Advanced Econometrics I

Jürgen Meinecke

Lecture 1 of 12

Research School of Economics, Australian National University

Welcome

Welcome Advanced Econometrics I

This is a PhD level course in econometric theory

The course makes heavy use of the following mathematical tools:

- linear algebra
- multivariate calculus
- concepts in analysis (real and functional)

If you don't feel familiar with these, then this course will be extremely demanding

Staff

You can seek help on matters *academic* from

- your friendly lecturer (me): Juergen Meinecke
- your amazing tutor: Shu Hu

Be nice to us!

Course Website

I'm not using Wattle very much (with few exceptions)

I've set up a public course website that contains pretty much everything you need to know

Let's take a look:

`https://juergenmeinecke.github.io/EMET8014`

Roadmap

Announcements

Vector Spaces, Hilbert Spaces, Projections

- Vector Spaces, Banach Spaces
- Inner Product Spaces, Hilbert Spaces
- Projection Theorem
- Linear Projections in L_2

Definition (Vector Space)

A **real vector space** is a triple $(V, +, \cdot)$, in which V is a set, and $+$ and $\cdot$ are binary operations such that, for any two elements, $X, Y \in V$ and scalar $\lambda \in \mathbb{R}$:

$$X + Y \in V \qquad \text{(closure under additivity)}$$

$$\lambda \cdot X \in V \qquad \text{(closure under scalar product)}$$

(Note: instead of writing $\lambda \cdot X$ we typically just write λX)

Big picture:
We define a notion of addition between any two elements of V, and we define a notion of multiplication between a constant and an element of V

These operations do not take us out of the vector space

With addition and multiplication there are some typical sensible ‘requirements’:

Let $X, Y, Z \in V$, and $\lambda, \mu \in \mathbb{R}$

- addition
  1. (i) commutativity: $X + Y = Y + X$
  2. (ii) associativity: $(X + Y) + Z = X + (Y + Z)$
  3. (iii) V contains a unique element 0 such that $X + 0 = X$
  4. (iv) V contains a unique element $-X$ such that $X + (-X) = 0$
- multiplication
  1. (i) distributivity: $\lambda \cdot (X + Y) = \lambda \cdot X + \lambda \cdot Y$
  2. (ii) distributivity: $(\lambda + \mu) \cdot X = \lambda \cdot X + \mu \cdot X$
  3. (iii) associativity: $\lambda \cdot (\mu \cdot X) = (\lambda \cdot \mu) \cdot X$
  4. (iv) $1 \cdot X = X$

The perhaps most intuitive illustration of a real vector space:

Example (Euclidian Space $(\mathbb{R}^n, +, \cdot)$)

- *elements are quite literally vectors or arrows*
- $X := (x_1, \dots, x_n)'$ *and* $Y := (y_1, \dots, y_n)'$
- *define* $X + Y := (x_1 + y_1, \dots, x_n + y_n)'$
- *define* $\lambda \cdot X := (\lambda x_1, \dots, \lambda x_n)'$
- *let* $n = 2$, $X = (24, 7)'$ *and* $Y = (18, 2)'$,
  then $X + Y = (42, 9)'$

When $X \in V$, I refer to X as an "*element*" of V

Some books use "*vector*", one could also say "*point*"

A less intuitive example...

Example (The Space of Continuous Functions)

Denote by $C[a,b]$ *the space of all real valued univariate and continuous functions on a closed interval* $[a,b]$.

- *each* $X \in C[a,b]$ *is a function* $X : [a,b] \to \mathbb{R}$
- *the points or elements of the space are functions*
- *let* $t \in [a,b]$ *and write* $X(t)$ *for the function value at* t
- *define* $(X+Y)(t) := X(t) + Y(t)$
- *define* $(\lambda \cdot X)(t) := \lambda \cdot X(t)$
- *let* $[a,b] = [2,3]$, $X(t) = 2 \cdot t$ *and* $Y(t) = 1 + 5 \cdot t$,
  then $(X+Y)(t) = 1 + 7 \cdot t$

Vector spaces of functions are very important, more examples:

- space of differentiable functions
- space of functions that are integrable
  (random variables live in this space)

Each element of a vector space can be given a 'length':

Definition (Norm)

A **norm** on a real vector space V is a real valued function, denoted by $\|\cdot\|$, on V with the properties

(i) $\|X\| \geq 0$

(ii) $\|X\| = 0 \Leftrightarrow X = 0$

(iii) $\|\lambda \cdot X\| = |\lambda| \cdot \|X\|$

(iv) triangle inequality: $\|X + Y\| \leq \|X\| + \|Y\|$

where X and Y are in V, and λ is a real constant

Any function $V \mapsto \mathbb{R}$ that satisfies these properties is a norm

Here are useful examples...

