-1} \le t \le a_j$ for some $1 < j \le k$. Notice that for each $1 < i \le k-1$ we have $m_i a_i + b_i = m_{i+1} a_i + b_{i+1}$ or $(m_{i+1} - m_i)a_i = -(b_{i+1} - b_i)$. Consequently, we have

$$\begin{aligned} b_1 + m_1 t + \sum_{i=1}^{k-1} (m_{i+1} - m_i)(t - a_i)^+ &= b_1 + m_1 t + \sum_{i=1}^{j-1} (m_{i+1} - m_i)(t - a_i)^+ \\ &= b_1 + m_1 t + \sum_{i=1}^{j-1} (m_{i+1} - m_i)(t - a_i) \\ &= b_1 + m_1 t + \Big[\sum_{i=1}^{j-1} (m_{i+1} - m_i)\Big] t - \sum_{i=1}^{j-1} (m_{i+1} - m_i) a_i \\ &= b_1 + m_1 t + \Big[\sum_{i=1}^{j-1} (m_{i+1} - m_i)\Big] t + \sum_{i=1}^{j-1} (b_{i+1} - b_i) \\ &= b_1 + m_1 t + (m_j - m_1)t + (b_j - b_1) = m_j t + b_j = f(t)\,, \end{aligned}$$

and the proof is finished. ■

Corollary 3.4 (Brown, Huijsmans, and de Pagter (1991)). *The vector subspace generated in $C[0,1]$ by the collection of continuous functions $\{\mathbf{1}, t\} \cup \{(\alpha - t)^+ :\ \alpha \in \mathbb{R}\}$ coincides with the Riesz subspace of all piecewise linear functions on $[0,1]$.*

Corollary 3.5. *Let $g\colon \mathbb{R} \to \mathbb{R}$ be a piecewise linear function. If $\{a_0, a_1, \ldots, a_k\}$ and (m_i, b_i), $i = 0, 1, \ldots, k, k+1$, is an arbitrary representation of g, then for each $t \in \mathbb{R}$ we have*

$$g(t) = b_0 + m_0 t + \sum_{i=0}^{k} (m_{i+1} - m_i)(t - a_i)^+\,.$$

In particular, a function $f\colon \mathbb{R} \to \mathbb{R}$ is piecewise linear if and only if there exist real constants $m_0, b_0, a_0, a_1, \ldots, a_k$ and $c_0, c_1, \ldots, c_k$ such that for each $t \in \mathbb{R}$ we have

$$f(t) = b_0 + m_0 t + \sum_{i=0}^{k} c_i (t - a_i)^+\,.$$

Proof. Consider the function $h\colon \mathbb{R} \to \mathbb{R}$ defined by $h(t) = b_1 + m_1 t + \sum_{i=1}^{k-1} (m_{i+1} - m_i)(t - a_i)^+$. As in the proof of Lemma 3.3, it is easy to see that $h(t) = f(t)$ for all $a_0 \le t \le a_k$. Moreover, $h(t) = m_1 t + b_1$ for all $t \le a_0$ and $h(t) = m_k t + b_k$ for all $t \ge a_k$. Since

$$\begin{aligned} m_1 t + b_1 + m_1(a_0 - t)^+ - m_0(a_0 - t)^+ &= b_0 + m_0 t \ \text{ for all } t \le a_0 \\ m_1 t + b_1 + m_1(a_0 - t)^+ - m_0(a_0 - t)^+ &= m_1 t + b_1 \ \text{ for all } t \ge a_0 \\ m_k t + b_k + (m_{k+1} - m_k)(t - a_k)^+ &= m_{k+1} + b_{k+1} \ \text{ for all } t \ge a_k\,, \ \text{ and} \\ m_k t + b_k + (m_{k+1} - m_k)(t - a_k)^+ &= m_k t + b_k \ \text{ for all } t \le a_k\,, \end{aligned}$$

it follows that

$$\begin{aligned}
g(t) &= m_1(a_0-t)^+ - m_0(a_0-t)^+h(t) + (m_{k+1}-m_k)(t-a_k)^+ \\
&= m_1(a_0-t)^+ - m_0(a_0-t)^+ + b_1 + m_1t + \sum_{i=1}^{k-1}(m_{i+1}-m_i)(t-a_i)^+ + (m_{k+1}-m_k)(t-a_k)^+ \\
&= m_1(a_0-t)^+ - m_0(a_0-t)^+ + b_1 + m_1t + \sum_{i=1}^{k}(m_{i+1}-m_i)(t-a_i)^+ \\
&= b_0 + m_0t + (m_1-m_0)(t-a_0)^+ + \sum_{i=1}^{k}(m_{i+1}-m_i)(t-a_i)^+ \\
&= b_0 + m_0t + \sum_{i=0}^{k}(m_{i+1}-m_i)(t-a_i)^+ ,
\end{aligned}$$

as desired. ■

We close the section with two results that will be useful for our study later.

Lemma 3.6. *Let $f\colon [a,b] \to \mathbb{R}$ be a piecewise linear function and let $\{a_0, a_1, \ldots, a_k\}$ and (m_i, b_i), $i = 1, \ldots, k$, be a representation of f. Also let $m = \frac{f(b)-f(a)}{b-a}$, the slope of the line segment joining the points $(a, f(a))$ and $(b, f(b))$.*

Then there exist $1 \le i \le k$ with $m_i \ge m$ and $a_{i-1} \le \xi \le a_i$ satisfying $f(\xi) = m(\xi - a) + f(a)$.

Proof. Assume by way of contradiction that if $m_i \ge m$, then $f(t) \ne m(t-a) + f(a)$ for all $a_{i-1} \le t \le a_i$. In particular, we have $m_1 < m$. Given that for $a \le t \le a_1$ we have $f(t) = m_1t + b_1 = m_1(t-a) + f(a)$, the latter implies $f(t) < m(t-a) + f(a)$ for all $a < t \le a_1$. Notice that for each $a_1 \le t \le a_2$ we have $f(t) = m_2t + b_2 = m_2(t-a_1) + f(a_1)$. So, if $m_2 < m$, then for each $a_1 \le t \le a_2$ we have $f(t) < m(t-a) + f(a)$. On the other hand, if $m_2 \ge m$, then for each $a_1 \le t \le a_2$ we must have $f(t) < m(t-a) + f(a)$; otherwise (by the intermediate value theorem) there should exist some $a_1 \le \xi \le a_2$ with $f(\xi) = m(t-a) + f(a)$, which contradicts our assumption. The same argument yields $f(t) < m(t-a) + f(a)$ for all $a_2 \le t \le a_3$. Continuing this way we see that $f(a_k) = f(b) < m(b-a) + f(a) = f(b)$, which is impossible. ■

As an immediate consequence we get the following result.

Corollary 3.7 (Ovchinnikov (2002))**.** *Let $f\colon [a,b] \to \mathbb{R}$ be a piecewise linear function and let $\{a_0, a_1, \ldots, a_k\}$ and (m_i, b_i), $i = 1, \ldots, k$, be the parameters of a representation of f. Then there exists some $1 \le i \le k$ such that $f(a) \ge m_ia + b_i$ and $f(b) \le m_ib + b_i$.*

Proof. According to Lemma 3.6 there exist some $1 \le i \le k$ and some $a_{i-1} \le \xi \le a_i$ satisfying $m_i \ge m = \frac{f(b)-f(a)}{b-a}$ and $f(\xi) = m(t-a) + f(a)$. Note that for each $a_{i-1} \le t \le a_i$ we have $m_it + b_i = m_i(t-\xi) + f(\xi)$ and that for all $a \le t \le b$ we have $m(t-a) + f(a) = m(t-\xi) + f(\xi)$. This implies $m_it + b_i \le m(t-a) + f(a)$ for all $a \le t \le \xi$ and $m_it + b_i \ge m(t-a) + f(a)$ for all $\xi \le t \le b$, and our conclusion follows. ■

4. Multivariate piecewise linear functions

Recall that any function $f\colon \mathbb{R}^m \to \mathbb{R}$ of the form $f(x) = \alpha + a \cdot x$, where $\alpha \in \mathbb{R}$ is a constant and $a \in \mathbb{R}^m$ is a fixed vector, is called an **affine function**. As usual, an affine function f is **linear** if $\alpha = 0$, i.e., $f(x) = a \cdot x$. A function $f\colon S \to \mathbb{R}$, where S is a subset of $\mathbb{R}^m$, is said to be an **affine function** if it is the restriction of an affine function defined on $\mathbb{R}^m$. Let **Aff** denote the collection of all affine functions on $\mathbb{R}^m$ and note that **Aff** is a vector subspace of $C(\mathbb{R}^m)$.

Lemma 4.1. *Regarding affine functions we have the following:*

1. *The vector space* **Aff** *of all affine functions is the linear span in* $C(\mathbb{R}^m)$ *of the functions* $\{\mathbf{1}, e_1, e_2, \ldots, e_m\}$, *where* $\mathbf{1}(x) = 1$ *and* $e_i(x) = x_i$ *for all* $x \in \mathbb{R}^m$. *That is, we have* $\mathbf{Aff} = \mathrm{Span}\,\{\mathbf{1}, e_1, e_2, \ldots, e_m\}$*; and so* **Aff** *is an* $(m+1)$*-dimensional vector space.*[3]
2. *Two affine functions* $f, g \in \mathbf{Aff}$ *coincide if and only if* $f(x) = g(x)$ *for all* x *in a non-empty open subset of* $\mathbb{R}^m$. *In particular, if a subset* S *of* $\mathbb{R}^m$ *has an interior point, then any affine function on* S *is the restriction of a unique affine function defined on* $\mathbb{R}^m$.

Proof. The proof of part (1) is obvious. The proof of part (2) follows easily from the following simple property: *If a non-zero linear functional* f *satisfies* $f(x) \ge \alpha$ *for all* x *in a non-empty open set* $\mathcal{O}$, *then* $f(x) > \alpha$ *must be the case for all* $x \in \mathcal{O}$.

To see this, fix $x \in \mathcal{O}$ and assume that $f(x) = \alpha$. Since $\mathcal{O}$ is an open set, there exists some $\epsilon > 0$ such that $x + B(0,\epsilon) \subseteq \mathcal{O}$. So, for each $y \in B(0,\epsilon)$ we have $\alpha + f(y) = f(x+y) \ge \alpha$ or $f(y) \ge 0$. This implies $f(y) = 0$ for all $y \in B(0,\epsilon)$ and so $f = 0$, which is impossible. ■

We are now ready to introduce the concept of a piecewise linear function.

Definition 4.2. *A function* $f\colon \mathbb{R}^m \to \mathbb{R}$ *is called* ***piecewise linear*** *(or* ***piecewise affine****) if there exist distinct affine functions* $f_1, f_2, \ldots, f_p$ *and subsets* $S_1, S_2, \ldots, S_p$ *of* $\mathbb{R}^m$ *such that:*

1. *Each* S_i *is closed with non-empty interior and* $\overline{\mathrm{Int}(S_i)} = S_i$.[4]
2. *If* $i \neq j$, *then* $\mathrm{Int}(S_i) \cap \mathrm{Int}(S_j) = \emptyset$.
3. $\bigcup_{i=1}^{p} S_i = \mathbb{R}^m$.
4. *If* $x \in S_i$, *then* $f(x) = f_i(x)$.

We also introduce the following terminology and notation.

(a) *The sets* S_i *are called the* ***regions*** *of* f *and the functions* f_i *will be referred to as the* ***components*** *of* f.
(b) *The pairs* $(S_1, f_1), \ldots, (S_p, f_p)$ *are the* ***characteristic pairs*** *of* f.
(c) *The set of all piecewise linear functions will be denoted by* **PL**.

A remark is in order here. The same definition of a piecewise linear function can be given for **solid** domains, i.e., for closed convex subsets of $\mathbb{R}^m$ with non-empty interior. All results in this section hold true for piecewise linear functions with solid domains. We assume that our functions have domain $\mathbb{R}^m$ for the sole purpose of simplifying the exposition. The reader can verify directly

[3] As a matter of fact, if we identify every vector $\mathbf{r} = (r_0, r_1, \ldots, r_m) \in \mathbb{R}^{m+1}$ with the affine function on $\mathbb{R}^m$ defined by $\mathbf{r}(x) = r_0 + r_1x_1 + \cdots + r_mx_m$, then it is not difficult to see that we can identify **Aff** with the vector space $\mathbb{R}^{m+1}$.

[4] If A is any subset of $\mathbb{R}^m$, then $\mathrm{Int}(A)$ denotes its interior and $\overline{A}$ its closure. We remark that the sets S_i are not assumed to be connected.

that when $m = 1$ the definitions for piecewise linear functions given in Definitions 3.1 and 4.2 are equivalent; see also Corollary 4.10 below.

Here is an example of an piecewise linear function with a solid domain in $\mathbb{R}^2$.

Example 4.3. Let $Q = [0,12] \times [0,12] = \{(x,y) \in \mathbb{R}^2 \colon\ 0 \le x \le 12 \text{ and } 0 \le y \le 12\}$. Consider the piecewise linear function $f \colon Q \to \mathbb{R}$ defined by

$$f(x_1,x_2) = \begin{cases} x_1 - 5 & \text{if } x_2 \ge x_1 \ \&\ 2x_1 \ge 17 - x_2\,, \\ x_2 - 5 & \text{if } x_2 \le x_1 \ \&\ x_1 \ge 17 - 2x_2\,, \\ -x_1 - x_2 - 12 & \text{if } x_2 \ge x_1 \ \&\ 2x_1 \le 17 - x_2 \ \&\ 2x_1 \ge 17 - 2x_2\,, \\ & \quad \text{or } x_2 \le x_1 \ \&\ x_1 \le 17 - 2x_2 \ \&\ 2x_1 \ge 17 - 2x_2\,, \\ x_1 + x_2 - 5 & \text{if } 2x_1 \le 17 - 2x_2\,. \end{cases}$$

The regions of this function are shown in Figure 3 and its graph is depicted in Figure 4.

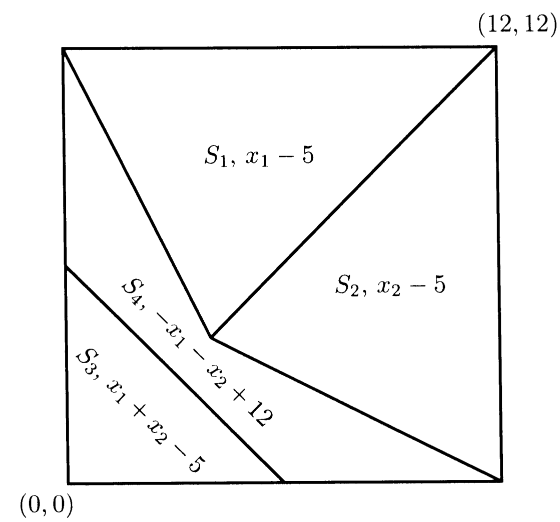


FIGURE 3. The regions of the function $f \colon \mathbb{R}^2 \to \mathbb{R}$.

Notice that the regions cannot be specified by separate thresholds on the variables x_1 and x_2. This would be the case only when the function f is itself separable. ■

The rest of the discussion in this section is devoted to the properties of piecewise linear functions. The fundamental result for our work will be obtained in the sequel (see Theorem 4.15) and it states that the collection of all piecewise linear functions is precisely the Riesz subspace generated in $C(\mathbb{R}^m)$ by the affine functions.

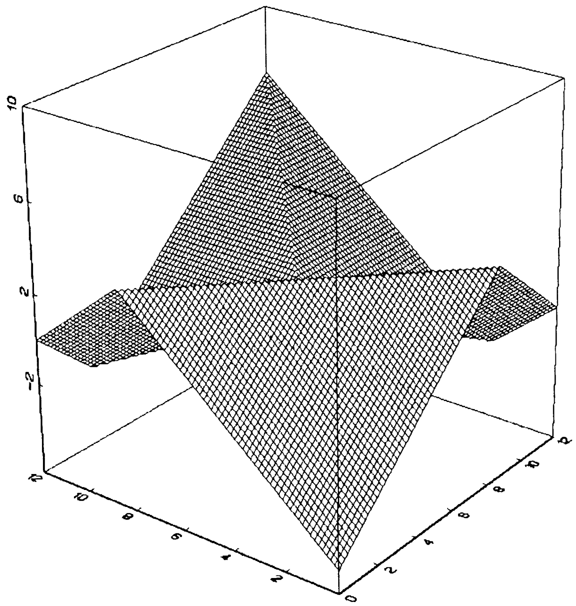


FIGURE 4. The graph of $f\colon \mathbb{R}^2 \to \mathbb{R}$

Lemma 4.4. *Every piecewise linear function is continuous.*

Proof. Let $f\colon \mathbb{R}^m \to \mathbb{R}$ be a piecewise linear function and let $x_n \to x$. If $f(x_n) \not\to f(x)$, then we can assume (by passing to a subsequence) that there exists some $\epsilon > 0$ such that $|f(x_n) - f(x)| \geq \epsilon$ for each n. Now notice that there exist some i and a subsequence $\{y_n\}$ of $\{x_n\}$ satisfying $y_n \in S_i$ for each n. But then we have $\epsilon \leq |f(y_n) - f(y)| = |f_i(y_n) - f_i(y)| \to 0$, which is impossible. This shows that f is continuous. ■

The following result presents an extremely simple characterization of piecewise linear functions.

