

Data Analysis: Projection

Mark Hendricks

August Review

UChicago Financial Mathematics

Outline

Projection

Environment

- ▶ We are interested in a random scalar variable, y .
- ▶ We observe a random $k \times 1$ column vector, \mathbf{x} .
- ▶ We do not put any restriction on the joint distribution of (y, \mathbf{x}) .

Conditional expectation

- ▶ Suppose we want to approximate y having observed \mathbf{x} .
- ▶ Let $f(\mathbf{x})$ denote an arbitrary function, resulting in the error $\varepsilon = y - f(\mathbf{x})$.
- ▶ If we specify a loss function that we wish to minimize, $\ell(\varepsilon)$, then we can consider the optimal approximating function, $f^*(\mathbf{x})$, that minimizes $\ell(\varepsilon)$.

MSE

For the **mean squared error** loss function,

$$\ell(\varepsilon) = \mathbb{E} \left[(y - f(\mathbf{x}))^2 \right]$$

the **conditional expectation** is the optimal approximation,

$$f^*(\mathbf{x}) = \mathbb{E} [y \mid \mathbf{x}]$$

Generality of Cond.Exp.

In general, the conditional expectation may,

- ▶ depend on the joint distribution of (y, \mathbf{x})
- ▶ be a nonlinear function of \mathbf{x} .

Decomposition

We can decompose y into two parts, the conditional expectation and the approximation error.

$$\begin{aligned}y &= \mathbb{E}[y \mid \mathbf{x}] + \varepsilon \\ 0 &= \mathbb{E}[\varepsilon \mid \mathbf{x}]\end{aligned}$$

Note that the second statement holds by construction. Implies,

$$\mathbb{E}[\varepsilon \mid \mathbf{x}] = 0 \iff 0 = \mathbb{E}[f(\mathbf{x})\varepsilon], \forall f(\mathbf{x})$$

so

$$\mathbb{E}[\mathbf{x}\varepsilon] = 0, \quad 0 = \mathbb{E}[\varepsilon]$$

Uncorrelated functions of \mathbf{x}

Furthermore, the last condition restricts the covariance and correlation:

$$\begin{aligned}\text{cov}[f(\mathbf{x}), \varepsilon] &= 0 = \text{corr}[f(\mathbf{x}), \varepsilon], \quad \forall f \\ \text{cov}[\mathbf{x}, \varepsilon] &= 0 = \text{corr}[\mathbf{x}, \varepsilon]\end{aligned}$$

ε cannot be improved through knowledge of \mathbf{x} : uncorrelated to any (nonlinear) function of \mathbf{x} .

Linear approximation

Suppose we restrict ourselves to linear approximation. If we again use the MSE loss function, we require a vector β that minimizes,

$$\ell(\varepsilon) = \mathbb{E} \left[(y - \mathbf{x}'\beta)^2 \right]$$

This is optimized for the choice, β which satisfies

$$\mathbb{E} [\mathbf{x} (y - \mathbf{x}'\beta)] = \mathbf{0}$$

That, the optimal β ensures the approximation error is orthogonal to \mathbf{x} .

Footnote: Optimality of orthogonal β

Consider the MSE for an arbitrary β ,

$$\begin{aligned}
 \ell(\epsilon) &= \mathbb{E} \left[(y - \mathbf{x}'\beta)^2 \right] \\
 &= \mathbb{E} \left[[(y - \mathbf{x}'\beta) + \mathbf{x}'(\beta - \beta)]^2 \right] \\
 &= \mathbb{E} \left[(y - \mathbf{x}'\beta)^2 \right] + 2(\beta - \beta)' \underbrace{\mathbb{E} [\mathbf{x}(y - \mathbf{x}'\beta)]}_{0 \text{ by orthogonality of } \beta} + \mathbb{E} \left[(\mathbf{x}'(\beta - \beta))^2 \right] \\
 &= \mathbb{E} \left[(y - \mathbf{x}'\beta)^2 \right] + \mathbb{E} \left[(\mathbf{x}'(\beta - \beta))^2 \right] \\
 &> \mathbb{E} \left[(y - \mathbf{x}'\beta)^2 \right]
 \end{aligned}$$

Thus, any choice, β has MSE at least as large as β .

Identification

Existence requires the following assumption

Assumption (Identified)

The $k \times k$ matrix, $\mathbb{E}[\mathbf{x}\mathbf{x}']$, is nonsingular, (and thus finite.)

- ▶ If this assumption does not hold, then the linear factors, \mathbf{x} are not **identified**.
- ▶ This assumption is not too restrictive; if one of the k elements of \mathbf{x} is an exact linear function of the others, simply drop it from consideration.