Examples of norms

Example (Euclidian Space $(\mathbb{R}^n, +, \cdot)$)

- *recall that elements are quite literally vectors or arrows*
- $X := (x_1, \ldots, x_n)'$
- $\|X\| := \sqrt{x_1^2 + \cdots + x_n^2}$

Example (The Space of Continuous Functions)

- *recall that each* $X \in C[a, b]$ *is a function* $X : [a, b] \to \mathbb{R}$
- $\|X\| := \max_{t \in [a,b]} |X(t)|$

Definition (Normed Space)

A **normed space** M is a vector space endowed with a norm $\|\cdot\|$.

The norm *induces* a metric, a notion of distance between two elements

Definition (Metric)

Given a normed space M and $X, Y \in M$, the **metric** is defined by $\|X - Y\|$.

A notion of distance between elements is crucial for the understanding of limiting behavior of sequences inside the vector space

For example, does a sequence $\{X_n, n = 1, 2, \ldots\}$ of elements of a normed space get "close" to a point?

The metric is key in determining what we mean by "closeness"

Definition (Convergence in Norm)

A sequence $\{X_n, n = 1, 2, \ldots\}$ of elements of a normed space M is said to **converge in norm** to $X \in M$ if for every $\varepsilon > 0$ there is an N_ε such that $\|X_n - X\| < \varepsilon$ for every $n > N_\varepsilon$.

Definition (Cauchy Sequence)

A sequence $\{X_n, n = 1, 2, \ldots\}$ of elements of a normed space M is said to be a **Cauchy sequence** if for every $\varepsilon > 0$ there is an N_ε such that $\|X_m - X_n\| < \varepsilon$ for every $m, n > N_\varepsilon$.

Idea: elements 'out there' become arbitrarily close to each other

Definition (Complete Space)

A space M is **complete** if every Cauchy sequence of elements of M converges to an element of M, that is, every Cauchy sequence in M has a limit which is an element of M.

We like our spaces to be complete because we can safely consider limits of elements within the space

Definition (Banach Space)

A **Banach space** B is a complete normed space, that is, a normed space in which every Cauchy sequence $\{X_n, n = 1, 2, \ldots\}$ converges in norm to some element $X \in B$.

In Banach spaces we can safely play with length, distance of elements and sequences and limits of sequences of elements

But something is still missing...

Roadmap

Announcements

Vector Spaces, Hilbert Spaces, Projections

- Vector Spaces, Banach Spaces
- Inner Product Spaces, Hilbert Spaces
- Projection Theorem
- Linear Projections in L_2

Given a vector space or Banach space, we can add elements and multiply them by scalars

We can also measure their length and distance via the metric

We would, in addition, like notions of

- multiplication between elements of a space
- angle, or *orthogonality*, or perpendicularity between elements of a space

The inner product comes to the rescue

Definition (Inner Product)

An **inner product** on a vector space V is a mapping, denoted by $\langle \cdot, \cdot \rangle$, of $V \times V$ into $\mathbb{R}$ such that

$$\begin{aligned}
\langle X, Y \rangle &= \langle Y, X \rangle && \text{(commutativity)} \\
\langle X + Y, Z \rangle &= \langle X, Z \rangle + \langle Y, Z \rangle && \text{(distributivity)} \\
\langle \lambda X, Y \rangle &= \lambda \langle X, Y \rangle && \\
\langle X, X \rangle &\geq 0 && \text{(positive semi-definiteness)} \\
\langle X, X \rangle &= 0 \iff X = 0 && \text{(point separating)}
\end{aligned}$$

where X, Y and Z are in V, and λ is a real constant.

Examples of inner products

Example (Euclidian Space $(\mathbb{R}^n, +, \cdot)$)

- *recall that elements are quite literally vectors or arrows*
- $X := (x_1, \dots, x_n)'$ *and* $Y := (y_1, \dots, y_n)'$
- $\langle X, Y \rangle := x_1 y_1 + \cdots x_n y_n$

Example (The Space of Continuous Functions)

- *let* X *and* Y *be real-valued functions on* $[a, b]$
- $\langle X, Y \rangle := \int_a^b X(t) Y(t) dt$

Inner products lend themselves naturally to the creation of a norm

Definition (Induced Norm)

Let M be a normed space. The **norm induced by the inner product** is $\|X\| := \sqrt{\langle X, X \rangle}$, for any $X \in M$.