Theorem 4.5. *A continuous function $f\colon \mathbb{R}^m \to \mathbb{R}$ is piecewise linear if and only if there exist affine functions $f_1, \dots, f_k$ such that for each $x \in \mathbb{R}^m$ there exists some $1 \leq i \leq k$ satisfying $f(x) = f_i(x)$.*

Moreover, the set of components of f is a subcollection of the collection of affine functions $\{f_1, \dots, f_k\}$.

Proof. If f is piecewise linear, then the condition is trivially true. So, for the converse, assume that there exist affine functions $f_1, \dots, f_k$ such that for each $x \in \mathbb{R}^m$ there exists some $1 \leq i \leq k$ such that $f(x) = f_i(x)$. We can assume that the affine functions $f_1, \dots, f_k$ are distinct. We claim the following:

- *For each non-empty open subset V of $\mathbb{R}^m$ there exist a non-empty open subset W of V and some $1 \leq i \leq k$ such that $f = f_i$ on W.*

To see this, assume by way of contradiction that the claim is false. This implies that $f \neq f_1$ on V, i.e., $f_1(v) \neq f(v)$ for some $v \in V$. Since f and f_1 are continuous, there exists some non-empty open subset V_1 of V such that $f_1(x) \neq f(x)$ for all $x \in V_1$. Similarly, since (by our hypothesis) $f \neq f_2$ on V_1 there exists some non-empty open subset V_2 of V_1 such that $f_2(x) \neq f(x)$ for all $x \in V_2$. Continuing this way, we see that there exist non-empty open sets $V_k \subseteq V_{k-1} \subseteq \cdots \subseteq V_1 \subseteq V$ such that for each $1 \leq i \leq k$ we have $f_i(x) \neq f(x)$ for all $x \in V_i$. But then for each $x \in V_k$ we have $f(x) \neq f_i(x)$ for all $1 \leq i \leq k$, which is impossible, and our claim has been established.

Now for each $1 \leq i \leq k$ let $\mathcal{O}_i = \bigcup\{U \subseteq \mathbb{R}^m\colon\ U \text{ is open and } f = f_i \text{ on } U\}$. That is, $\mathcal{O}_i$ is the largest open set on which $f = f_i$. By the preceding discussion $\mathcal{O}_i \neq \emptyset$ for at least one i. (To see this take $V = \mathbb{R}^m$ and apply $(\bullet)$.) Deleting the $\mathcal{O}_i$ with $\mathcal{O}_i = \emptyset$, we can assume that $\mathcal{O}_i \neq \emptyset$ for each i. Put $S_i = \overline{\mathcal{O}_i}$, and note that $f = f_i$ on S_i. We shall verify that the closed sets $S_1, \ldots, S_k$ satisfy the conditions of Definition 4.2. Start by observing that condition (4) is obvious.

For (1) note that from $\mathcal{O}_i \subseteq S_i$, we get that $\operatorname{Int}(S_i) \neq \emptyset$ and that $\mathcal{O}_i \subseteq \operatorname{Int}(S_i)$. Moreover, $\mathcal{O}_i = \operatorname{Int}(S_i)$ must be the case, since otherwise the maximality property of $\mathcal{O}_i$ will be violated. The condition $\mathcal{O}_i \cap \mathcal{O}_j = \emptyset$ for $i \neq j$ should be obvious and the validity of (2) follows. If $\bigcup_{i=1}^k S_i \neq \mathbb{R}^m$, then by the above discussion there exists some non-empty open subset Q of $\mathbb{R}^m \setminus \bigcup_{i=1}^k S_i$ and some $1 \leq \ell \leq k$ such that $f = f_\ell$ on Q. But then the open set $\mathcal{O}_\ell \cup Q$ violates the maximality property of $\mathcal{O}_\ell$. Hence, $\bigcup_{i=1}^k S_i = \mathbb{R}^m$.

That the components of f are among the affine functions $f_1, \ldots, f_k$ should be obvious from the above discussion. ■

An immediate consequence of the preceding result is that **PL** is a Riesz subspace.

Corollary 4.6. *The collection of all piecewise linear functions on $\mathbb{R}^m$ is a Riesz subspace of $C(\mathbb{R}^m)$. In particular, we have $\mathcal{R}(\mathbf{Aff}) = \mathbf{Aff}^{\vee\wedge} = \mathbf{Aff}^{\wedge\vee} \subseteq \mathbf{PL}$.*

Recall that an **affine transformation** from $\mathbb{R}^k$ to $\mathbb{R}^m$ is any function $T\colon \mathbb{R}^k \to \mathbb{R}^m$ of the form $T(t) = At + b$, where A is an $m \times k$ real matrix and $b \in \mathbb{R}^m$ is a fixed vector. Now if T is an affine transformation and $f\colon \mathbb{R}^m \to \mathbb{R}$ is an affine function, then the function $f \circ T\colon \mathbb{R}^k \to \mathbb{R}$ is also an affine function. To see this, assume that f is defined as $f(x) = \alpha + u \cdot x$ and note that for each $t \in \mathbb{R}^k$ we have

$$[f \circ T](t) = f(T(t)) = \alpha + u \cdot (At + b) = (\alpha + u \cdot b) + (A'u) \cdot t\,.$$

This conclusion in connection with Theorem 4.5 yields the following result.

Corollary 4.7. *If $f\colon \mathbb{R}^m \to \mathbb{R}$ is a piecewise linear function and $T\colon \mathbb{R}^k \to \mathbb{R}^m$ is an affine transformation, then the function $f \circ T\colon \mathbb{R}^k \to \mathbb{R}$ is piecewise linear. Moreover, if f has the components $f_1, \ldots, f_p$, then the components of $f \circ T$ are among the affine functions $f_1 \circ T, \ldots, f_p \circ T$.*

In particular, for any two fixed vectors $a, b \in \mathbb{R}^m$ the function $\theta\colon \mathbb{R} \to \mathbb{R}$, defined via the formula $\theta(t) = f(ta + (1-t)b)$, is (one-dimensional) piecewise linear.

A **hyperplane** of $\mathbb{R}^m$ is any subset of the form $H = \{x \in \mathbb{R}^m\colon\ a \cdot x = \alpha\}$, where $a \in \mathbb{R}^m$ is a non-zero vector and $\alpha \in \mathbb{R}$ is a constant. Clearly, every hyperplane is a closed set and has Lebesgue measure zero. Notice that two affine functions $f, g\colon \mathbb{R}^m \to \mathbb{R}$ either do not agree at any point or the set that they agree is a hyperplane, i.e., the set $[f = g] = \{x \in \mathbb{R}^m\colon\ f(x) = g(x)\}$ is either empty or a hyperplane.

The boundaries of the regions of a piecewise linear function are parts of hyperplanes.

Lemma 4.8. *Let $(S_1, f_1), \dots, (S_p, f_p)$ be the characteristic pairs of a piecewise linear function $f\colon \mathbb{R}^m \to \mathbb{R}$. For each i let $\mathcal{I}_i = \{j \in \{1, \dots, p\}\colon\ j \neq i$ and $S_i \cap S_j \neq \emptyset\}$. Then the boundary of the region S_i has the following property:*

$$\partial S_i = \bigcup_{j \in \mathcal{I}_i} S_i \cap S_j \subseteq \bigcup_{j \in \mathcal{I}_i} [f_i = f_j]\,.$$

In particular,

(a) *each boundary ∂S_i has Lebesgue measure zero and consists of "parts" of hyperplanes, and*
(b) *if $x \in \operatorname{Int}(S_i)$ for some i, then $x \notin S_j$ for all $j \neq i$.*

Proof. Let $x \in \partial S_i$. Since $B(x, \frac{1}{n}) \cap (\mathbb{R}^m \setminus S_i) \neq \emptyset$, there exists for each n some $x_n \in \bigcup_{r \neq i} S_r$ such that $x_n \in B(x, \frac{1}{n})$. It follows that for some $j \neq i$ we have $x_n \in S_j$ for infinitely many n. This implies $x \in \overline{S_j} = S_j$ and so $x \in S_i \cap S_j$.

Now assume that $x \in S_i \cap S_j$ for some $j \neq i$. If $x \notin \partial S_i$, then $x \in \operatorname{Int}(S_i)$ and so there exists some $\delta > 0$ such that $B(x, \delta) \subseteq \operatorname{Int}(S_i)$. Since $\operatorname{Int}(S_i) \cap \operatorname{Int}(S_j) = \emptyset$, we infer that $x \in \partial S_j$. From $\overline{\operatorname{Int}(S_j)} = S_j$, it follows that there exists some $y \in \operatorname{Int}(S_j)$ such that $y \in B(x, \delta)$. This implies $y \in \operatorname{Int}(S_i) \cap \operatorname{Int}(S_j)$, which is impossible. Consequently, $x \in \partial S_i$, and the proof is finished. ■

The characteristic pairs of a piecewise linear function are uniquely determined.

Lemma 4.9. *The regions and the components of a piecewise linear function $f\colon \mathbb{R}^m \to \mathbb{R}$ are uniquely determined in the following sense: If another collection of pairs $\{(S'_1, g_1), \dots, (S'_q, g_q)\}$ satisfies properties (1)–(4) of Definition 4.2, then $q = p$ and $\{(S'_1, g_1), (S'_2, g_2), \dots, (S'_q, g_q)\}$ is a permutation of the collection of pairs $\{(S_1, f_1), (S_2, f_2), \dots, (S_p, f_p)\}$.*

Proof. Fix some $1 \leq i \leq p$. Since $\operatorname{Int}(S_i)$ is non-empty (and hence it has positive Lebesgue measure), it follows from Lemma 4.8 that there exists some $1 \leq j \leq q$ such that the open set $V = \operatorname{Int}(S_i) \cap \operatorname{Int}(S'_j)$ is non-empty. In particular, since $f_i(x) = g_j(x) = f(x)$ holds true for each $x \in V$, it follows from part (2) of Lemma 4.1 that $f_i = g_j$.

Now let $x \in \operatorname{Int}(S_i)$. Fix $\delta > 0$ such that $B(x, \delta) \subseteq \operatorname{Int}(S_i)$ and let $0 < \epsilon < \delta$. As above, $B(x, \epsilon) \cap \operatorname{Int}(S'_r) \neq \emptyset$ must hold true for some index $1 \leq r \leq q$. But then (as above again) $g_j = f_i = g_r$ must be the case. Since the affine functions $g_1, \dots, g_q$ are all distinct, we infer that $r = j$. Therefore, $B(x, \epsilon) \cap \operatorname{Int}(S'_j) \neq \emptyset$ for all $0 < \epsilon < \delta$. This implies $x \in \overline{S'_j} = S'_j$, and so $\operatorname{Int}(S_i) \subseteq S'_j$. Consequently, $S_i = \overline{\operatorname{Int}(S_i)} \subseteq S'_j$.

By the symmetry of the situation, there exists some $1 \leq m \leq p$ such that $S'_j \subseteq S_m$. This implies $\operatorname{Int}(S_i) \cap \operatorname{Int}(S_m) = \operatorname{Int}(S_i) \neq \emptyset$, from which it follows that $m = i$. Therefore, $S_i = S'_j$ and so $(S_i, f_i) = (S'_j, g_j)$. From the last result, the desired conclusion now easily follows. ■

Another consequence of Theorem 4.5 is that for real functions defined on $\mathbb{R}$ the definitions for piecewise linear functions given in Definitions 3.1 and 4.2 are equivalent.

Corollary 4.10. *A function $f\colon \mathbb{R} \to \mathbb{R}$ is piecewise linear according to Definition 3.1 if and only if it is piecewise linear according to Definition 4.2.*

Proof. Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function. If f is piecewise linear according to Definition 3.1, then f is clearly piecewise linear according to Definition 4.2. For the converse, assume that f is piecewise linear according to Definition 4.2. Let $\{(S_1, f_1), (S_2, f_2), \ldots, (S_p, f_p)\}$ be the collection of characteristic pairs of f. Notice that every f_i is of the form $f_i(t) = m_i t + b_i$. So, every non-empty set of the form $[f_i = f_j]$ is simply a point of $\mathbb{R}$. This is connection with Lemma 4.8 shows that the boundary of each S_i is a finite set. Now each $\mathrm{Int}(S_i)$ is the union of an at most countable collection of pairwise disjoint open intervals. Since ∂S_i is a finite set, a moment's thought reveals that $\mathrm{Int}(S_i)$ is a union of a finite number of pairwise disjoint open intervals. From this it follows that S_i is the union of the closures of these intervals. Now it is easy to see that f is a piecewise linear function according to Definition 3.1. ∎

In order to further study piecewise linear functions, we shall need the theory of arrangements of hyperplanes and oriented matroids, which are well studied combinatorial constructions that are closely related to vector lattices and the simplex methods in linear programming; see Chapter 4 of Björner, Las Vergnas, Sturmfels, White, and Ziegler (1999).

Recall once more that any subset of $\mathbb{R}^m$ of the form $H = \{x \in \mathbb{R}^m\colon\ a \cdot x = \alpha\}$, where $a \in \mathbb{R}^m$ is a non-zero fixed vector and $\alpha \in \mathbb{R}$ is a constant, is called a **hyperplane** of $\mathbb{R}^m$. We can assume without loss of generality that $\|a\| = 1$ and refer to a as a (unit) vector **normal** to H. Since $H = \{x \in \mathbb{R}^m\colon\ (-a) \cdot x = -\alpha\}$, we see that $-a$ is also another (unit) normal vector to H. In other words, H has essentially two unit normal vectors, each of which defines an **orientation** in the sense that it divides $\mathbb{R}^m$ into three parts: a "positive" part $\{x \in \mathbb{R}^m\colon\ a \cdot x > \alpha\}$, a "zero" part $\{x \in \mathbb{R}^m\colon\ a \cdot x = \alpha\}$, and a "negative" part $\{x \in \mathbb{R}^m\colon\ a \cdot x < \alpha\}$. Of course, if we let $H = \{x \in \mathbb{R}^m\colon\ (-a) \cdot x = -\alpha\}$, then the orientation changes: the positive part is now negative and the negative part is positive. Thus, writing H in the form $H = \{x \in \mathbb{R}^m\colon\ a \cdot x = \alpha\}$, the vector a defines automatically an orientation, and H is called an **oriented hyperplane**.

Now let E be a finite index set and let $(H_e)_{e\in E}$, where $H_e = \{x \in \mathbb{R}^m\colon\ a_e \cdot x = \alpha_e\}$, be a family of (oriented) hyperplanes in $\mathbb{R}^m$. The family $(H_e)_{e\in E}$, is called an ***oriented arrangement of hyperplanes*** (or simply an ***arrangement***). Every arrangement of hyperplanes $(H_e)_{e\in E}$ "almost" subdivides $\mathbb{R}^m$ into a finite number of non-empty convex regions. The subdivisions are obtained by means of the "sign" mapping $x \mapsto \sigma_x$, from $\mathbb{R}^m$ to $\{+, -, 0\}^E$, that is defined by

$$\sigma_x(e) = \begin{cases} + & \text{if} \quad a_e \cdot x > \alpha_e\,, \\ - & \text{if} \quad a_e \cdot x < \alpha_e\,, \\ 0 & \text{if} \quad a_e \cdot x = \alpha_e\,, \end{cases}$$

i.e., $\sigma_x = \big(\mathrm{Sign}\,(a_e \cdot x - \alpha_e)\big)_{e\in E}$. Let $\mathcal{M}$ denote the range of σ, i.e., $\mathcal{M} = \sigma(\mathbb{R}^m) \subseteq \{+, -, 0\}^E$.