Formula for β

Rearranging, we have

$$\beta^* = (\mathbb{E} [\mathbf{x}\mathbf{x}'])^{-1} \mathbb{E} [\mathbf{x}y] \quad (1)$$

Thus, $\mathbb{L}(y \mid \mathbf{x})$. is the optimal linear approximation of y given \mathbf{x} .

$$\mathbb{L}(y \mid \mathbf{x}) = \mathbf{x}'\beta$$

Example: Linear dependence

- ▶ Suppose in analyzing mean excess returns for S&P 500 stocks, we wanted to also condition on whether the stock is in the Information Technology (IT) sector.
- ▶ We set two indicator variables in $\mathbf{x} = \begin{bmatrix} x^{[1]} \\ x^{[2]} \end{bmatrix}$.

$$x^{[1]} = \mathbb{I}_{\text{IT sector}}, \quad x^{[2]} = \mathbb{I}_{\text{not IT sector}}$$

- ▶ But clearly, $x^{[1]}$ and $x^{[2]}$ are not linearly independent:

$$x_i^{[2]} = 1 - x_i^{[1]}, \quad \forall i$$

- ▶ We need to drop either linearly dependent variable.

Dummy variables

- ▶ This does happen when setting indicator variables for many categories, and forget to exclude one category to make it the baseline.
- ▶ For instance, we might look at how membership in each of the 11 GICS-defined sectors impacts a stock's mean excess returns.
- ▶ Then y would be mean excess return, and \mathbf{x} should have indicator variables for just 10 of the 11 sectors.
- ▶ Whichever sector we exclude is seen as the baseline, for which each of the elements of β is measuring relative impact.

Including a constant

Suppose we allow an affine approximation by including a constant as one of the elements of \mathbf{x} .

$$\mathbf{x} = \begin{bmatrix} 1 \\ \tilde{\mathbf{x}} \end{bmatrix}$$

where $\tilde{\mathbf{x}}$ are the non-constant elements of \mathbf{x} . The second moments are then equivalent to centered second moments:

$$\begin{aligned} \mathbb{E} [\mathbf{x}\varepsilon] &= \text{cov} [\mathbf{x}\varepsilon] \\ \mathbb{E} [\mathbf{x}\mathbf{x}'] &= \text{var} [\mathbf{x}] \end{aligned}$$

Projection w/ constant as demean

It can be shown that the linear projection is simply,

$$\mathbb{L}(y | \mathbf{x}) = \mathbb{E}[y] + (\tilde{\mathbf{x}} - \mathbb{E}[\tilde{\mathbf{x}}])' (\text{var}[\tilde{\mathbf{x}}])^{-1} \text{cov}[\tilde{\mathbf{x}}y]$$

which in the case of scalar $\tilde{\mathbf{x}} = x$ reduces to

$$\mathbb{L}\left[y \mid \begin{bmatrix} 1 \\ x \end{bmatrix}\right] = \mathbb{E}[y] + (x - \mathbb{E}[x]) \frac{\text{cov}[x, y]}{\text{var}[x]}$$

This makes clear that including a constant is equivalent to de-meaning all the variables and then approximating.

Projection

By definition, a projection is a linear operator, P , such that $P^2 = P$.

- ▶ $\mathbb{L}[y | \mathbf{x}]$ is decomposing y into a portion spanned by \mathbf{x} and a portion orthogonal to \mathbf{x} .
- ▶ Thus, it is formally a projection of y onto \mathbf{x} .
- ▶ Operator is a linear function of \mathbf{x} .
- ▶ β , depends only on the second moments of (y, \mathbf{x}) , not the entire joint distribution.

Projection decomposition

y into two parts,

$$y = \mathbb{L}[y \mid \mathbf{x}] + \epsilon$$

$$y = \mathbf{x}'\boldsymbol{\beta} + \epsilon$$

This is indeed a projection, and we can verify that

$$0 = \mathbb{L}[\epsilon \mid \mathbf{x}]$$

alternately stated as

$$0 = \mathbb{E}[\mathbf{x}\epsilon]$$

Contrast this with the approximation error of the conditional expectation, ϵ , which is not just orthogonal to x but has zero conditional expectation.

LP as CE

Theorem

If $\mathbb{E}[\epsilon | \mathbf{x}] = 0$, then the linear projection is the conditional expectation.

$$\mathbb{E}[y | \mathbf{x}] = \mathbb{L}[y | \mathbf{x}]$$

Equivalently, if we assume that the conditional expectation is of the form,

$$\mathbb{E}[y | \mathbf{x}] = \mathbf{x}\boldsymbol{\theta} + \varepsilon$$

then the linear projection is the conditional expectation, with

$$\boldsymbol{\beta} = \boldsymbol{\theta}$$

Nonlinearity

Example

- ▶ the value of a call option, denoted c ,
- ▶ and the value of the underlying stock, denoted s .
- ▶ Suppose call options have the following relationship to underlying stock price:

$$c = f(s) + \epsilon$$

- ▶ f is a nonlinear equation
- ▶ ϵ is the impact of transaction costs, has zero mean, and is independent of s .