Likewise there is a metric $\|X - Y\|$ induced by the inner product

Definition (Inner Product Space)

An **inner product space** is a vector space endowed with an inner product $\langle \cdot, \cdot \rangle$.

Definition (Hilbert Space)

A **Hilbert space** is a complete inner product space.

It is clear that completeness is with respect to the norm induced by the inner product

It follows that Hilbert spaces are Banach spaces

Equipped with the inner product, we can now define the notion of angle between elements of an inner product space

Definition (Orthogonality)

Two elements X, Y of a Hilbert space are **orthogonal** if $\langle X, Y \rangle = 0$.

We write $X \perp Y$.

Think of vectors that are perpendicular

Hilbert spaces are useful for econometrics because they offer us powerful tools to address the following optimization problem:

- given an element Y in a Hilbert space H,
- and a subspace S of H,
- find the element $\widehat{Y} \in S$ closest to Y in the sense that $\left\| Y - \widehat{Y} \right\|$ is minimal

Key questions

- is there such an element $\widehat{Y}$?
- is it unique?
- what is it, or how can it be characterized?

The projection theorem answers these questions

Roadmap

Definition (Subspace)

A subset S of a Hilbert space H is called a **subspace** of H if S itself is a vector space.

We will focus on *complete* subspaces of Hilbert spaces:
subspaces that contain all of their limit points
(i.e., addition and multiplication don't take us out of the subspace)

Aside:

- the notions of *completeness* and *closedness* coincide in Banach spaces
- will use complete = closed while in Banach space

Theorem (Projection Theorem)

Let H be a Hilbert space and S be a closed subspace of H.

(i) *For any element $Y \in H$ there is a unique element $\widehat{Y} \in S$ such that $\left\|Y - \widehat{Y}\right\| \leq \|Y - s\|$ for all $s \in S$.*

(ii) *$\widehat{Y} \in S$ is the unique minimizer if and only if $Y - \widehat{Y} \perp S$.*

The element $\widehat{Y}$ is called the **orthogonal** projection of Y onto S, also denoted $\mathbb{P}_S Y$

$\mathbb{P}_S$ is the projection operator of H onto S

Existence of a unique minimizer sounds great, but how to obtain $\widehat{Y}$?

Obtaining $\widehat{Y}$ turns out to be straightforward, when the subspace onto which we are projecting is 'generated' by a finite set of elements of H

What do I mean?

A **linear combination** of elements $X_1, \dots, X_K$ of a vector space is an expression of the form $b_1 X_1 + \cdots + b_K X_K$
(where the b_i are real numbers)

Definition (Span)

Let X be a nonempty subset of a vector space. The **span** of X is the set of all linear combinations of elements of X.

We simply write $\mathrm{sp}(X)$ for the span of X

While X might not be a subspace, $\mathrm{sp}(X)$ is a subspace by construction

Put differently:
if V is a vector space, then $\mathrm{sp}(X)$ is the smallest subspace of V containing X

In econometrics, we are usually interested in the subspace spanned by 'regressors' $X_1, \dots, X_K$

Collect them all in $X = (X_1, \dots, X_K)$

Theorem (Existence of Orthonormal Basis (Gram-Schmidt))

There exists is a collection $\tilde{X}_1, \dots, \tilde{X}_K$ such that

(i) $\langle \tilde{X}_j, \tilde{X}_l \rangle = \begin{cases} 0 & \text{for } j \neq l \\ 1 & \text{for } j = l. \end{cases}$

(ii) $sp(\tilde{X}) = sp(X)$ *for* $\tilde{X} := (\tilde{X}_1, \dots, \tilde{X}_K)$.

The collection $\tilde{X}_1, \dots, \tilde{X}_K$ is called an **orthonormal basis** for $\mathrm{sp}(X)$

Why do we consider orthonormal bases?

Projections on $\mathrm{sp}(X)$ are easy to characterize via orthonormal bases

Theorem

Let $X_1, \dots, X_K$ and Y be elements from a Hilbert space. The projection of Y on $sp(X_1, \dots, X_K)$ is

$$\hat{Y} = \mathbb{P}_{sp(X)} Y = \sum_{i=1}^{K} \langle \tilde{X}_i, Y \rangle \tilde{X}_i,$$

where $\tilde{X}_1, \dots, \tilde{X}_K$ is an orthonormal basis for $sp(X)$.