A vector $T \in \mathcal{M}$ satisfying $T(e) \neq 0$ for all $e \in E$ is called a ***tope*** of $\mathcal{M}$. Note that σ_x is a tope if and only if $x \notin \bigcup_{e\in E} H_e$. Let $T_1, T_2 \ldots, T_J$ be an enumeration of the topes of $\mathcal{M}$. For each $1 \leq h \leq J$ let

$$K_h = \big\{x \in \mathbb{R}^m\colon\ \sigma_x = T_h\big\} = \sigma^{-1}(\{T_h\})\,.$$

Obviously, each K_h is a non-empty open convex set and from $\bigcup_{h=1}^J K_h = \mathbb{R}^m \backslash \bigcup_{e\in E} H_e$, we see that $\overline{\bigcup_{h=1}^J K_h} = \mathbb{R}^m$. The sets $K_1, K_2, \ldots, K_J$ are called the ***cells*** induced by the arrangement of the hyperplanes $(H_e)_{e\in E}$. It should not be difficult to see that the collection of cells $\{K_1, K_2, \ldots, K_J\}$ is independent of the orientation of the planes H_e, and so we can refer to $\{K_1, K_2, \ldots, K_J\}$ as the **collection of cells generated** (or **induced**) by the family of hyperplanes $(H_e)_{e\in E}$. For an example of an arrangement of hyperplanes see Figure 5.

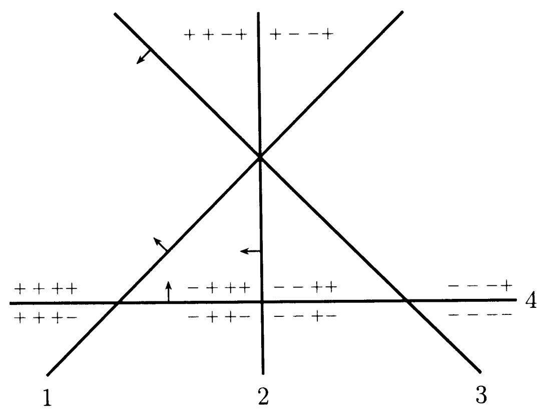


FIGURE 5. An arrangement of 4 oriented hyperplanes in $\mathbb{R}^2$.

Now let $\{f_1, \ldots, f_p\}$, where $p \geq 2$, be a collection of distinct affine functions on $\mathbb{R}^m$. If for each $1 \leq i < j \leq p$ we let $H_{i,j} = [f_i = f_j]$, then the set $E = \{(i,j)\colon\ 1 \leq i < j \leq p \text{ and } H_{i,j} \neq \emptyset\}$ is a finite set. Letting $H_e = [f_i = f_j] = \{x \in \mathbb{R}^m\colon\ a_e \cdot x = \alpha_e\}$ for each $e = (i,j) \in E$, we see that the family $(H_e)_{e \in E}$ is an arrangement of hyperplanes, called **an arrangement generated** by $\{f_1, \ldots, f_p\}$. The collection of cells collection of cells generated by $(H_e)_{e\in E}$ is called the **collection of cells generated** (or **induced**) by $\{f_1, \ldots, f_p\}$.

With this terminology at hand, we are now ready to state several extra properties of piecewise linear functions.

Lemma 4.11. *Let $F = \{f_1, \ldots, f_k\}$ be a finite collection of distinct affine functions of $\mathbb{R}^m$ and let $\{K_1, K_2, \ldots, K_J\}$ be the cells induced by F. Assume also that $f\colon \mathbb{R}^m \to \mathbb{R}$ is a continuous function such that for each $x \in \mathbb{R}^m$ there exists some $1 \leq i \leq k$ satisfying $f(x) = f_i(x)$.*[5] *Then for a vector $x \in K_h$ we have the following:*

1. *If $f(x) = f_i(x)$, then $f(y) = f_i(y)$ for all $y \in K_h$.*
2. *If $f(x) > f_i(x)$, then $f(y) > f_i(y)$ for all $y \in K_h$.*
3. *If $f(x) < f_i(x)$, then $f(y) < f_i(y)$ for all $y \in K_h$.*

Moreover, for each $1 \leq h \leq J$ there exists a unique $1 \leq i_h \leq k$ such that $f = f_{i_h}$ on K_h.

Proof. We shall prove (1) first. To this end, suppose that some $x \in K_h$ satisfies $f(x) = f_i(x)$.

Let $X = \bigcup_{h=1}^{J} K_h$ and note that X is an open dense subset of $\mathbb{R}^m$. Notice that for each $z \in X$ any pair of distinct functions $f_i, f_j \in F$ we have $f_i(z) \neq f_j(z)$. So for each $z \in X$ there exists a unique $1 \leq i_z \leq k$ such that $f(z) = f_{i_z}(z)$. Since f and the f_i are continuous functions and $f(z) = f_{i_z}(z) \neq f_j(z)$ for each $j \neq i_z$, there exists an open neighborhood $N_z \subseteq X$ of z such that for each $y \in N_z$ and all $j \neq i_z$ we have $f(y) \neq f_j(y)$ and $f_{i_z}(y) \neq f_j(y)$. This implies that for each $y \in N_z$ we have $f(y) = f_{i_z}(y)$, i.e., $i_y = i_z$.

[5]Keep in mind that this implies (by Theorem 4.5) that f is piecewise linear.

Now fix $y \in K_h$. Let $L(x,y)$ be the line segment joining x and y and notice that $L(x,y) \subseteq K_h$ since K_h is convex. Since $L(x,y)$ is compact, there exists a finite set $Z = \{z_1, \ldots, z_r\} \subseteq L(x,y)$ such that $L(x,y) \subseteq \bigcup_{z\in Z} N_z$. We can assume that the neighborhoods $\{N_z\colon\ z \in Z\}$ form a chain, i.e., $N_{z_t} \cap N_{z_{t+1}} \neq \emptyset$ for each $t = 1, \ldots, r-1$; see (Abramovich and Aliprantis, 2002b, Problem 1.5.7, p. 50). This easily implies that for each $z \in L(x,y)$ we have $f(z) = f_{i_x}(z) = f_{i_y}(z)$. In particular, $i_x = i_y$.

Therefore, we have shown that for each K_h there exists a unique index $1 \le i_h \le k$ such that $y \in K_h$ implies $f(y) = f_{i_h}(y)$. This proves (1) and the last part of the lemma.

To establish (2), assume that $f(x) > f_i(x)$ holds true for some $x \in K_h$ and that some other $y \in K_h$ satisfies $f(y) \le f_i(y)$. If $f(y) = f_i(y)$, then according to (1) we must have $f(x) = f_i(x)$, which is impossible. If $f(y) < f_i(y)$, then there exists some z in the line segment joining x and y (and hence $z \in K_h$) satisfying $f(z) = f_i(z)$. But then (according to (1) again) we get $f(x) = f_i(x)$, a contradiction. This establishes (2) and the validity of (3) can be proven in a similar fashion. ■

From Theorem 4.5 we know that if for a continuous function $f\colon \mathbb{R}^m \to \mathbb{R}$ and affine functions $f_1, \ldots, f_k$ for each $x \in \mathbb{R}^m$ there exists some $1 \le i \le k$ satisfying $f(x) = f_i(x)$, then f is piecewise linear. The next result constructs the characteristic pairs of such a piecewise linear function from a given collection of affine functions.

Theorem 4.12. *Assume that a continuous function $f\colon \mathbb{R}^m \to \mathbb{R}$ and a finite set of distinct affine functions $F = \{f_1, \ldots, f_k\}$ are such that for each $x \in \mathbb{R}^m$ there exists some $1 \le i \le k$ satisfying $f(x) = f_i(x)$. Let $\{K_1, K_2, \ldots, K_J\}$ be the cells generated by F. For each $1 \le i \le k$ let*

$$E_i = \big\{h \in \{1, \ldots, J\}\colon\ f = f_i \text{ on } K_h\big\}$$

and then define $S_i = \overline{\bigcup_{h\in E_i} K_h}$. We have the following.

(a) *If $\{E_i\}_{i\in\mathcal{I}}$ is the family of non-empty E_i, then the family $\{(S_i, f_i)\}_{i\in\mathcal{I}}$ is precisely the family of characteristic pairs of the piecewise linear function f.*
(b) *For each $1 \le h \le J$ there exists exactly one $i \in \mathcal{I}$ such that $K_h \subseteq \mathrm{Int}(S_i)$.*
(c) *For each $i \in \mathcal{I}$ the non-empty set $\mathrm{Int}(S_i)$ is a union of a finite collection of pairwise disjoint non-empty open and connected subsets of $\mathbb{R}^m$.*

Proof. (a) We know from Theorem 4.5 that the function f is piecewise linear whose components are among the $f_1, \ldots, f_k$. The proof below will present also an alternate constructive proof of Theorem 4.5. Let $\{K_1, \ldots, K_J\}$ be the collection of cells generated by F and for each $1 \le i \le k$ define E_i and S_i as in the statement of the lemma.

According to Lemma 4.11 at least one of the E_i is non-empty; relabeling, we can assume that $E_1, \ldots, E_p$ are the non-empty E_i, i.e., $\mathcal{I} = \{1, \ldots, p\}$. Clearly, $f = f_i$ on S_i. Since the affine functions $f_1, \ldots, f_k$ are distinct, it follows from part (2) of Lemma 4.1 that $E_r \cap E_s = \emptyset$ for $r \neq s$ and from Lemma 4.11 we see that $\bigcup_{i=1}^p E_i = \{1, \ldots, J\}$. The latter yields

$$\bigcup_{i=1}^p S_i = \bigcup_{i=1}^p \overline{\bigcup_{h\in E_i} K_h} = \overline{\bigcup_{i=1}^p \bigcup_{h\in E_i} K_h} = \overline{\bigcup_{h=1}^J K_h} = \mathbb{R}^m\,.$$

Next notice that since for each $1 \le i \le p$ we have $\bigcup_{h\in E_i} K_h \subseteq \mathrm{Int}(S_i)$, it follows on one hand that $\mathrm{Int}(S_i) \neq \emptyset$ and on the other hand that $\overline{\mathrm{Int}(S_i)} = S_i$. Moreover, using part (2) of Lemma 4.1, it is easy to see that $\mathrm{Int}(S_r) \cap \mathrm{Int}(S_s) = \emptyset$ for $r \neq s$. Since $f = f_i$ holds true for each

$1 \le i \le p$, it follows from Definition 4.2 that f is a piecewise linear function with characteristic pairs $(S_1, f_1), \dots, (S_p, f_p)$.

(b) Now let $1 \le h \le J$. According to Lemma 4.11 there exists a unique $1 \le i_h \le k$ such that $f = f_{i_h}$ on K_h. This implies that $E_{i_h} \ne \emptyset$ and $K_h \subseteq \mathrm{Int}(S_{i_h})$.

(c) Observe that for each $i \in \mathcal{I}$ every component of $\mathrm{Int}(S_i)$, i.e., every maximal (with respect to $\supseteq$) non-empty and connected subset of $\mathrm{Int}(S_i)$, is open. Now notice that every $K_h \subseteq \mathrm{Int}(S_i)$ is open and connected (as being a convex set) and so is included in some component of $\mathrm{Int}(S_i)$. Moreover, from the definition of S_i, it is not difficult to see that every component of $\mathrm{Int}(S_i)$ includes some K_h. Thus, the number of components of $\mathrm{Int}(S_i)$ is at most J, and the proof is finished. ■

To continue our study, we need one more property of piecewise linear functions.

Lemma 4.13 (Ovchinnikov (2002)). *If $f\colon \mathbb{R}^m \to \mathbb{R}$ is a piecewise linear function with components $f_1, \dots, f_p$, then for any pair $a, b \in \mathbb{R}^m$ there exists a component f_i of f satisfying $f_i(a) \le f(a)$ and $f_i(b) \ge f(b)$.*

Proof. Fix $a, b \in \mathbb{R}^m$ and consider the function $g\colon \mathbb{R} \to \mathbb{R}$ defined by $g(t) = f(tb + (1-t)a)$. By Corollary 4.7, g is a one-dimensional piecewise linear function whose components are among the affine functions $g_1, g_2, \dots, g_p$, where $g_i(t) = f_i(tb + (1-t)a)$. Consider g restricted to $[0,1]$ and then use Lemma 3.2 in conjunction with Corollary 3.7 to see that there exists a component g_i satisfying $f(a) = g(0) \ge g_i(0) = f_i(a)$ and $f(b) = g(1) \le g_i(1) = f_i(b)$. ■

We are now ready to state and prove one of the major results of this work. It includes the basic structural properties of piecewise linear functions. The proof is based on the discussion by Ovchinnikov of referees' comments concerning his paper Ovchinnikov (2002).

Theorem 4.14. *Assume that $f\colon \mathbb{R}^m \to \mathbb{R}$ is a piecewise linear function with characteristic pairs $\{(S_1, f_1), \dots, (S_p, f_p)\}$ and let $\{K_1, K_2, \dots, K_J\}$ be the set of cells induced by $\{f_1, \dots, f_p\}$.*

1. *If for each h we pick some $x_h \in K_h$ and let $E_h = \big\{i \in \{1, \dots, p\}\colon\ f_i(x_h) \ge f(x_h)\big\}$, then E_h is non-empty and*

$$f = \bigvee_{h=1}^{J} \bigwedge_{i \in E_h} f_i\,.$$

In particular, $f \in \{f_1, f_2, \dots, f_p\}^{\vee\wedge}$.

2. *If J^* is the subset of $\{1, \dots, J\}$ having the property that for each $1 \le h \le J$ there exists a $j \in J^*$ such that $E_j \subseteq E_h$, then we have*

$$f = \bigvee_{j \in J^*} \bigwedge_{i \in E_j} f_i\,.$$

Proof. (1) For each $1 \le h \le J$ fix some $x_h \in K_h$ and then use Theorem 4.12 to choose some $1 \le j \le p$ such that $K_h \subseteq \mathrm{Int}(S_j)$. Clearly, $f_j(x_h) = f(x_h)$. This implies that if for each $1 \le h \le J$ we let

$$E_h = \big\{i \in \{1, \dots, p\}\colon\ f_i(x_h) \ge f(x_h)\big\}\,,$$

then on one hand $E_h \neq \emptyset$ and on the other hand a glance at Lemma 4.11 guarantees that for each $i \in E_h$ and each $y \in K_h$ we have $f_i(y) \geq f(y)$. Now for each $1 \leq h \leq J$ consider the function

$$F_h = \bigwedge_{i \in E_h} f_i\,, \tag{$\star$}$$

and note that $F_h(y) \geq f(y)$ for each $y \in K_h$. Since for some $j \in E_h$ we have $f_j(x_h) = f(x_h)$, it follows from Lemma 4.11 that $f_j(y) = f(y)$ for all $y \in K_h$. Thus, $F_h(y) = f(y)$ for all $y \in K_h$.

Next, fix $y \in \mathbb{R}^m$. For each $1 \leq h \leq J$ there exists (according to Lemma 4.13) some f_j satisfying $f_j(y) \leq f(y)$ and $f_j(x_h) \geq f(x_h)$. In particular, it follows that we have $j \in E_h$ and consequently $F_h(y) = \big[\bigwedge_{i \in E_h} f_i\big](y) \leq f(y)$ for all $1 \leq h \leq J$. This implies $\big[\bigvee_{h=1}^J F_h\big](y) \leq f(y)$ for all $y \in \mathbb{R}^m$.

On the other hand, since for each $x \in K_h$ we have $F_h(x) = f(x)$, it must be the case that

$$\bigvee_{h=1}^{J} F_h = \bigvee_{h=1}^{J} \bigwedge_{i \in E_h} f_i = f\,,$$

on $\bigcup_{h=1}^J K_h$. Since $\bigcup_{h=1}^J K_h$ is dense in $\mathbb{R}^m$ and $\bigvee_{h=1}^J \bigwedge_{i \in E_h} f_i$ and f are both continuous functions, it follows that $\bigvee_{h=1}^J \bigwedge_{i \in E_h} f_i = f$ holds true on $\mathbb{R}^m$.

(2) To establish this identity, note first that if $E_j \subseteq E_h$, then $\bigwedge_{i \in E_h} f_i \leq \bigwedge_{i \in E_j} f_i \leq f$. This implies $\bigwedge_{i \in E_h} f_i \leq \bigvee_{j \in J^*} \bigwedge_{i \in E_j} f_i \leq f$ for each $1 \leq h \leq J$, and consequently we have $f = \bigvee_{h=1}^J \bigwedge_{i \in E_h} f_i \leq \bigvee_{j \in J^*} \bigwedge_{i \in E_j} f_i \leq f$, and the proof is finished. ■

Combining Corollary 4.6 and Theorem 4.14 we are now ready to state the fundamental result for this work.