Example: continued

- ▶ Then by assumption, $\mathbb{E}[c | s] = f(s)$, with $\mathbb{E}[\varepsilon | s] = 0$.
- ▶ However, the linear projection of c onto $\mathbf{x} = \begin{bmatrix} 1 \\ s \end{bmatrix}$ is

$$\begin{aligned}\mathbb{L}[c | \mathbf{x}] &= (\mathbb{E}[\mathbf{xx}'])^{-1} \mathbb{E}[\mathbf{x}c] \\ &= \mathbb{E}[c] + \frac{\text{cov}(s, c)}{\text{var}(s)} (s - \mathbb{E}[s])\end{aligned}$$

$$c = \mathbb{E}[c] + \frac{\text{cov}(s, c)}{\text{var}(s)} (s - \mathbb{E}[s]) + \epsilon$$

where the only restriction on ϵ is that

$$0 = \mathbb{E}[\mathbf{x}\epsilon] \iff 0 = \text{corr}(s, \epsilon), 0 = \mathbb{E}[\epsilon]$$

Example: illustration

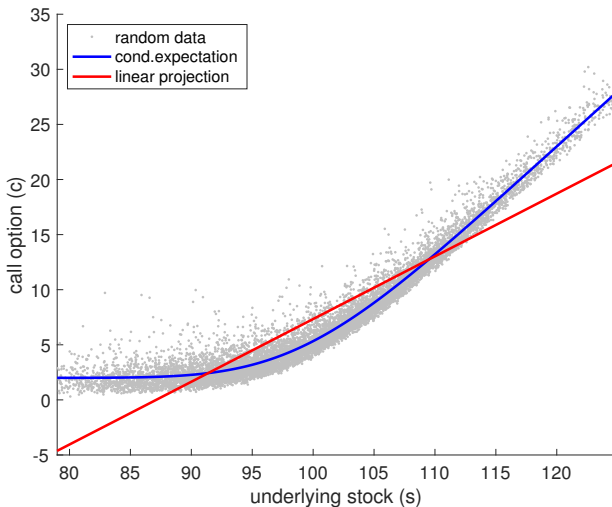


Figure: Illustration of Example 1

Example: nonlinear prediction for ϵ

- ▶ We do not have $\mathbb{E}[\epsilon | s] = 0$.
- ▶ So ϵ may predictably be small or large for certain values of s .
- ▶ However, the ways in which s predicts ϵ are not linear, given the zero correlation.

Partial derivative

- ▶ β gives predicted linear effect from observed change in \mathbf{x} .
- ▶ Not inferring a causal impact from \mathbf{x} to y .
- ▶ Suppose some other variable, z , causes change in y , but z is correlated with \mathbf{x} .
- ▶ The projection coefficients on x may be due to the implied change to z .
- ▶ For the conditional expectation, this indirect channel is wider.

Identification

If we do not condition on z , yet it is correlated with both \mathbf{x} and y , then it is known as a **confounding variable**.

- ▶ It causes the projection vector to identify not just causal impact, but inferred impact from z .
- ▶ This question of **identification** is simple in theory: we should just condition on z in addition to \mathbf{x} .
- ▶ But in realized data, we'll see this may not be possible.

Causation?

In many areas, researchers are focused on identifying causal impacts:

- ▶ y = salary, x = education, z = intelligence, responsibility, etc.
- ▶ y = crime rate, x = policing, z = lawfulness
- ▶ y = profitability, x = leverage, z = operating margin

Often not concerned with causation

However, there are many cases where this is not an issue.

- ▶ Tax authorities predicting whether someone has unreported income.
- ▶ Netflix predicting your rating of a movie.
- ▶ Assessing portfolio risk to certain factors.

Confounding variables

Example (Equity Premia)

Suppose we want to understand how a stock's characteristics, \mathbf{x} , impacts its premium, y .

$$\mathbb{E}[y | \mathbf{x}] = \mathbb{L}[y | \mathbf{x}] = \mathbf{x}\beta$$

- ▶ Population is the stocks of the S&P 500 from Jan 2005 to July 2018.
- ▶ Let \mathbf{x} be a scalar variable, the stock's dividend-yield.
- ▶ Table 1 gives the projection coefficients.

Correlation and causation

- ▶ Only using \mathbf{x} , the vector β is our best (linear) estimate of the impact of a stock return.
- ▶ However, it is not causal. It is based on the population's covariance between dividend-yield and other relevant factors.
- ▶ Should a firm take this as a prescription that if it increases the dividend yield by one point, the stock's mean excess return will change as indicated by β ?

Adding a correlated regressor

Let z denote the stock's return volatility, and include it in the projection:

$$\mathbb{L}[y \mid \mathbf{x}, z] = \mathbf{x}'\boldsymbol{\beta} + z\gamma$$

See Table 1 for the projection coefficients.

Estimates

	cond. on \mathbf{x}	cond. on (\mathbf{x}, \mathbf{z})
dividend-yield	0.0393	(-0.0009)
volatility		0.0796

Table: Projection of all S&P 500 mean excess returns on the stock's own dividend yield and return volatility.

Example conclusions

- ▶ Now, the estimated impact of dividend-yield on mean stock return almost disappears.
- ▶ Positive covariance between dividend-yield and volatility.
- ▶ Thus, in the projection without volatility, dividend-yield is measuring its own impact as well as the associated impact via volatility.
- ▶ Specifically, for every point of dividend yield, we expect 0.5 points increase in standardized volatility.