This gives us a constructive method for obtaining $\widehat{Y}$

We will use it soon

Roadmap

Announcements

Vector Spaces, Hilbert Spaces, Projections

- Vector Spaces, Banach Spaces
- Inner Product Spaces, Hilbert Spaces
- Projection Theorem
- Linear Projections in L_2

Definition (The Space $L_2(\Omega, \mathcal{F}, P)$)

Let $(\Omega, \mathcal{F}, P)$ be a probability space. Denote by $L_2(\Omega, \mathcal{F}, P)$ the set of all random variables X defined on Ω with the property

$$\mathrm{E}\left(X^2\right) < \infty.$$

I'll refer to $L_2(\Omega, \mathcal{F}, P)$ simply as L_2

The condition $\mathrm{E}\left(X^2\right) < \infty$ is sometimes referred to as finite second moment or X being square integrable; it implies that $\mathrm{Var}(X) < \infty$

L_2 is a huge space

Can it be a Hilbert space?

Need an inner product: let $\langle X, Y \rangle = \mathrm{E}(X \cdot Y)$

Let $X, Y \in L_2$ with $\mathrm{E}(X) = \mathrm{E}(Y) = 0$, then $\langle X, Y \rangle = \mathrm{Cov}(X, Y)$

In other words, the inner product we're using here is related to the familiar notion of covariance

Proposition

The space L_2 with $\langle X, Y \rangle = E(X \cdot Y)$ is a Hilbert space.

See Brockwell and Davis, "*Time Series: Theory and Methods*" for proof

What is our overarching objective?

We want to "*predict*" one random variable (the dependent variable) using a bunch of other random variables (independent variables, exogenous variables, regressors)

Mapping this into the framework of the projection theorem

- the dependent variable Y is a "point" in the Hilbert space L_2
- the regressors make up a subspace onto which we "project" Y
- the projection theorem tells us that there exists a unique optimal $\hat{Y}$

We need to be precise about the subspace created by the regressors

Let there be a *finite* collection $X_1, \ldots, X_K \in L_2$

Let $X := (X_1, \ldots, X_K)'$

Proposition

$sp(X)$ is a complete subspace of L_2.

Recall our theorem a few slides earlier:

$$\mathbb{P}_{\mathrm{sp}(X)} Y = \sum_{i=1}^{K} \langle \tilde{X}_i, Y \rangle \tilde{X}_i$$

Using our inner product

$$\mathbb{P}_{\mathrm{sp}(X)} Y = \sum_{i=1}^{K} \mathsf{E}\left(\tilde{X}_i \cdot Y\right) \tilde{X}_i$$

Now, let's go slow and set $K = 1$

That is, we only have one regressor to predict Y

Going back to the question: What is $\widehat{Y}$ equal to?

Answer, of course, depends on choice of subspace to project on

In the current example we have $\mathrm{sp}(X_1)$
(so only one random variable)

Let's make the problem even simpler and pick $X_1 = 1$
(the degenerate rv that is almost surely equal to 1)

What is the orthonormal basis for $\mathrm{sp}(1)$?

Easy: X_1 already is an orthonormal basis (because $\mathrm{E}(1 \cdot 1) = 1$)

It follows $\mathbb{P}_1 Y = \mathrm{E}(1 \cdot Y) \cdot 1 = \mathrm{E}Y = \mu_Y$

Of course you knew this already:
The expected value of Y is the projection of Y onto a constant

What if we use a more sophisticated space for the projection?

Let $X_1 = 1$ and $X_2 \in L_2$ and project on $\text{sp}(1, X_2)$

Let's find an orthonormal basis of $\text{sp}(1, X_2)$

We need to find a version of X_2 that has length 1

This is easy: $\tilde{X}_2 := (X_2 - \mu_2)/\sigma_2$ achieves this:
$\|\tilde{X}_2\| = \sqrt{\mathsf{E}\left(\frac{X_2-\mu_2}{\sigma_2}\right)^2} = 1$

Turns out that $\{1, \tilde{X}_2\}$ form an orthonormal basis of $\text{sp}(1, X_2)$ (confirm this!)