Theorem 4.15. *The vector space* **PL** *of all piecewise linear functions is a vector sublattice of* $C(\mathbb{R}^m)$ *and coincides with* $\mathbf{Aff}^{\vee\wedge}$, *i.e.,* $\mathbf{PL} = \mathbf{Aff}^{\vee\wedge}$.

In other words, **PL** *is precisely the Riesz subspace of* $C(\mathbb{R}^m)$ *generated by the* $(m+1)$*-dimensional vector subspace* **Aff** *of all affine functions.*

The next example reported in Ovchinnikov (2002) shows that piecewise polynomial functions need not admit a sup-inf representation.

Example 4.16. Define the piecewise quadratic function $f\colon \mathbb{R} \to \mathbb{R}$ as follows:

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0\,, \\ x^2 & \text{if } x > 0\,. \end{cases}$$

Notice that $x^2 \vee 0 = x^2$ and $x^2 \wedge 0 = 0$. ■

Noting that the set $\{f_1, f_2, \ldots, f_p\}^{\vee\wedge}$ is finite, Theorem 4.15 yields also the following.

Corollary 4.17. *If* $F = \{f_1, f_2, \ldots, f_p\}$ *is a finite set of affine functions on* $\mathbb{R}^m$, *then a function* $f \in C(\mathbb{R}^m)$ *is piecewise linear with components in* F *if and only if* f *belongs to the finite set* $F^{\vee\wedge}$.

Theorem 4.14 also provides an algorithm for constructing the sup-inf representation of a piecewise linear function with components $f_1, f_2, \ldots, f_p$ and unknown regions. The next example is a rudimentary algorithm illustrating this.

Example 4.18 (From $f, f_1, f_2, \ldots, f_p$ to $\mathbf{Aff}^{\vee\wedge}$)**.** Take $f \in \mathbf{PL}$ with components $f_1, f_2, \ldots, f_p$. Following Theorem 4.14 the function f can be re-constructed following the steps below:

Step I: Determine $E = \{(i,j)\colon\ 1 \le i < j \le p \text{ and } [f_i = f_j] \neq \emptyset\}$ and then for each $e = (i,j) \in E$ pick $\alpha_e \in \mathbb{R}$ and $a_e \in \mathbb{R}^m$ such that $H_e = \{x \in \mathbb{R}^m\colon\ a_e \cdot x = \alpha_e\} = [f_i = f_j]$.

Step II: Using the hyperplane arrangement $(H_e)_{e\in E}$ determine that cells $K_1, \ldots, K_J$.

Step III: For each $h = 1, 2, \ldots, J$ choose some x_h from the cell K_h.

Step IV: For each $h = 1, 2, \ldots, J$ determine $E_h = \{i \in \{1, \ldots, p\}\colon\ f_i(x_h) \ge f(x_h)\}$.

Step V: Select a "minimal" set $J^* \subseteq \{1, 2, \ldots, J\}$ so that it satisfies property (2) of Theorem 4.14. Then we have

$$f = \bigvee_{j\in J^*} \bigwedge_{i\in E_j} f_i\,.$$

The above procedure gives a desired sup-inf representation of f. ■

The next example illustrates the preceding algorithm. It also shows how in applying this algorithm, we can restrict our attention to a closed convex domain with non-empty interior.

Example 4.19. Consider once again Example 4.3 but with the restricted domain shown in Figure 6. Take the four affine components of the function f:

$$\begin{aligned} f_1(x_1,x_2) &= x_1 - 5\,,\\ f_2(x_1,x_2) &= x_2 - 5\,,\\ f_3(x_1,x_2) &= x_1 + x_2 - 5\,,\\ f_4(x_1,x_2) &= -x_1 - x_2 + 12\,. \end{aligned}$$

These four affine functions induce eight cells. They are the eight regions of the oriented matroid in Step I of the algorithm of Example 4.18 and they are depicted in Figure 6.

Notice that $E_1 = \{1,2,3\}$, $E_2 = \{1,2,3\}$, $E_3 = E_4 = E_5 = E_6 = E_7 = E_8 = \{3,4\}$. Therefore, if we take $J^* = \{1,3\}$, then we can write

$$f = (f_3 \wedge f_4) \vee (f_1 \wedge f_2 \wedge f_3)\,.$$

Since we have restricted the domain, we can now write $f = (f_3 \wedge f_4) \vee (f_1 \wedge f_2)$; compare Figures 4 and 7. ■

A rudimentary algorithm for computing the regions of the functions in $\mathbf{Aff}^{\vee\wedge}$ by means of Theorem 4.12 is presented next.

Example 4.20 (From $\mathbf{Aff}^{\vee\wedge}$ to $\mathbf{PL}$)**.** Take $f \in \{f_1, f_2, \ldots, f_p\}^{\vee\wedge}$, where $f_1, f_2, \ldots, f_p$ are affine functions on R^m. Following Theorem 4.12 the regions of the function f can be obtained following the steps below:

Step I: Determine $E = \{(i,j)\colon\ 1 \le i < j \le p \text{ and } [f_i = f_j] \neq \emptyset\}$ and for each $e = (i,j) \in E$ and then pick $\alpha_e \in \mathbb{R}$ and $a_e \in \mathbb{R}^m$ such that $H_e = \{x \in \mathbb{R}^m\colon\ a_e \cdot x = \alpha_e\} = [f_i = f_j]$.

Step II: Use the hyperplane arrangement $(H_e)_{e\in E}$ to determine the cells $K_1, \ldots, K_J$.

Step III: For each $h = 1, 2, \ldots, J$ choose some $x_h \in K_h$ and then let

$$i_h = \min\{i \in \{1, \ldots, p\}\colon\ f_i(x_h) = f(x_h)\}\,.$$

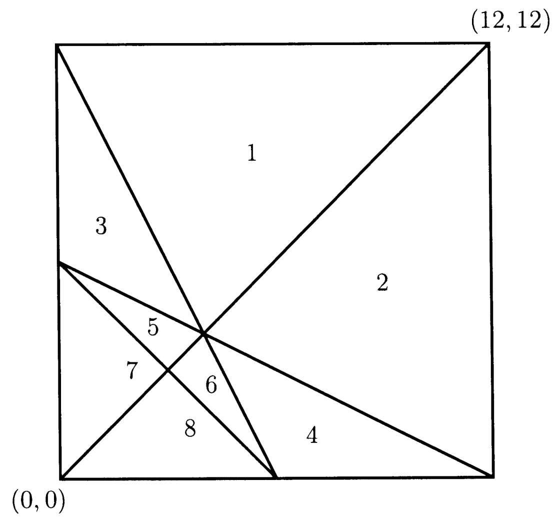


FIGURE 6. The eight regions of the oriented matroid.

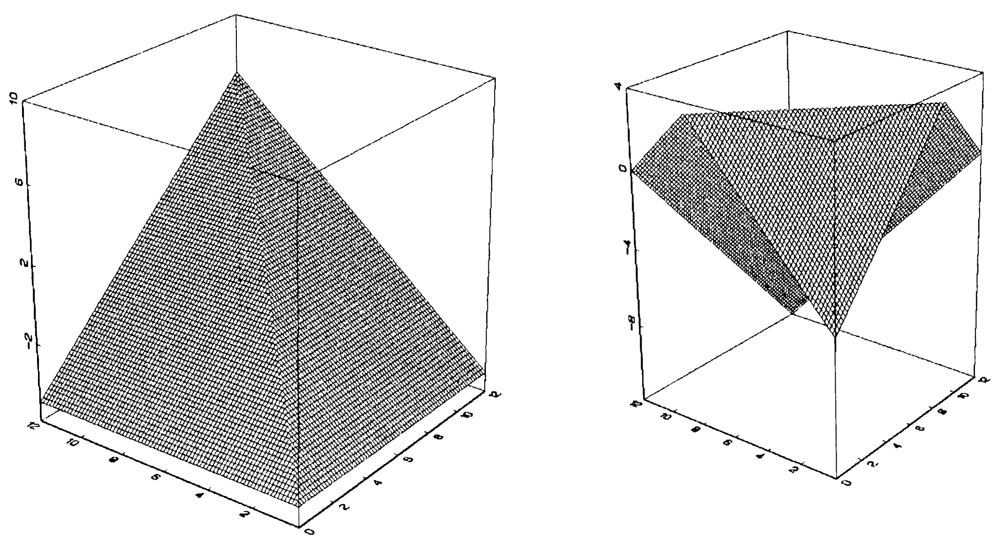


FIGURE 7. The graphs of $f_1 \wedge f_2$ and $f_3 \wedge f_4$.

Step IV: For each $h = 1, 2, \ldots, J$ determine the set $I_h = \{j \in \{1, \ldots, J\} \colon \ i_j = i_h\}$.
Step V: For each $h = 1, 2, \ldots, J$ let $S_{i_h} = \overline{\bigcup_{i \in I_h} K_j}$.

The characteristics pairs of f are distinct members of the family $\{(f_{i_h}, S_{i_h})\}_{h \in \{1, \ldots, J\}}$. ∎

5. The space of Riesz Estimators

In this section we study the space of Riesz estimators, establish approximation results for Riesz estimators, and alternative formulations of the estimators. As mentioned earlier, we are working in the Hilbert space (Banach lattice) $L_2(\pi) = L_2(\Omega, \mathcal{F}, \pi)$. Recall that a random variable $\widehat{Y}$ is said to be an ***estimator*** based on a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_m)$, if we can write $\widehat{Y}$ as a function of realizations of $\mathbf{X}$ in the sense that there exists a Borel measurable function $g\colon \mathbb{R}^m \to \mathbb{R}$ satisfying $\widehat{Y} = g \circ \mathbf{X}$. The function g is called the ***estimating*** function of the random vector $\mathbf{X}$.

Now let $\mathbf{X}$ be a random vector and consider the vector subspace of $L_2(\pi)$ given by

$$A_{\mathbf{X}} = \operatorname{Span}\{\mathbf{1}, X_1, X_2, \ldots, X_m\}.$$

Notice that each $\widehat{Z} \in A_{\mathbf{X}}$ can be written in the form

$$\widehat{Z} = r^0 + r^1 X_1 + r^2 X_2 + \cdots + r^m X_m,$$

where the coefficients r^j are real scalars. If we consider the continuous function $g\colon \mathbb{R}^m \to \mathbb{R}$ defined by $g(t_1, \ldots, t_m) = r^0 + r^1 t_1 + r^2 t_2 + \cdots + r^m t_m$, then $\widehat{Z} = g \circ \mathbf{X}$, and so every random variable in $A_{\mathbf{X}}$ is an estimator based on $\mathbf{X}$ having an affine estimating function. This justifies the name we reserve for the estimators in $A_{\mathbf{X}}$.

Definition 5.1. *The estimators in $A_{\mathbf{X}}$ are called* ***affine estimators*** *based on* $\mathbf{X}$.

Given affine estimators $\widehat{Z}_1, \widehat{Z}_2, \ldots, \widehat{Z}_k \in A_{\mathbf{X}}$ we shall write

$$\bigvee_{i=1}^{k} \widehat{Z}_i \quad \text{and} \quad \bigwedge_{i=1}^{k} \widehat{Z}_i,$$

for the π-almost everywhere pointwise supremum and infimum of the set $\{\widehat{Z}_1, \widehat{Z}_2, \ldots, \widehat{Z}_k\}$, respectively. That is, for π-almost all $\omega \in \Omega$ we let

$$\Big[\bigvee_{i=1}^{k} \widehat{Z}_i\Big](\omega) = \max\{\widehat{Z}_1(\omega), \widehat{Z}_2(\omega), \ldots, \widehat{Z}_k(\omega)\}, \text{ and}$$

$$\Big[\bigwedge_{i=1}^{k} \widehat{Z}_i\Big](\omega) = \min\{\widehat{Z}_1(\omega), \widehat{Z}_2(\omega), \ldots, \widehat{Z}_k(\omega)\}.$$

Notice that $\bigvee_{i=1}^{k} \widehat{Z}_i$ and $\bigwedge_{i=1}^{k} \widehat{Z}_i$ are not usually in $A_{\mathbf{X}}$ but they always lie in the vector space of estimators $L_2(\sigma(\mathbf{X}))$.

As before, we shall denote by $A_{\mathbf{X}}^{\wedge}$ the collection of all random variables that can be written as infima of finite subsets of $A_{\mathbf{X}}$. That is, a random vector $\widehat{Z} \in L_2(\pi)$ belongs to $A_{\mathbf{X}}^{\wedge}$ if and only if there exist random variables $\widehat{Z}_1, \widehat{Z}_2, \ldots, \widehat{Z}_h \in A_{\mathbf{X}}$ such that $\widehat{Z} = \bigwedge_{i=1}^{k} \widehat{Z}_i$. Similarly, $A_{\mathbf{X}}^{\vee}$ is the set consisting of all suprema of the finite subsets of $A_{\mathbf{X}}$. Write $A_{\mathbf{X}}^{\vee\wedge}$ for $(A_{\mathbf{X}}^{\vee})^{\wedge}$ and $A_{\mathbf{X}}^{\wedge\vee}$ for $(A_{\mathbf{X}}^{\wedge})^{\vee}$.

We now introduce the notion of a Riesz estimator.

Definition 5.2. *Any random variable in $A_{\mathbf{X}}^{\vee\wedge}$ is called a* ***Riesz estimator*** *based on* $\mathbf{X}$.

A direct application of Lemma 2.2 yields the following.

Lemma 5.3. *The collection $A_{\mathbf{X}}^{\vee\wedge}$ of all Riesz estimators based on* $\mathbf{X}$ *coincides with the vector sublattice $\mathcal{R}(A_{\mathbf{X}})$ generated by the vector space of all affine estimators $A_{\mathbf{X}}$. That is, we have*

$$\mathcal{R}(A_{\mathbf{X}}) = A_{\mathbf{X}}^{\vee\wedge} = A_{\mathbf{X}}^{\wedge\vee}\,.$$

In other words, a random variable $\widehat{Z}$ is a Riesz estimator if and only if there exist affine estimators $\{\widehat{Z}_{ij}\colon\ i \in \{1,2,\ldots,p\};\ j \in \{1,2,\ldots,q\}\}$ in $A_{\mathbf{X}}$ such that

$$\widehat{Z} = \bigwedge_{i=1}^{p}\bigvee_{j=1}^{q}\widehat{Z}_{ij}\,.$$

We list below several important and useful consequences of Lemma 5.3.

Theorem 5.4. *A random variable $\widehat{Y}$ is a Riesz estimator based on* $\mathbf{X}$ *if and only if there exists a piecewise linear estimating function $f\colon \mathbb{R}^m \to \mathbb{R}$ such that $\widehat{Y} = f \circ \mathbf{X}$.*

Proof. Observing that for any collection of affine functions $\{f_1, f_2, \ldots, f_p\}$ on $\mathbb{R}^m$) and any family of non-empty subsets $\{E_h\colon\ h = 1, \ldots, J\}$ from this collection we have

$$\bigvee_{h=1}^{J}\bigwedge_{i\in E_h} f_i \circ \mathbf{X} = \Big[\bigvee_{h=1}^{J}\bigwedge_{i\in E_h} f_i\Big] \circ \mathbf{X}\,,$$

the conclusion follows from Theorem 4.14. ■

Corollary 5.5. *For a random vector* $\mathbf{X}$ *the following hold true:*

(a) *Each Riesz estimator $\widehat{Y}$ based on the random vector* $\mathbf{X}$ *can be written as*

$$\widehat{Y} = \bigvee_{i=1}^{p}\bigwedge_{i=1}^{q}\big(r_{ij}^0 + r_{ij}^1 X_1 + r_{ij}^2 X_2 + \cdots + r_{ij}^m X_m\big)\,. \qquad (\mathcal{R})$$

(b) *Every estimator of the form*

$$\widehat{Z} = \alpha + \sum_{i=1}^{m}\beta_i X_i + \sum_{j=1}^{k}\gamma_j\Big[\bigvee_{h=1}^{p_j}\bigwedge_{\ell=1}^{q_j}\Big(r_{jh\ell}^0 + r_{jh\ell}^1 X_1 + r_{jh\ell}^2 X_2 + \cdots + r_{jh\ell}^m X_m\Big)\Big] \qquad (\mathcal{SR})$$

is a Riesz estimator based on the random vector $\mathbf{X}$.

