Linear SEL Prediction Using Basis Function Transformations: *p=1* Wavelet Bases

Stephen Vardeman
Analytics Iowa LLC
ISU Statistics and IMSE

Linear prediction using basis transformations

We consider some set of (basis) functions $\{h_m\}$ and predictors of the form or depending upon the form

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^{p} \hat{\beta}_m h_m(\mathbf{x}) = \mathbf{h}(\mathbf{x})' \hat{\beta} \tag{1}$$

for $\mathbf{h}(\mathbf{x})' = (h_1(\mathbf{x}), \ldots, h_p(\mathbf{x}))$. $\mathbf{h}(\mathbf{x})$ is being used here in place of the general transformation notation $T(\mathbf{x})$ in honor of the slight specialization to cases where the p component functions of $\mathbf{h}(\mathbf{x})$ are "basis" functions for a function space, whereby any function of interest can be approximated as a linear combination of them. Periodic functions of a single variable can be approximated by linear combinations of sine (basis) functions of various frequencies. General differentiable functions can be approximated by polynomials (linear combinations of monomial basis functions). Etc.

SEL fitting

SEL fitting of form (1) can be done using any of the linear methods just discussed based on the $N \times p$ matrix of inputs

$$\mathbf{X_h} = (h_j(\mathbf{x}_i)) = \begin{pmatrix} \mathbf{h}(\mathbf{x}_1)' \\ \mathbf{h}(\mathbf{x}_2)' \\ \vdots \\ \mathbf{h}(\mathbf{x}_N)' \end{pmatrix}$$

(i indexing rows and j indexing columns). Large sets of basis functions can give highly flexible prediction methods.

Motivation and the Haar basis

The idea here is to consider a set of basis functions rich/large enough to provide effective approximations of essentially any function on $[0, 1]$. For example, it is well known that periodic functions can be effectively represented as linear combinations of sines and cosines of various periods. Here we consider sets of "wavelet" basis functions that can be effective even in the absence of periodicities.

The simplest wavelet basis for $L_2[0, 1]$ is the Haar basis. It is built on the time-honored mathematical notion of approximating functions with ones constant over short intervals. The Haar basis consists of (orthonormal in $L_2[0, 1]$) step functions constant on natural intervals of length $1/2^m$.

Haar basis development

Define the so-called Haar **"father" wavelet**

$$\varphi(x) = I[0 < x \leq 1]$$

and the so-called Haar **"mother" wavelet**

$$\psi(x) = \varphi(2x) - \varphi(2x-1) = I\left[0 < x \leq \frac{1}{2}\right] - I\left[\frac{1}{2} < x \leq 1\right]$$

Linear combinations of these functions provide all elements of $L_2[0,1]$ that are constant on $\left(0, \frac{1}{2}\right]$ and on $\left(\frac{1}{2}, 1\right]$. Write

$$\Psi_0 = \{\varphi, \psi\}$$

Haar basis development cont.

Next, define

$$\psi_{1,0}(x) = \sqrt{2}\left(I\left[0 < x \leq \frac{1}{4}\right] - I\left[\frac{1}{4} < x \leq \frac{1}{2}\right]\right) \quad \text{and}$$

$$\psi_{1,1}(x) = \sqrt{2}\left(I\left[\frac{1}{2} < x \leq \frac{3}{4}\right] - I\left[\frac{3}{4} < x \leq 1\right]\right)$$

and let

$$\Psi_1 = \{\psi_{1,0}, \psi_{1,1}\}$$

Using the set of functions $\Psi_0 \cup \Psi_1$ one can build (as linear combinations) all elements of $L_2[0,1]$ that are constant on $\left(0, \frac{1}{4}\right]$ and on $\left(\frac{1}{4}, \frac{1}{2}\right]$ and on $\left(\frac{1}{2}, \frac{3}{4}\right]$ and on $\left(\frac{3}{4}, 1\right]$.

Haar basis development cont.