(c) *If $\widehat{Z}_1, \widehat{Z}_2$ are Riesz estimators based on the random vector* $\mathbf{X}$, *then $\widehat{Z}_1 \vee \widehat{Z}_2$, $\widehat{Z}_1 \wedge \widehat{Z}_2$, $\widehat{Z}_1^+$, $\widehat{Z}_1^-$, and $|\widehat{Z}_1|$ are also Riesz estimators based on the random vector* $\mathbf{X}$.

The Riesz estimators enjoy the following important density property.

Theorem 5.6. *The vector space $A_{\mathbf{X}}^{\vee\wedge}$ of all Riesz estimators based on a random vector* $\mathbf{X}$ *is norm dense in the vector space of all estimators based on* $\mathbf{X}$.

Proof. We know that $A_{\mathbf{X}}^{\vee\wedge}$ is the Riesz subspace of $L_2(\pi)$ generated by $\{\mathbf{1}, X_1, X_2, \ldots, X_m\}$ and that $A_{\mathbf{X}}^{\vee\wedge} \subseteq L_2(\sigma(\mathbf{X}))$. Now the closure $\overline{A^{\vee\wedge}}$ is a closed Riesz subspace of $L_2(\mathcal{F})$ and satisfies $\overline{A_{\mathbf{X}}^{\vee\wedge}} \subseteq L_2(\sigma(\mathbf{X}))$.

According to Lemma 2.4 there exists some sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$ such that $\overline{A_{\mathbf{X}}^{\vee\wedge}} = L_2(\mathcal{A})$. In particular, we have $L_2(\mathcal{A}) \subseteq L_2(\sigma(\mathbf{X}))$. Since each X_i belongs to $L_2(\mathcal{A})$, it follows that each X_i is

$\mathcal{A}$-measurable. Therefore, $\sigma(\mathbf{X}) \subseteq \mathcal{A}$ and so $L_2(\sigma(\mathbf{X})) \subseteq L_2(\mathcal{A}) = \overline{A_{\mathbf{X}}^{\vee\wedge}}$ is also true. Consequently, $\overline{A_{\mathbf{X}}^{\vee\wedge}} = L_2(\mathcal{A}) = L_2(\sigma(\mathbf{X}))$. ■

An immediate consequence of the preceding result is the following.

Corollary 5.7. *Assume that $\mathbf{X}$ is a random vector and Y is a random variable. Also, assume that an estimator $\widehat{Z}$ based on $\mathbf{X}$, i.e., $\widehat{Z} \in L_2(\sigma(\mathbf{X}))$, is not the minimum variance estimator of Y, i.e., $\widehat{Z} \neq \mathbb{E}(Y|\mathbf{X})$. Then there exists a Riesz estimator $\widehat{Z}_1$ based on $\mathbf{X}$ of the form*

$$\widehat{Z}_1 = \bigvee_{j=1}^{p} \bigwedge_{k=1}^{q} \left(r_{jk}^0 + r_{jk}^1 X_1 + r_{jk}^2 X_2 + \cdots + r_{jk}^m X_m\right)$$

satisfying $\|\widehat{Z}_1 - Y\| < \|\widehat{Z} - Y\|$.

Some special estimators can be approximated by increasing sequences of Riesz estimators.

Lemma 5.8. *Assume that a random vector $\mathbf{X}$ has support on an order interval $[\mathbf{a}, \mathbf{b}]$ of $\mathbb{R}^m$ and that $g\colon [\mathbf{a}, \mathbf{b}] \to \mathbb{R}$ is a continuous function. If $\widehat{Y} = g \circ \mathbf{X}$, then there exists a sequence of finite Riesz estimators $\{\widehat{Y}_t\}$ based on $\mathbf{X}$ such that $\widehat{Y}_t \uparrow \widehat{Y}$ and $\widehat{Y}_t \to \widehat{Y}$ uniformly on Ω.*

Proof. We can assume that $\mathbf{X} = X \in L_1(\pi)$ and $[\mathbf{a}, \mathbf{b}] = [0, 1]$. Let $g\colon [0,1] \to \mathbb{R}$ be an arbitrary continuous function. Since the vector space of all continuous piecewise linear functions is norm dense in $C[0,1]$, there exists a sequence $\{u_t\} \subseteq C[0,1]$ of piecewise linear functions such that $0 \le u_t \uparrow g$ holds in $C[0,1]$ and $u_t \to g$ uniformly on $[0,1]$. Letting $\widehat{Y}_t = u_t \circ X$, we see that $\widehat{Y}_t \uparrow \widehat{Y}$ and $\widehat{Y}_t \to \widehat{Y}$ uniformly on Ω. ■

We conclude this section with a discussion of additive estimators. Recall that a random variable $\widehat{Y}$ is an ***additive*** or ***separable*** estimator based on $\mathbf{X}$, if we can write $\widehat{Y}$ as a separable function of the realizations of $\mathbf{X}$. That is, if there exist Borel measurable functions $g_i\colon \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \ldots, m$, satisfying

$$\widehat{Y} = g_1 \circ X_1 + g_2 \circ X_2 + \cdots + g_m \circ X_m\,.$$ [6]

For additive models we introduce the following series estimator.

Definition 5.9. *Any random variable $\widehat{Y}$ of the form*

$$\widehat{Y} = \alpha + \sum_{i=1}^{m} \beta_i X_i + \sum_{i=1}^{m} \sum_{j=1}^{k} \gamma_{ij} (X_i - r_{ij})^+\,, \qquad (\mathcal{AR})$$

is said to be an ***additive Riesz estimator*** *based on $\mathbf{X} = (X_1, X_2, \ldots, X_m)$.*

Notice the following characterization of additive Riesz estimators, which is a simple consequence of Lemma 3.5.

[6] When we only have a single observation X, then each estimator based on X is trivially separable. Notice also that the vector space of all additive estimators based on $\mathbf{X}$ is the vector space $L_2(\sigma(X_1)) + L_2(\sigma(X_2)) + \cdots + L_2(\sigma(X_m))$.

Lemma 5.10. *A random variable $\widehat{Y}$ is an additive Riesz estimator based on $\mathbf{X}$ if an only if there exist piecewise (one-dimensional) linear functions $g_i\colon \mathbb{R} \to \mathbb{R}$, $i = 1, 2, \dots, m$ satisfying*

$$\widehat{Y} = g_1 \circ X_1 + g_2 \circ X_2 + \cdots + f_m \circ X_m .$$

The next result shows that additive Riesz estimators approximate all other additive estimators.

Theorem 5.11. *The vector space of all additive Riesz estimators based on $\mathbf{X}$ is norm dense in the vector space of all additive estimators based on $\mathbf{X}$.*

Proof. We know that the vector space of all additive Riesz estimators based on X_i is norm dense in $L_2(\sigma(X_i))$. This implies that the vector space of all additive Riesz estimators is likewise dense in $L_2(\sigma(X_1)) + L_2(\sigma(X_2)) + \cdots + L_2(\sigma(X_m))$. ■

Recall that the ***best affine* (*minimum*) *variance estimator*** $\widehat{Y}$ of Y based on a random vector $\mathbf{X} = (X_1, X_2, \dots, X_m)$ is the unique random variable of the form

$$\widehat{Y} = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m$$

that minimizes $\|Z - Y\|$. That is, $\widehat{Y}$ is the unique solution to the minimization problem:

$$\begin{aligned} &\min \|Z - Y\| \\ &\text{s.t.: } (\alpha, \beta) \in \mathbb{R} \times \mathbb{R}^m \text{ and } Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m . \end{aligned}$$

In other words $\widehat{Y}$ is the best variance estimator based on $A_{\mathbf{X}}$.

Theorem 5.12. *Let Y and $\widehat{Z}$ be random variables and let $\mathbf{X} = (X_1, X_2, \dots, X_m)$ be a random vector such that $\widehat{Z}$ and the conditional expectation $\mathbb{E}(Y|\mathbf{X})$ are additive estimators based on $\mathbf{X}$, i.e.,*

$$\begin{aligned} \widehat{Z} &= f_1 \circ X_1 + f_2 \circ X_2 + \cdots + g_m \circ X_m , \quad \text{and} \\ \mathbb{E}(Y|\mathbf{X}) &= g_1 \circ X_1 + g_2 \circ X_2 + \cdots + g_m \circ X_m , \end{aligned}$$

where the g_i and f_i are Borel measurable functions from $\mathbb{R}$ to $\mathbb{R}$.

Then $\widehat{Z}$ coincides with $\mathbb{E}(Y|\mathbf{X})$, i.e., $\widehat{Z} = \mathbb{E}(Y|\mathbf{X})$, if and only if for any i and any real number r the best affine variance estimator $\widehat{Y}$ of Y based on the random vector $\mathbf{X}_{i,r} = (\widehat{Z}, X_i, (X_i - r)^+)$ is $\widehat{Z}$, i.e., $\|\widehat{Z} - Y\| \le \|\widehat{Y} - Y\|$ for all $\widehat{Y} \in L_2(\sigma(\mathbf{X}_{i,r}))$.

Proof. Assume by way of contradiction that the conclusion is false. This means that if for each $1 \le i \le m$ and each $r \in \mathbb{R}$ the nearest point of to Y in the vector space Span $\{\mathbf{1}, X_i, \widehat{Z}, (X_i - r)^+\}$ is precisely $\widehat{Z}$. This implies that the random variable $Y - \widehat{Z}$ is orthogonal to $\mathbf{1}$, X_i and $(X_i - r)^+$ for for each $1 \le i \le m$ and each $r \in \mathbb{R}$. Consequently, for each $1 \le i \le m$ the random variable $Y - \widehat{Z}$ is orthogonal to

$$V_i = \text{Span}\,\big\{\mathbf{1}, X_i\big\} \bigcup \big\{(X_i - r)^+ \colon\ r \in \mathbb{R}\big\} .$$

Since (by Theorem 5.6) the vector subspace V_i is dense in $L_2(\sigma(X_i))$, it follows that $Y - \widehat{Z}$ is orthogonal to $\sum_{i=1}^m L_2(\sigma(X_i))$. Since $\mathbb{E}(Y|\mathbf{X})$ and $\widehat{Z}$ are elements of $\sum_{i=1}^m L_2(\sigma(X_i))$, we see that

$$\langle Y - \widehat{Z}, \mathbb{E}(Y|\mathbf{X}) - \widehat{Z} \rangle = 0 \quad \text{and} \quad \langle \mathbb{E}(Y|\mathbf{X}) - Y, \mathbb{E}(Y|\mathbf{X}) - \widehat{Z} \rangle = 0 .$$

This implies

$$\begin{aligned}\|\mathbb{E}(Y|\mathbf{X}) - \widehat{Z}\|^2 &= \langle \mathbb{E}(Y|\mathbf{X}) - \widehat{Z}, \mathbb{E}(Y|\mathbf{X}) - \widehat{Z}\rangle \\ &= \langle Y - \widehat{Z}, \mathbb{E}(Y|\mathbf{X}) - \widehat{Z}\rangle + \langle \mathbb{E}(Y|\mathbf{X}) - Y, \mathbb{E}(Y|\mathbf{X}) - \widehat{Z}\rangle = 0\,,\end{aligned}$$

and so $\mathbb{E}(Y|\mathbf{X}) - \widehat{Z}$ or $\widehat{Z} = \mathbb{E}(Y|\mathbf{X})$, which is impossible. ■

Rewriting the preceding result in its contra-positive form we have the following

Corollary 5.13. *Let Y and $\widehat{Z}$ be random variables and let $\mathbf{X}$ be a random vector such that $\widehat{Z}$ and the conditional expectation $\mathbb{E}(Y|\mathbf{X})$ are additive estimators based on $\mathbf{X}$.*

Then $\widehat{Z} \neq \mathbb{E}(Y|\mathbf{X})$ if and only if for there exist some i, some $r \in \mathbb{R}$ and some estimator $\widehat{Y}$ of Y based on the random vector $(\widehat{Z}, X_i, (X_i - r)^+)$ satisfying $\|\widehat{Y} - Y\| < \|\widehat{Z} - Y\|$.

To explain this corollary let us apply it to the case of a single observed random variable.

Corollary 5.14. *If $\widehat{Z}$ is an estimator based on a random variable X and $\widehat{Z} \neq \mathbb{E}(Y|X)$, then there exists $r \in \mathbb{R}$ such that the best affine variance estimator $\widehat{Y}$ of Y based on the random vector $(\widehat{Z}, X, (X - r)^+)$ satisfies $\|\widehat{Y} - Y\| < \|\widehat{Z} - Y\|$.*

The idea is that if $\widehat{Z}$ is not the minimum variance estimator based on X, then there always exists at least one kink or structural break in the estimation function which reduces variance; see Figure 8.

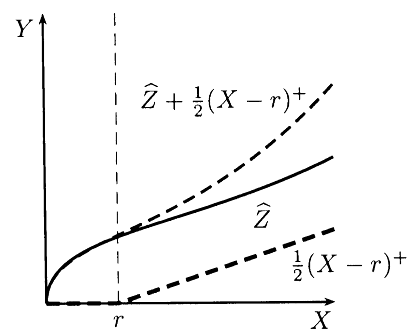


FIGURE 8. The estimator $\widehat{Z} + \frac{1}{2}(X - r)^+$ is $\widehat{Z}$ with a structural break at r.

6. Multivariate piecewise linear regression

We turn to defining the statistical model in which we estimate a multivariate piecewise affine regression. To do this we need the following definition.

Definition 6.1. *Two sets E_1 and E_2 are said to be* **non-comparable** *if neither $E_1 \nsubseteq E_2$ nor $E_2 \nsubseteq E_1$. For any natural number k:*

(i) *The letter Ψ_k will denote the collection of all families of subsets of $\{1, 2, \ldots, k\}$ whose members are pairwise non-comparable.*
(ii) *Arbitrary elements of Ψ_k will be denoted with the usual family notation $\{E_j\}_{j \in J}$.*
(iii) *We let $\#k = \max |J|$ such that $\{E_j\}_{j \in J} \in \Psi_k$.*
(iv) *For any number $q \leq k$, we let $\Psi_{qk} = \{\{E_j\}_{j \in J} \in \Psi_k : |J| \leq q\}$.*

Clearly, $\Psi_{qk} \subseteq \Psi_k$, and so Ψ_{qk} is a finite set. However, as the next example shows, Ψ_k is quite large.

Example 6.2. Consider the case $k = 3$. Note that

$$\Psi_3 = \left\{\begin{array}{llllllll} E_1 & = & \{1\} & & & & & \\ E_1 & = & \{2\} & & & & & \\ E_1 & = & \{3\} & & & & & \\ E_1 & = & \{1,2\} & & & & & \\ E_1 & = & \{1,3\} & & & & & \\ E_1 & = & \{2,3\} & & & & & \\ E_1 & = & \{1,2,3\} & & & & & \\ E_1 & = & \{1\} & E_2 & = & \{2\} & & \\ E_1 & = & \{1\} & E_2 & = & \{3\} & & \\ E_1 & = & \{2\} & E_2 & = & \{3\} & & \\ E_1 & = & \{1\} & E_2 & = & \{2,3\} & & \\ E_1 & = & \{2\} & E_2 & = & \{1,3\} & & \\ E_1 & = & \{3\} & E_2 & = & \{2,1\} & & \\ E_1 & = & \{1,2\} & E_2 & = & \{2,3\} & & \\ E_1 & = & \{1,3\} & E_2 & = & \{2,3\} & & \\ E_1 & = & \{1,2\} & E_2 & = & \{1,3\} & & \\ E_1 & = & \{1,2\} & E_2 & = & \{1,3\} & E_3 = \{2,3\} & \\ E_1 & = & \{1\} & E_2 & = & \{2\} & E_3 = \{3\} & \end{array}\right\}$$

Also for $q = 2$ we have

$$\Psi_{23} = \left\{\begin{array}{llllll} E_1 & = & \{1\} & & & \\ E_1 & = & \{2\} & & & \\ E_1 & = & \{3\} & & & \\ E_1 & = & \{1,2\} & & & \\ E_1 & = & \{1,3\} & & & \\ E_1 & = & \{2,3\} & & & \\ E_1 & = & \{1\} & E_2 & = & \{2\} \\ E_1 & = & \{1\} & E_2 & = & \{3\} \\ E_1 & = & \{1\} & E_2 & = & \{2\} \\ E_1 & = & \{2\} & E_2 & = & \{3\} \\ E_1 & = & \{1\} & E_2 & = & \{2,3\} \\ E_1 & = & \{2\} & E_2 & = & \{1,3\} \\ E_1 & = & \{3\} & E_2 & = & \{2,1\} \\ E_1 & = & \{1,2\} & E_2 & = & \{2,3\} \\ E_1 & = & \{1,3\} & E_2 & = & \{2,3\} \\ E_1 & = & \{1,2\} & E_2 & = & \{1,3\} \end{array}\right\}$$

Notice also that in this case $\#3 = 3$. ■

In this terminology, Corollary 4.17 and Theorem 4.14 can be stated as follows.

- *Let $F = \{f_1, f_2, \ldots, f_p\}$ be a finite set of affine functions. Then a function $f \in C(\mathbb{R}^m)$ belongs to $F^{\vee\wedge}$ if and only if there exists a family $\{E_j\}_{j \in J} \in \Psi_p$ such that $f = \bigvee_{j \in J} \bigwedge_{i \in E_j} f_i$.*
- *If $f : \mathbb{R}^m \to \mathbb{R}$ is a piecewise linear function with affine components $f_1, f_2, \ldots, f_p$, then there exists a family $\{E_j\}_{j \in J} \in \Psi_p$ such that $f = \bigvee_{j \in J} \bigwedge_{i \in E_j} f_i$.*

Recall that every random vector $\mathbf{X} = (X_1, \ldots, X_m)$ defines (in the usual manner) a Borel measure $\pi_{\mathbf{X}}$ on $\mathbb{R}^m$ by letting $\pi_{\mathbf{X}}(B) = \pi\big(\mathbf{X}^{-1}(B)\big)$ for each Borel subset B of $\mathbb{R}^m$. Also, recall that the support $\mathrm{Supp}(\mu)$ of a Borel measure μ is the complement of the largest open set on which μ vanishes. A Borel measure has ***compact support*** if its support is a compact set. Two random vectors $\mathbf{X}$ and $\mathbf{Y}$ are said to be **identically distributed** if $\pi_{\mathbf{X}} = \pi_{\mathbf{Y}}$. In this case notice that for π-almost all ω the values $X(\omega)$ and $Y(\omega)$ lie in $\mathrm{Supp}(\pi_{\mathbf{X}})$.

Definition 6.3. *A* ***piecewise linear regression model*** *consists of the following:*

(1) *Two sequences $\{Y_t\}$ and $\{\mathbf{X}_t\}$ each of which consists of independent and identically distributed (i.i.d.) random variables, where $Y_t\colon \Omega \to \mathbb{R}$ and $\mathbf{X}_t\colon \Omega \to \mathbb{R}^m$.*
(2) *The common support $\mathcal{X}$ of the sequence of i.i.d. random variables $\{\mathbf{X}_t\}$ is compact and has non-empty interior.*
(3) *Affine functions $f_1, f_2, \ldots, f_p\colon \mathbb{R}^m \to \mathbb{R}$ and some $\{E_j\}_{j\in J} \in \Psi_p$ such that*

$$\mathbb{E}(Y_t|\mathbf{X}_t) = \bigvee_{j\in J} \bigwedge_{i\in E_j} f_i \circ \mathbf{X}_t \,. \qquad \textbf{(PL)}$$

For the rest of our discussion we shall assume that the sequences $\{Y_t\}$ and $\{\mathbf{X}_t\}$ are as in Definition 6.3. Notice that since $\mathcal{X}$ has a non-empty interior any affine function defined on $\mathcal{X}$ is the restriction of a unique affine function defined on $\mathbb{R}^m$. Therefore, we can talk about affine functions on $\mathbb{R}^m$ as being in $C(\mathcal{X})$ without creating any confusion.

By Theorem 4.14, the model of Definition 6.3 includes all piecewise linear specifications for $\mathbb{E}(Y_t|\mathbf{X}_t)$. So if we consider each $X_{i,t}$ as a column vector of the matrix $\mathbf{X_t}$, then in matrix form our statistical model is the following multivariate threshold model

$$\mathbb{E}(Y_t|\mathbf{X_t}) = \begin{cases} [\mathbf{1}, \mathbf{X_t}]\beta_1 & \text{if } \mathbf{X_t} \in S_1\,, \\ [\mathbf{1}, \mathbf{X_t}]\beta_2 & \text{if } \mathbf{X_t} \in S_2\,, \\ \quad\vdots & \\ [\mathbf{1}, \mathbf{X_t}]\beta_p & \text{if } \mathbf{X_t} \in S_p\,, \end{cases}$$

where $\beta_i \in \mathbb{R}^{m+1}$, each S_i is a closed subset of $\mathbb{R}^m$, $\mathrm{Int}\,(S_i) \cap \mathrm{Int}\,(_j) = \emptyset$ for $i \neq j$, and $S_1, S_2, \ldots, S_p$ cover $\mathbb{R}^m$. An equivalent specification of the model is that

$$\mathbb{E}(Y_t|\mathbf{X_t}) = f \circ \mathbf{X_t}\,,$$

where $f\colon \mathbb{R}^m \to \mathbb{R}$ is continuous and agrees with a finite set of affine functions $f_1, f_2, \ldots, f_p$.

Notice, however, as with other parametric representations of piecewise linear functions there could be several representations of $\mathbb{E}(Y_t|\mathbf{X}_t)$ in (**PL**). In our case this identification problem is a problem of choosing one of the finite number of equivalent representations in Ψ_p. There are several natural ways to solve this identification problem. One way is to order Ψ_p by size (and lexicographically) and choose the smallest representation of $\mathbb{E}(Y_t|\mathbf{X}_t)$. In this section we will avoid this finite identification problem and consider consistent algorithms for the estimation of $\mathbb{E}(Y_t|\mathbf{X}_t)$ using one of the sup-inf representations.

We now introduce the notion of a consistent estimator.

Definition 6.4. *A* ***consistent estimator*** *of $\mathbb{E}(Y_t|\mathbf{X}_t)$ in* (**PL**) *is a pair of sequences*

$$\left(\{\widehat{E}_j^n\}_{j\in \hat{J}^n}, (\hat{g}_1^n, \hat{g}_2^n, \ldots, \hat{g}_k^n)\right),$$

where k is a fixed natural number, satisfying the following properties:

(a) *For each n and each $1 \le i \le k$ we have $\hat{g}_i^n\colon \Omega \to C(\mathcal{X})$ and $\hat{g}_i^n(\omega)$ is an affine function of $\mathbb{R}^m$ for π-almost all $\omega \in \Omega$.*
(b) $\{\widehat{E}_j^n\}_{j\in \hat{J}^n}\colon \Omega \to \Psi_k$.

(c) *If for each n we define $\hat{f}^n \colon \Omega \to C(\mathcal{X})$ by*

$$\hat{f}^n(\omega) = \bigvee_{j \in \hat{J}^n(\omega)} \bigwedge_{h \in \hat{E}_j^n(\omega)} \hat{g}_h^n(\omega),$$

then the sequence $\{\hat{f}^n\} \colon \Omega \to C(\mathcal{X})$ satisfies $\hat{f}^n \xrightarrow[\|\cdot\|_\infty]{a.s.} \bigvee_{j \in J} \bigwedge_{i \in E_j} f_i$.[7]

For simplicity, in the preceding definition, the function $\widehat{Y}^n \colon \Omega \to C(\mathcal{X})$ will also be written as

$$\hat{f}^n = \bigvee_{j \in \hat{J}^n} \bigwedge_{h \in \hat{E}_j^n} \hat{g}_h^n,$$

where it is understood that $\bigvee_{j \in \hat{J}^n} \bigwedge_{h \in \hat{E}_j^n}$ varies with the states in Ω. Noting that if $f^n \in C(\mathcal{X})$ converges almost surely to f in $C(\mathcal{X})$, then $f^n \circ \mathbf{X_t} \in L_2(\pi)$ converges almost surely to $f \circ \mathbf{X_t}$ in $L_2(\pi)$. In other words, our definition of consistency implies that the estimator converges almost surely to the conditional expectations in $L_2(\pi)$.

Out next task is to introduce two methods of computing consistent estimators of the estimation problem (**PL**). The first is `RIESZVAR(i)`, which is easy to implement on a computer, and the second, `RIESZVAR(ii)`, is a non-linear estimator that requires an efficient algorithm for minimization of piecewise quadratic functions.

7. `RIESZVAR(i)`

We now introduce our first algorithm for the estimation of (**PL**). Theorem 4.14 divides the problem of estimating a solid piecewise linear function into two parts. The first is estimating the affine components of the functions and then searching over the finite number of sup-inf operations on these affine components to find the sup-inf operations that best fits the function. This gives us a "meta-algorithm" for the parametric estimation of piecewise linear functions, which we call `RIESZVAR(i)`. We call it a "meta-algorithm" since it takes as given that we have a method for consistently estimating a superset of functions in $C(\mathcal{X})$ that contains the true piecewise linear functions of $\mathbb{E}(Y_t|\mathbf{X}_t)$. The idea is that this initialization problem is related to standard affine estimation problems in statistics; examples of such algorithms are given in the sequel. The algorithm is shown in Table 1. It highlights the role of the Riesz operation in providing consistent estimation.

The idea in **Module I** of `RIESZVAR(i)` is to estimate a set of affine functions that contains members that consistently estimate the affine components of the piecewise linear function. We provide examples of the implementation of **Module I**.

Example 7.1. Suppose that $\mathbb{E}(Y_t|\mathbf{X}_t)$ is given by (**PL**) as in Definition 6.3. Denote the regions of this piecewise linear function by $S_1, S_2, \ldots, S_p$. Assume for convenience that $\mathbf{X}_t$ has support $[0,1]^m \subseteq \mathbb{R}^m$. We can assume without loss of generality that the interior of each S_h intersects $[0,1]^m$.

Let G be an "regular" grid of $[0,1]^m$ whose mesh size is less than some given $\epsilon > 0$. For ϵ small enough, there exists a cell in the grid G contained in the interior of S_h for every h. For each cell

[7] For a sequence $\hat{x}^n \colon \Omega \to C(\mathcal{X})$ and $x \in C(\mathcal{X})$ we write $\hat{x}^n \xrightarrow[\|\cdot\|_\infty]{a.s.} x$ or simply $\hat{x}^n \xrightarrow{a.s.} x$ if for π-almost all ω $\lim_{n\to\infty} \|\hat{x}^n(\omega) - x\|_\infty = 0$.

RIESZVAR(i)

Consider the estimation of (**PL**) in Definition 6.3 using $\{Y_t, \mathbf{X}_t\}_{t=1}^n$, $(n = 1, 2, \ldots)$:

Module I:

Estimate k affine functions $\hat{g}_1^n, \hat{g}_2^n, \ldots, \hat{g}_k^n : \Omega \to \mathbf{Aff} \subseteq C(\mathcal{X})$ such that

(1) for each $1 \leq i \leq k$ there is some $g_i \in C(\mathcal{X})$ such that $\hat{g}_i^n \xrightarrow{a.s.} g_i$, and

(2) $f_j \in \{g_1, g_2, \ldots, g_k\}$ for each true affine component f_j in (**PL**).

Module II:

(a) Compute $\{\hat{E}_j^n\}_{j \in \hat{J}^n} : \Omega \to \Psi_k$ such that for each $\omega \in \Omega$ we have

$$\sum_{t=1}^{n} \Big[Y_t(\omega) - \bigvee_{j \in \hat{J}^n(\omega)} \bigwedge_{h \in \hat{E}_j^n(\omega)} \hat{g}_h^n(\omega, \mathbf{X}_t(\omega)) \Big]^2 = \min \Big\{ \sum_{t=1}^{n} \Big[Y_t(\omega) - \bigvee_{j \in J} \bigwedge_{h \in E_j} \hat{g}_h^n(\omega, \mathbf{X}_t(\omega)) \Big]^2 : \ \{E_j\}_{j \in J} \in \Psi_k \Big\} .$$

(b) Let $\hat{f}^n : \Omega \to C(\mathcal{X})$ be the function

$$\hat{f}^n = \bigvee_{j \in \hat{J}^n} \bigwedge_{h \in \hat{E}_j^n} \hat{g}_h^n .$$

Module II′: If we know an upper bound q (satisfying $p \leq q \leq k$) of the number p of affine components of $\mathbb{E}(Y_t|\mathbf{X}_t)$, then (a) of **Module II** can be replaced with the following.

(a) Compute $\{\hat{E}_j^n\}_{j \in \hat{J}^n} : \Omega \to \Psi_{qk}$ such that for each ω we have

$$\sum_{t=1}^{n} \Big[Y_t(\omega) - \bigvee_{j \in \hat{J}^n(\omega)} \bigwedge_{h \in \hat{E}_j^n(\omega)} \hat{g}_h^n(\omega, \mathbf{X}_t(\omega)) \Big]^2 = \min \Big\{ \sum_{t=1}^{n} \Big[Y_t(\omega) - \bigvee_{j \in J} \bigwedge_{h \in E_j} \hat{g}_h^n(\omega, \mathbf{X}_t(\omega)) \Big]^2 : \ \{E_j\}_{j \in J} \in \Psi_{qk} \Big\} .$$

TABLE 1. The RIESZVAR(i) algorithm.

C_ℓ of the grid, use OLS to estimate the piecewise linear regression of Y_t on $\mathbf{X}_t$ using only those observations for which $\mathbf{X}_t \in C_\ell$. This gives the estimated function

$$\hat{g}_\ell(\mathbf{X}_t) = \hat{b}_{0,\ell} + \hat{b}_{1,\ell} X_{1,t} + \cdots + \hat{b}_{m,\ell} X_{m,t} .$$

Obviously the true regions where $\mathbb{E}(Y_y|\mathbf{X}_t)$ is affine are unknown so some arbitrary cell size must be chosen to commence estimation. The basic trade-off in the choice of the cell size is that it is desirable that it be small enough so each affine component can be estimated, but not too small so that some cells contain few observations resulting in imprecise regression estimates. The number of observations in each cell will go to infinity as $n \to \infty$. Thus, by the strong consistency of OLS for i.i.d. observations and the compactness of the support of $\mathbf{X}_t$, the functions $\{\hat{g}_\ell\}$ converge almost surely and uniformly to functions $\{g_\ell\}$. It then follows that, for small enough mesh size ϵ, each true f_j is equal to at least one of the affine functions g_ℓ. ■

Example 7.2. Under the assumptions of the previous example, consider an alternative grid G of $[0,1]^m$ consisting of cells C_ℓ, for $\ell = 1, \dots, k$, defined so that each cell contains at least $\lfloor \frac{k_n}{n} \rfloor$ observations. Compared to the previous example, this approach guarantees that each cell contains a specified number of observations, allowing some control over the precision of the regression estimates. However, this approach allows no control over the actual locations of the cells in $[0,1]^m$. The two approaches are equivalent when $\mathbf{X}_t$ is uniformly distributed on $[0,1]^m$ but can produce different results for other distributions. A possible extension is to consider $k = k_n$ where $k_n \to \infty$ and $\frac{k_n}{n} \to 0$ as $n \to \infty$, although this is not included in our consistency results below. ■

To prove the consistency of the `RIESZVAR(i-ii)` estimators we shall need the following uniform law of large numbers.

Lemma 7.3 (Jennrich (1969)). *Let $\{\mathbf{Z}_t\}$ be a sequence of i.i.d random k-vectors, let $\mathcal{C}$ be a compact subset of $\mathbb{R}^\ell$, and let $g\colon \mathbb{R}^k \times \mathcal{C} \to \mathbb{R}$ be a Carathéodory function, i.e.,*

1. *$g(\cdot, c)$ is Borel-measurable for each $c \in \mathcal{C}$, and*
2. *$g(x, \cdot)$ is continuous for each $x \in \mathbb{R}^k$.*

If $\mathbb{E}\big(\sup_{c\in\mathcal{C}} |g(\mathbf{Z}_t, c)|\big) = \int_\Omega \sup_{c\in\mathcal{C}} |g(\mathbf{Z}_t(\omega), c)| d\pi(\omega) < \infty$, then for π-almost all ω we have

$$\lim_{n\to\infty} \Big(\sup_{c\in\mathcal{C}} \Big| \frac{1}{n} \sum_{t=1}^{n} g(\mathbf{Z}_t(\omega), c) - \mathbb{E}\big(g(\mathbf{Z}_t, c)\big) \Big| \Big) = 0\,.$$

The next theorem shows that $\hat{Y}^n$ defined in **Module II** of `RIESZVAR(i)` provides a consistent estimator of $\mathbb{E}(Y_t|\mathbf{X}_t)$ as in Definition 6.4.