And so on. In general,

$$\psi_{m,j}(x) = \sqrt{2^m}\psi\left(2^m\left(x - \frac{j}{2^m}\right)\right) \quad \text{for } j = 0, 1, 2, \ldots, 2^m - 1$$

and

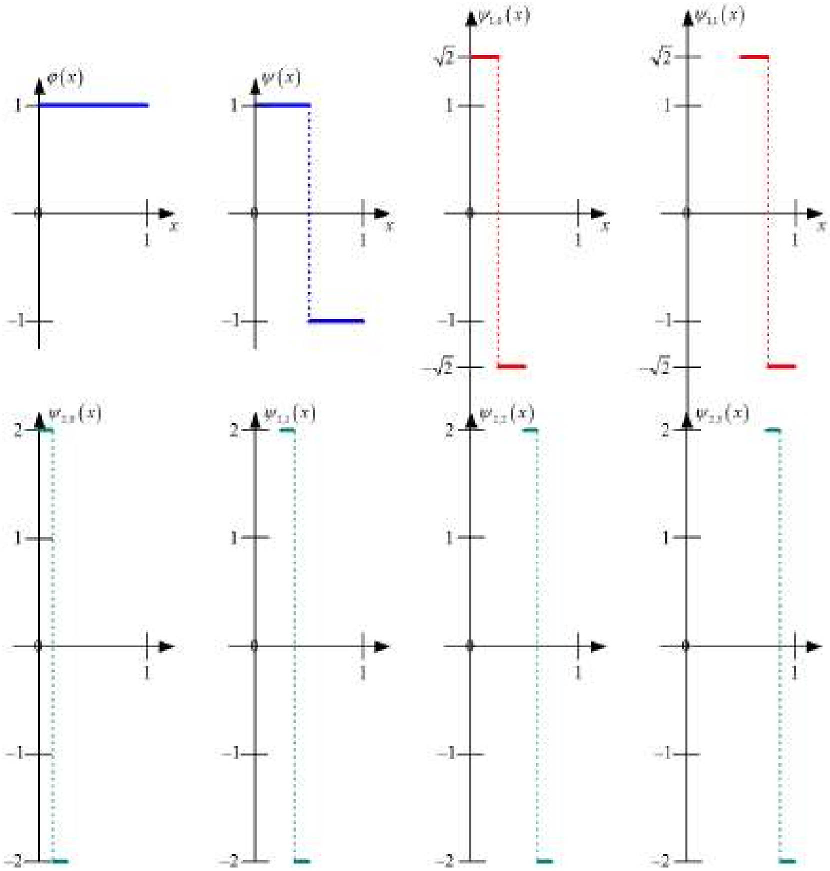
$$\Psi_m = \{\psi_{m,0}, \psi_{m,1}, \ldots, \psi_{m,2^m-1}\}$$

The Haar basis of $L_2[0,1]$ is then

$$\cup_{m=0}^{\infty}\Psi_m$$

Plots of some Haar basis functions

Below are the sets of functions Ψ_0 (blue), Ψ_1 (red), and Ψ_2 (green).



Application

Then, one can use the Haar basis functions through order M in constructing a form

$$f(x) = \beta_0 + \sum_{m=0}^{M} \sum_{j=0}^{2^m - 1} \beta_{mj} \psi_{m,j}(x) \qquad (2)$$

(with the understanding that $\psi_{0,0} = \psi$). Such a function is constant on consecutive intervals of length $1/2^{M+1}$. It can be fit by any of the linear fitting methods (especially involving thresholding/selection, as a typically very large number, 2^{M+1}, of basis functions is employed in (2)). Large absolute values of coefficients β_{mj} indicate *scales* at which important variation in response occur *through the value of the index* m. The *location* in $[0, 1]$ where that variation occurs is *encoded in the value* $j/2^m$. Where only a relatively few fitted coefficients are important, the corresponding scales and locations provide an informative and compact summary of the fit.

Application cont.

A nice visual summary of the results of the fit to form (2) can be made by plotting for each m (plots arranged vertically, from M through 0, aligned and to the same scale) spikes of length $|\beta_{mj}|$ pointed in the direction of $\text{sign}(\beta_{mj})$ along an "x" axis at positions (say) $(j/2^m) + 1/2^{m+1}$.

Where $N = 2^K$ and

$$x_i = i\left(\frac{1}{2^K}\right) \text{ for } i = 1, 2, \ldots, 2^K$$

and one uses the Haar basis functions through order $K-1$, the fitting of (2) is computationally clean, since the vectors

$$\left(\psi_{m,j}\left(x_1\right), \ldots, \psi_{m,j}\left(x_N\right)\right)'$$

(together with the column vector of 1s) are orthogonal. (Upon normalization, i.e. division by $\sqrt{N} = 2^{K/2}$, they form an orthonormal basis for $\Re^N$.)

“Symmlet” wavelets

The Haar wavelet basis functions are easy to describe and understand. But they are discontinuous, and from some points of view that is unappealing. Other sets of wavelet basis functions have been developed that are smooth. The construction begins with a smooth "mother wavelet" in place of the step function used above. HTF make some discussion of the smooth "symmlet" wavelet basis at the end of their Chapter 5.