Theorem 7.4. *Assume that $\mathbb{E}(Y_t|\mathbf{X}_t)$ has the form* (**PL**) *as in Definition* 6.3 *and*

$$\mathbb{E}(Y_t|\mathbf{X}_t) = \bigvee_{j\in J^0} \bigwedge_{h\in E_j^0} g_h\,.$$

If the sequence of estimators

$$\big(\{\hat{E}_j^n\}_{j\in \hat{J}^n}, (\hat{g}_1^n, \hat{g}_2^n, \dots, \hat{g}_k^n)\big)$$

is from `RIESZVAR(i)`, *then*

$$\bigvee_{j\in \hat{J}^n} \bigwedge_{h\in \hat{E}_j^n} \hat{g}_h^n \xrightarrow{a.s.} \bigvee_{j\in J^0} \bigwedge_{h\in E_j^0} g_h\,.$$

Moreover, for π-almost all $\omega \in \Omega$, there exists n_ω^* such that $n \geq n_\omega^*$ implies

$$\mathbb{E}(Y_t|\mathbf{X}_t) = \bigvee_{j\in \hat{J}^n(\omega)} \bigwedge_{h\in E_j^n(\omega)} g_h \circ \mathbf{X}_t\,.$$

In summary, `RIESZVAR(i)` *consistently estimates* (**PL**).

Proof. Let Λ be the non-empty subset of Ψ_k defined by

$$\Lambda = \left\{ \{E_j\}_J \in \Psi_k : \ \mathbb{E}(Y_t|\mathbf{X}_t) = \bigvee_{j\in J} \bigwedge_{h\in E_j} g_h \circ \mathbf{X}_t \right\}.$$

We show that for π-almost all $\omega \in \Omega$, there exists n^* such that if $n \geq n^*$, then $\{\hat{E}_j^n\}_{j\in \hat{J}^n}(\omega) \in \Lambda$. We know from Lemma 7.3 that for π-almost all $\omega \in \Omega$ and each $\{E_j\}_J \in \Psi_k$ we have

$$\lim_{n\to\infty} \tfrac{1}{n} \sum_{t=1}^{n} \Big[Y_t(\omega) - \bigvee_{j\in J} \bigwedge_{h\in E_j} g_h(\mathbf{X}_t(\omega)) \Big]^2 = \Big\| Y_t - \bigvee_{j\in J} \bigwedge_{h\in E_j} g_h \circ \mathbf{X} \Big\|^2.$$

Now by assumption for π-almost all ω we have $\hat{g}_h^n(\omega)$ converges to g_h in $C(\mathcal{X})$. Therefore, for π-almost all ω there exists n^* large enough such that if $\{E_j\}_J \in \Lambda$ and $\{E'_j\}_{J'} \notin \Lambda$, then $n \geq n^*$ implies

$$\sum_{t=1}^{n} \Big[Y_t(\omega) - \bigvee_{j\in J} \bigwedge_{h\in E_j} \hat{g}_h^n(\omega, \mathbf{X}_t(\omega)) \Big]^2 < \sum_{t=1}^{n} \Big[Y_t(\omega) - \bigvee_{j\in J'} \bigwedge_{h\in E'_j} \hat{g}_h^n(\omega, \mathbf{X}_t(\omega)) \Big]^2.$$

This of course implies that if $n \geq n^*$, then $\{\hat{E}_j^n\}_{j\in \hat{J}^n}(\omega) \in \Lambda$.

Notice now that if $\{E_j\}_J, \{E'_j\}_{J'} \in \Lambda$, then

$$\bigvee_{j\in J} \bigwedge_{h\in E_j} g_h(x) = \bigvee_{j\in J'} \bigwedge_{h\in E'_j} g_h(x)$$

for all $x \in \mathcal{X}$. That is, the two functions in $C(\mathcal{X})$ are identical. This is so since otherwise the two functions differ on an open subset of $\mathcal{X}$, and since $\mathcal{X}$ is the support of $\mathbf{X}_t$ this open set is not $\pi_{\mathbf{X}}$-null. In view of this, the fact that Ψ_k is finite, and that $C(\mathcal{X})$ is a Banach lattice, we see from the last paragraph that $\hat{f}^n \xrightarrow{a.s.} \bigvee_{j\in J} \bigwedge_{h\in E_j} g_h$ for each $\{E_j\}_J \in \Lambda$, and the proof is finished. ■

We move to two computer generated examples of `RIESZVAR(i)`.

Example 7.5. In the first example we have

$$Y_t = \bigwedge_{k=1}^{2} f_k\left(X_{1,t}, X_{2,t}\right) + \varepsilon_t \tag{7.1}$$

where $\varepsilon_t \sim N(0,1)$, $X_{1,t}, X_{2,t}$ are independent and uniformly distributed on $[0,12]^2$, and

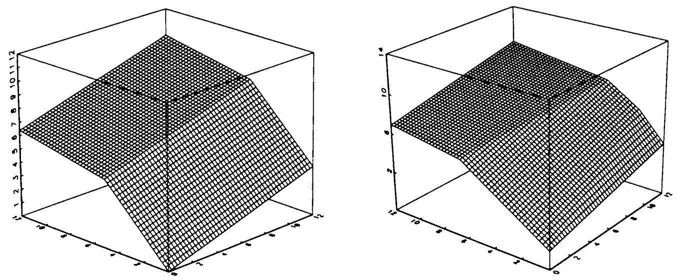
$$f_1\left(X_{1,t}, X_{2,t}\right) = 4 + 0.2X_{1,t} + 0.3X_{2,t}, \quad f_2\left(X_{1,t}, X_{2,t}\right) = X_{1,t} + 0.3X_{2,t}.$$

A plot of $\bigwedge_{k=1}^{2} f_k\left(X_{1,t}, X_{2,t}\right)$ is given in Figure 9. This example exhibits what is usually referred to as a structural break or threshold in the regression function, specifically

$$\mathbb{E}(Y_t|X_t) = \begin{cases} 4 + 0.2X_{1,t} + 0.3X_{2,t} & \text{if } X_{1,t} > 3.2, \\ X_{1,t} + 0.3X_{2,t} & \text{if } X_{1,t} \leq 3.2\,. \end{cases}$$

This shows that a continuous threshold regression with a single threshold is a very simple example of the general model (**PL**).

We generate a sample of $n = 160$ observations from (7.1), denoted $\{y_t, x_{1,t}, x_{2,t}\}_{t=1}^{n}$ and implement `RIESZVAR(i)` with **Module II′**. We choose $k = 16$ in Module I and, following Example 7.2,



$f_1 \wedge f_2$ $\hat{f} = \hat{f}_1 \wedge \hat{f}_2 \wedge \hat{f}_3$

FIGURE 9. The graph of the functions in Example 7.5.

divide the space $[0,1]^2$ into 16 regions each containing 10 observations. OLS regressions are computed for each of these 16 regions. We consider a maximum of $q = 3$ functions to be included in the final piecewise linear estimate and implement the least squares search specified in **Module II$'$**. The fitted piecewise linear regression plotted in Figure 9 can be seen to resemble the graph of the true function. The point estimates are

$$\hat{f}_1(x_{1,t}, x_{2,t}) = 4.39 + 0.21x_{1,t} + 0.27x_{2,t},$$

$$\hat{f}_2(x_{1,t}, x_{2,t}) = -0.16 + 0.91x_{1,t} + 0.40x_{2,t},$$

$$\hat{f}_3(x_{1,t}, x_{2,t}) = -0.14 + 1.57x_{1,t} + 0.29x_{2,t},$$

and

$$\hat{f}(x_{1,t}, x_{2,t}) = \hat{f}_1(x_{1,t}, x_{2,t}) \wedge \hat{f}_2(x_{1,t}, x_{2,t}) \wedge \hat{f}_3(x_{1,t}, x_{2,t}).$$

As would be expected, **Module II$'$** selects a continuous piecewise linear estimate with $q = 3$ functions, even though the truth is $p = 2$. Clearly, model selection methods to consistently estimate p are an important area for further research. It can be seen that $\hat{f}_1$ provides an estimate of f_1 and both $\hat{f}_2$ and $\hat{f}_3$ provide estimates of f_2. The imprecision is due to the fact that these coefficient estimates are computed from subsamples of just 10 observations each, which is a disadvantage of this sample splitting approach. ■

Example 7.6. Now consider (**PL**) in which

$$f(X_{1,t}, X_{2,t}) = [f_1(X_{1,t}, X_{2,t}) \wedge f_2(X_{1,t}, X_{2,t})] \vee f_3(X_{1,t}, X_{2,t})$$

and

$$\begin{aligned} f_1(X_{1,t}, X_{2,t}) &= -5 + X_{1,t} + X_{2,t}, \\ f_2(X_{1,t}, X_{2,t}) &= 12 - X_{1,t} - X_{2,t}, \\ f_3(X_{1,t}, X_{2,t}) &= -5 + X_{2,t}. \end{aligned}$$

A plot of this function is given in Figure 10. This function implies considerable changes in the behavior of $\mathbb{E}(Y_t|\mathbf{X}_t)$ over the support of $\mathbf{X}_t$.

Data generation and estimation are carried out as in the previous example. Figure 10 shows the estimated piecewise linear regression for one of the samples. The point estimates are

$$\begin{aligned}\hat{f}_1\left(x_{1,t}, x_{2,t}\right) &= 2.89 + 0.70x_{1,t} + 0.60x_{2,t}, \\ \hat{f}_2\left(x_{1,t}, x_{2,t}\right) &= 10.00 - 0.90x_{1,t} - 0.64x_{2,t}, \\ \hat{f}_3\left(x_{1,t}, x_{2,t}\right) &= -6.46 + 0.11x_{1,t} + 1.11x_{2,t},\end{aligned}$$

and

$$\hat{f}\left(x_{1,t}, x_{2,t}\right) = \left[\hat{f}_1\left(x_{1,t}, x_{2,t}\right) \wedge \hat{f}_2\left(x_{1,t}, x_{2,t}\right)\right] \vee \hat{f}_3\left(x_{1,t}, x_{2,t}\right).$$

The number of functions is correctly specified before estimation, i.e., $q = p$. For this sample, the sup-inf arrangement is correctly estimated and the point estimates are suggestive of the true parameters.

For another sample, an interesting outcome is also given in Figure 10. In this case the point estimates are

$$\begin{aligned}\hat{f}_1\left(x_{1,t}, x_{2,t}\right) &= -6.52 + 1.31x_{1,t} + 0.97x_{2,t}, \\ \hat{f}_2\left(x_{1,t}, x_{2,t}\right) &= -3.88 - 0.065x_{1,t} - 1.11x_{2,t}, \\ \hat{f}_3\left(x_{1,t}, x_{2,t}\right) &= 10.00 - 0.90x_{1,t} - 0.64x_{2,t},\end{aligned}$$

and $\hat{f} = (\hat{f}_1 \wedge \hat{f}_2) \vee (\hat{f}_2 \wedge \hat{f}_3) \vee (\hat{f}_1 \wedge \hat{f}_3)$. This case highlights the finite identification problem in (**PL**). The consequence of this alternative arrangement of sup-inf is shown graphically in Figure 10. It shows that having 3 affine functions does not necessarily mean that the resulting piecewise linear regression function will have 3 connected regions—complex behavior is possible by taking sup-inf operations over three affine functions. ■

8. `RIESZVAR(ii)`

We consider again the piecewise linear regression model (**PL**) as described in Definition 6.3. One of the basic problems of `RIESZVAR(i)` is that we need to divide the data into small regions of $\mathcal{X}$, the common support of the measures $\pi_{\mathbf{X}_t}$. This results in efficiency losses. In this section, we define a non-linear estimator of (**PL**) that avoids the division of data into regions. The idea is to estimate the sup-inf representation and the affine components in a single shot. It turns out that this can be done by minimizing a piecewise quadratic function of the data.

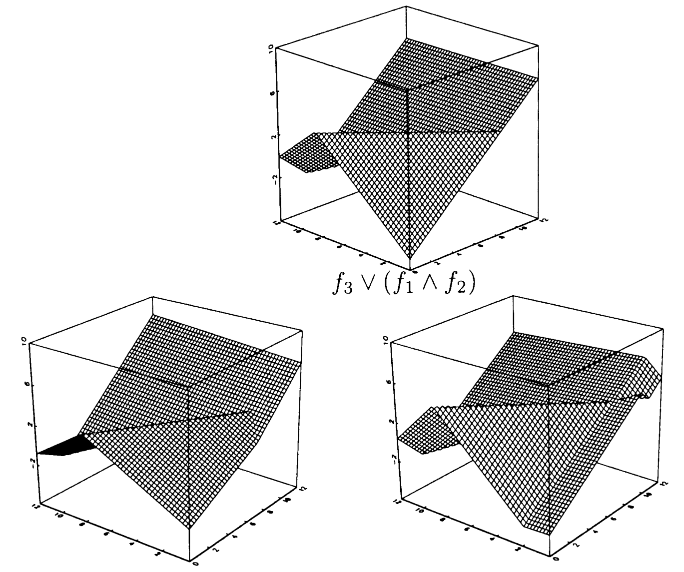
To this end, notice that every vector $\mathbf{r} = (r_0, r_1, \ldots, r_m) \in \mathbb{R}^{m+1}$ generates an affine function via the formula

$$\mathbf{r}(x) = r_0 + r_1 x_1 + \cdots + r_m x_m\,.$$

To avoid introducing extra notation, we shall view $\mathbf{r}$ both as a vector in $\mathbb{R}^{m+1}$ and as an affine function in $\mathbf{Aff} \subseteq C(\mathcal{X})$.

Now fix two numbers p, q and let $\Phi_{pq} = \mathbb{R}^{(m+1)pq}$. An arbitrary element R of Φ_{pq} is denoted by a matrix of the form

$$R = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} & \cdots & \mathbf{r}_{1p} \\ \mathbf{r}_{21} & \mathbf{r}_{22} & \cdots & \mathbf{r}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{r}_{q1} & \mathbf{r}_{p2} & \cdots & \mathbf{r}_{qp} \end{bmatrix},$$



$\hat{f}_3 \vee (\hat{f}_1 \wedge \hat{f}_2)$ $(\hat{f}_1 \wedge \hat{f}_2) \vee (\hat{f}_2 \wedge \hat{f}_3) \vee (\hat{f}_1 \wedge \hat{f}_3)$

FIGURE 10. The graph of the functions in Example 7.6.

where the entries $\mathbf{r}_{ij} = (r_{ij}^0, r_{ij}^1, \dots, r_{ij}^m) \in \mathbb{R}^{m+1}$. For any $\mathbf{x} \in \mathcal{X}$ (and as a matter of fact for any $\mathbf{x} \in \mathbb{R}^m$) let

$$R \diamond \mathbf{x} = \Big[\bigvee_{i=1}^{q} \bigwedge_{j=1}^{p} \mathbf{r}_{ij}\Big](x) = \bigvee_{i=1}^{q} \bigwedge_{j=1}^{p} (r_{ij}^0 + r_{ij}^1 x_1 + \cdots + r_{ij}^m x_m)\,.$$

Therefore, $R\diamond$ defines a function from $\mathcal{X}$ to $\mathbb{R}$ (which is also the restriction of $R\diamond\colon \mathbb{R}^m \to \mathbb{R}$ to $\mathcal{X}$). Notice that $R\diamond \in \mathbf{Aff}^{\vee\wedge} \subseteq C(\mathcal{X})$.

Next, for the random vector $\mathbf{X}_t$ write

$$R \diamond \mathbf{X}_t = R \diamond (\mathbf{X}_t) = \bigvee_{i=1}^{q} \bigwedge_{j=1}^{p} (r_{ij}^0 + r_{ij}^1 X_{1,t} + \cdots + r_{ij}^m X_{m,t})\,.$$

We have the following simple lemma.

Lemma 8.1. *Regarding the function $\diamond$ the following hold true:*

(1) *The function $R \mapsto R\diamond$, from Φ_{pq} to $C(\mathcal{X})$, continuous.*
(2) *The function $R \mapsto R \diamond \mathbf{X}_t$, from Φ_{pq} to $L_2(\pi)$, is continuous.*
(3) *If $\{\mathbf{x}_t\}_{t=1}^n$ is a finite collection of points in $\mathbb{R}^m$, then the function*

$$R \mapsto (R \diamond \mathbf{x}_1, R \diamond \mathbf{x}_2, \dots, R \diamond \mathbf{x}_n)\,,$$

RIESZVAR(ii) Consider the estimation of (**PL**) assuming that we know of a compact subset Θ of Φ_{qp} satisfying $$\{R \in \Theta\colon\ \mathbb{E}(Y_t\vert\mathbf{X}_t) = R \diamond \mathbf{X}_t\} \neq \emptyset\,.$$ Using $\{Y_t, \mathbf{X}_t\}_{t=1}^n$: ($\star$) Choose $\hat{R}^n\colon \Omega \to \Theta$ such that $\hat{R}^n(\omega)$ is a solution to the following minimization problem: $$\min\Big\{\sum_{t=1}^n \big[R \diamond \mathbf{X}_t(\omega) - Y_t(\omega)\big]^2\colon\ R \in \Theta\Big\}\,.$$

TABLE 2. The RIESZVAR(ii) algorithm

from Φ_{qp} to $\mathbb{R}^n$, is positively homogeneous and piecewise linear.

The next result is an easy consequence of Theorem 4.14. It tells us that we can express the estimation problem (**PL**) in terms of the functions of the form $R\diamond$.

Lemma 8.2. *If $\mathbb{E}(Y_t\vert\mathbf{X}_t) = f \circ \mathbf{X}_t$ where f is piecewise linear with at most q components, then there exists $R \in \Phi_{q(\#q)}$ such that $\mathbb{E}(Y_t\vert\mathbf{X}_t) = R \diamond \mathbf{X}_t$.*

Table 2 describes our second algorithm for estimating piecewise linear functions from data. We are ready to prove the main result of the section.

Theorem 8.3. *If $\mathbb{E}(Y_t\vert\mathbf{X}_t) = R^0 \diamond \mathbf{X}_t$, then for $\hat{R}_n$ in* RIESZVAR(ii) *we have*

$$\hat{R}_n \diamond \xrightarrow{a.s.} R^0 \diamond\,.$$

That is, RIESZVAR(ii) *consistently estimates* (**PL**).

Proof. Notice that all conditions of Lemma 7.3 are satisfied for $g\colon \mathbb{R} \times \mathbb{R}^m \times \Theta \to \mathbb{R}$ defined by

$$g(y, \mathbf{x}, R) = (y - R \diamond \mathbf{x})^2\,.$$

This implies that almost surely

$$\lim_{n\to\infty} \sup_{R\in\Theta} \Big\vert \frac{1}{n} \sum_{t=1}^n \big[Y_t(\omega) - R \diamond \mathbf{X}_t(\omega)\big]^2 - \Vert Y_t - R \diamond \mathbf{X}_t \Vert^2 \Big\vert = 0\,.$$

Clearly, therefore, for π-almost all $\omega \in \Omega$ we have

$$\lim_{n\to\infty} \big\Vert Y_t - \hat{R}^n(\omega) \diamond \mathbf{X}_t \big\Vert = \big\Vert Y_t - \mathbb{E}(Y_t\vert\mathbf{X}_t) \big\Vert\,.$$

This (in connection with Lemma 2.9) yields $\lim_{n\to\infty} \big\Vert \hat{R}^n(\omega) \diamond \mathbf{X}_t - \mathbb{E}(Y_t\vert\mathbf{X}_t) \big\Vert = 0$ for π-almost all $\omega \in \Omega$.

Now let

$$\Lambda = \left\{R \in \Theta\colon\ \mathbb{E}(Y_t|\mathbf{X}_t) = R \diamond \mathbf{X}_t\right\}.$$

Since Θ is compact and $R \mapsto R \diamond \mathbf{X}_t$, from Φ_{pq} to $L_2(\pi)$, is continuous it must be the case that for π-almost all ω, the cluster points of $\hat{R}^n(\omega)$ are in Λ. Now the continuous function $R \mapsto R\diamond$, from Φ_{pq} to $C(\mathcal{X})$, maps Λ to a single point $R^0\diamond$ of $C(\mathcal{X})$. Therefore, it must be the case that $\hat{R}^n \diamond \xrightarrow{a.s.} R^0\diamond$, and the proof is finished. ■

We now give a computer generated example of the application of `RIESZVAR(ii)`.

Example 8.4. Suppose the data generating process is (**PL**) where

$$f(X_{1,t}, X_{2,t}) = \left[f_1(X_{1,t}, X_{2,t}) \wedge f_2(X_{1,t}, X_{2,t})\right] \vee \left[f_3(X_{1,t}, X_{2,t}) \wedge f_4(X_{1,t}, X_{2,t})\right]$$

and

$$\begin{aligned} f_1(X_{1,t}, X_{2,t}) &= -5 + X_{1,t} + X_{2,t}, \\ f_2(X_{1,t}, X_{2,t}) &= 12 - X_{1,t} - X_{2,t}, \\ f_3(X_{1,t}, X_{2,t}) &= -5 + X_{2,t}, \\ f_4(X_{1,t}, X_{2,t}) &= 2. \end{aligned}$$

A plot of this function is given in Figure 11. The matrix R that defines the function $R \diamond \mathbf{X}_t$ for this example is

$$R = \begin{bmatrix} \begin{pmatrix} -5 \\ 1 \\ 1 \end{pmatrix} & \begin{pmatrix} 12 \\ -1 \\ -1 \end{pmatrix} \\ \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix} & \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} \end{bmatrix}$$

As in Example 7.5, we generate $n = 160$ observations $x_{1,t}$ and $x_{2,t}$ as being uniformly distributed on $[0, 12]^2$ and generate $\varepsilon_t = Y_t - \mathbb{E}(Y_t|\mathbf{X}_t)$ from a standard normal distribution.

We implement `RIESZVAR(ii)` using $q = p = 2$, which are the smallest values of p and q that include our data generating process. The numerical minimization with respect to R specified by `RIESZVAR(ii)` is implemented using the BFGS algorithm in the "optimum" library of Gauss v6.0.

For the first sample we report, the point estimates are given by

$$\hat{R}^n = \begin{bmatrix} \begin{pmatrix} -4.42 \\ 1.11 \\ 1.02 \end{pmatrix} & \begin{pmatrix} 11.07 \\ -0.84 \\ -0.82 \end{pmatrix} \\ \begin{pmatrix} -5.32 \\ 0.02 \\ 1.13 \end{pmatrix} & \begin{pmatrix} 2.48 \\ 0.04 \\ 0.02 \end{pmatrix} \end{bmatrix},$$

and Figure 11 shows a plot of the resulting estimate (Estimation 1) of $\mathbb{E}(Y_t|\mathbf{X}_t)$ given by $\hat{R}^n \diamond \mathbf{x}$ for $\mathbf{x} \in [0,12]^2$. Clearly both the point estimates and the resulting estimate of $\mathbb{E}(Y_t|\mathbf{X}_t)$ strongly resemble the true values.

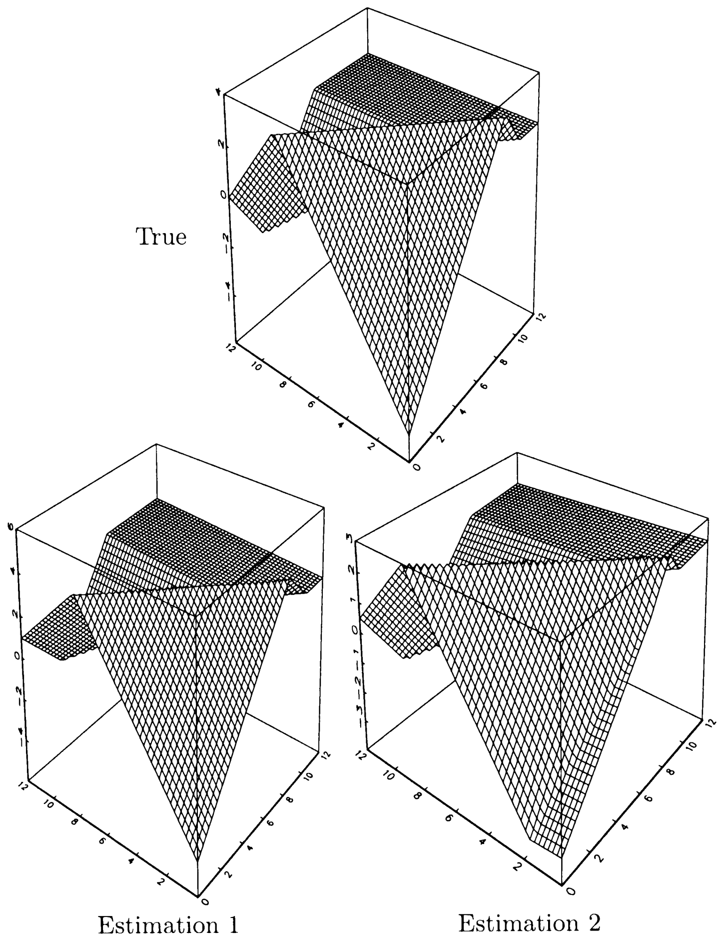


FIGURE 11. The graph of the functions in Example 8.4.

A plot (Estimation 2) of the estimates of $\mathbb{E}(Y_t|\mathbf{X}_t)$ from additional replications are given in Figure 11. This plot show qualitatively that $\mathbb{E}(Y_t|\mathbf{X}_t)$ is generally estimated very well globally, with just small local deviations due to the usual sampling variation. ■

REFERENCES

Y. A. Abramovich and C. D. Aliprantis. *An invitation to operator theory*, volume 50 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002a.

Y. A. Abramovich and C. D. Aliprantis. *Problems in operator theory*, volume 51 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002b.

G. G. Agarwal and W. J. Studden. Asymptotic integrated mean square error using least squares and bias minimizing splines. *Ann. Statist.*, 8(6):1307–1325, 1980.

C. D. Aliprantis and K. C. Border. *Infinite-dimensional analysis*. Springer-Verlag, Berlin, second edition, 1999. A hitchhiker's guide.

C. D. Aliprantis, D. J. Brown, and O. Burkinshaw. *Existence and optimality of competitive equilibria*. Springer-Verlag, Berlin, 1990.

C. D. Aliprantis, D. J. Brown, I. A. Polyrakis, and J. Werner. Portfolio dominance and optimality in infinite security markets. *Journal of Mathematical Economics*, 30(3):347–366, 1998.

C. D. Aliprantis, D. J. Brown, and J. Werner. Minimum-cost portfolio insurance. *J. Econom. Dynam. Control*, 24(11-12):1703–1719, 2000. Computational aspects of complex securities.

C. D. Aliprantis and O. Burkinshaw. *Locally solid Riesz spaces with applications to economics*, volume 105 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, second edition, 2003.

C. D. Aliprantis, B. Cornet, and R. Tourky. Economic equilibrium: optimality and price decentralization. *Positivity*, 6(3):205–241, 2002a. Special issue of the mathematical economics.

C. D. Aliprantis, Y. A. Polyrakis, and R. Tourky. The cheapest hedge. *J. Math. Econom.*, 37(4): 269–295, 2002b.

A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. M. Ziegler. *Oriented matroids*, volume 46 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 1999.

I. M. Bomze. *A functional analytic approach to statistical experiments*, volume 237 of *Pitman Research Notes in Mathematics Series*. Longman Scientific & Technical, Harlow, 1990.

D. J. Brown, C. B. Huijsmans, and B. de Pagter. Approximating derivative securities in f-algebras. In *Positive operators, Riesz spaces, and economics (Pasadena, CA, 1990)*, pages 171–177. Springer, Berlin, 1991.

D. J. Brown and S. A. Ross. Spanning, valuation and options. *Econom. Theory*, 1(1):3–12, 1991.

R. W. Cottle, J.-S. Pang, and R. E. Stone. *The linear complementarity problem*. Computer Science and Scientific Computing. Academic Press Inc., Boston, MA, 1992.

G. B. Dantzig and A. S. Manne. A complementary algorithm for optimal capital path with invariant proportions. *Journal of Economic Theory*, 9:312–323, 1974.

B. C. Eaves and C. E. Lemke. Equivalence of LCP and PLS. *Math. Oper. Res.*, 6(4):475–484, 1981.

B. C. Eaves and C. E. Lemke. On the equivalence of the linear complementarity problem and a system of piecewise linear equations. II. In *Homotopy methods and global convergence (Porto Cervo, 1981)*, volume 13 of *NATO Conf. Ser. II: Systems Sci.*, pages 79–90. Plenum, New York, 1983.

B. C. Eaves and H. Scarf. The solution of systems of piecewise linear equations. *Math. Oper. Res.*, 1(1):1–27, 1976.

J. H. Friedman. Multivariate adaptive regression splines. *Ann. Statist.*, 19(1):1–141, 1991.

J. Gonzalo and J. Y. Pitarakis. Estimation and model selection based inference in single and multiple threshold models. *J. Econometrics*, 110(2):319–352, 2002. Long memory and nonlinear time series (Cardiff, 2000).

R. E. Groff. *Piecewise Linear Homeomorphisms for Approximation of Invertible Maps*. PhD thesis, The University of Michigan, 2003.

R. E. Groff, P. P. Khargonekar, and D. E. Koditschek. A local convergence proof for the minvar algorithm for computing continuous piecewise linear approximations. *SIAM J. Numer. Anal.*, 41(3):983–1007 (electronic), 2003.

B. E. Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–430, 1996.

B. E. Hansen. Threshold effects in non-dynamic panels: estimation, testing, and inference. *J. Econometrics*, 93(2):345–368, 1999.

B. E. Hansen and B. Seo. Testing for two-regime threshold cointegration in vector error-correction models. *J. Econometrics*, 110(2):293–318, 2002. Long memory and nonlinear time series (Cardiff, 2000).

R. I. Jennrich. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.*, 40:633–643, 1969.

L. Le Cam. *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986.

C. E. Lemke and J. T. Howson. Equilibrium points of bimatrix games. *J. Soc. Indust. Appl. Math.*, 12:413–423, 1964.

W. A. J. Luxemburg and A. C. Zaanen. *Riesz spaces. Vol. I.* North-Holland Publishing Co., Amsterdam, 1971.

A. Mas-Colell and W. R. Zame. Equilibrium theory in infinite-dimensional spaces. In *Handbook of mathematical economics, Vol. IV*, volume 1 of *Handbooks in Econom.*, pages 1835–1898. North-Holland, Amsterdam, 1991.

S. Ovchinnikov. Max-min representation of piecewise linear functions. *Beiträge Algebra Geom.*, 43 (1):297–302, 2002.

S. A. Ross. Options and efficiency. *Quarterly Journal of Economics*, 90:75–89, 1976.

P. A. Samuelson. Spatial price equilibrium and linear programming. *American Economic Review*, 42:283–303, 1952.

H. E. Scarf. The approximation of fixed points of a continuous mapping. *SIAM J. Appl. Math.*, 15:1328–1343, 1967.

H. H. Schaefer. *Banach lattices and positive operators.* Springer-Verlag, New York, 1974. Die Grundlehren der mathematischen Wissenschaften, Band 215.

L. L. Schumaker. On spaces of piecewise polynomials in two variables. In *Approximation theory and spline functions (St. John's, Nfld., 1983)*, volume 136 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 151–197. Reidel, Dordrecht, 1984.

H. Tong. *Threshold models in nonlinear time series analysis*, volume 21 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1983.

R. Wilson. The bilinear complementarity problem and competitive equilibria of piecewise linear economic models. *Econometrica*, 46(1):87–103, 1978.

S. Zhou, X. Shen, and D. A. Wolfe. Local asymptotics for regression splines and confidence regions. *Ann. Statist.*, 26(5):1760–1782, 1998.

C. D. ALIPRANTIS, DEPARTMENT OF ECONOMICS, KRANNERT SCHOOL OF MANAGEMENT, PURDUE UNIVERSITY, W. LAFAYETTE IN 47907-1310, USA
E-mail address: aliprantis@mgmt.purdue.edu

D. HARRIS, DEPARTMENT OF ECONOMICS, UNIVERSITY OF MELBOURNE, MELBOURNE, VICTORIA 3010, AUSTRALIA
E-mail address: harrisd@unimelb.edu.au

R. TOURKY, DEPARTMENT OF ECONOMICS, UNIVERSITY OF MELBOURNE, MELBOURNE, VICTORIA 3010, AUSTRALIA; AND DEPARTMENT OF ECONOMICS, KRANNERT SCHOOL OF MANAGEMENT, PURDUE UNIVERSITY, W. LAFAYETTE IN 47907-1310, USA
E-mail address: rtourky@unimelb.edu.au