The example of $\tilde{X}_2$ offers you some intuition about orthonormal bases: it is the *standardized version* of X_2

With $\tilde{X}_1 := 1$ and $\tilde{X}_2 = (X_2 - \mu_2)/\sigma_2$, it follows,

$$
\begin{aligned}
\widehat{Y} = \mathbb{P}_{\mathrm{sp}(1,X_2)} Y &= \sum_{i=1}^{2} \mathsf{E}(\tilde{X}_i \cdot Y)\tilde{X}_i \\
&= \mathsf{E}(1 \cdot Y) \cdot 1 + \mathsf{E}(\tilde{X}_2 Y)\tilde{X}_2 \\
&= \mathsf{E}Y + \mathsf{E}(((X_2 - \mu_2)/\sigma_2)Y)(X_2 - \mu_2)/\sigma_2 \\
&= \mathsf{E}Y + \frac{\mathsf{E}(X_2 Y) - \mu_2 \mathsf{E}Y}{\sigma_2^2}(X_2 - \mu_2) \\
&= \mathsf{E}Y + \frac{\mathrm{Cov}(X_2, Y)}{\sigma_2^2}(X_2 - \mu_2) \\
&= \beta_1^* + \beta_2^* X_2,
\end{aligned}
$$

where

$$
\begin{aligned}
\beta_2^* &:= \sigma_{2Y}/\sigma_2^2 \\
\beta_1^* &:= \mathsf{E}Y - \beta_2^* \mathsf{E}X_2
\end{aligned}
$$

Let's generalize this once more

What if we're projecting on $\text{sp}(1, X_2, X_3)$?

You might think that $\{1, \tilde{X}_2, \tilde{X}_3\}$ form an orthonormal basis,
where $\tilde{X}_2 := (X_2 - \mu_2)/\sigma_2$ and $\tilde{X}_3 := (X_3 - \mu_3)/\sigma_3$

Not so, sorry!

Convince yourself that $\mathrm{E}(\tilde{X}_2\tilde{X}_3) \neq 0$ unless $\tilde{X}_2 \overset{\text{as}}{=} \tilde{X}_3$ which is a boring case

Intuitively, the problem is that X_2 and X_3 have nonzero covariance

How do we construct orthonormal bases out of two random variables that have nonzero covariance?

Answer: Gram-Schmidt orthogonalization!

Gram-Schmidt orthogonalization applied to the current context, involves these simple steps:

1. Let $\tilde{X}_1 := 1$
2. create $\ddot{X}_2 := X_2 - \mathrm{E}(X_2\tilde{X}_1)\tilde{X}_1$
   normalize by its length: $\tilde{X}_2 := \ddot{X}_2/\sqrt{\mathrm{Var}\,\ddot{X}_2}$
   (notice that because we include a constant term, $\mathrm{Var}\,\ddot{X}_2 = \mathrm{E}\ddot{X}_2^2$ and therefore $\sqrt{\mathrm{Var}\,\ddot{X}_2} = \|\ddot{X}_2\|$)
3. create $\ddot{X}_3 := X_3 - \mathrm{E}(X_3\tilde{X}_1)\tilde{X}_1 - \mathrm{E}(X_3\tilde{X}_2)\tilde{X}_2$
   normalize by its length: $\tilde{X}_3 := \ddot{X}_3/\sqrt{\mathrm{Var}\,\ddot{X}_3}$
   (notice again that $\sqrt{\mathrm{Var}\,\ddot{X}_3} = \|\ddot{X}_3\|$)

You can view Gram-Schmidt orthogonalization as an iterative algorithm to construct orthonormal bases

By the way, the order in which you are doing this does not matter

When you work this out (and you should!), you get

$$\tilde{X}_2 = (X_2 - \mu_2)/\sigma_2$$
$$\ddot{X}_3 = (X_3 - \mu_3) - \frac{\sigma_{23}}{\sigma_2^2}(X_2 - \mu_2)$$

where $\sigma_{23} := \text{Cov}(X_2, X_3)$

What's going on here?

$\tilde{X}_2$ is the same as before, it's the standardized version of X_2

$\ddot{X}_3$ is a particular version of X_3: it is the part of X_3 that has zero covariance with X_2; it's been orthogonolized

$\tilde{X}_3$ is an appropriately normalized version of $\ddot{X}_3$ so that its length is 1

Also, convince yourself that if $\text{Cov}(X_2, X_3) = 0$ then $\tilde{X}_3 = (X_3 - \mu_3)/\sigma_3$

With the orthonormal basis it's easy to construct the projection:

$$\mathbb{P}_{\mathsf{sp}(1,X_2,X_3)}Y = \sum_{i=1}^{3} \mathsf{E}(\tilde{X}_i \cdot Y)\tilde{X}_i$$

It is tedious but not difficult to show that

$$\begin{aligned}\mathbb{P}_{\mathsf{sp}(1,X_2,X_3)}Y &= \mathsf{E}Y + \beta_2^*(X_2 - \mu_2) + \beta_3^*(X_3 - \mu_3) \\ &= \beta_1^* + \beta_2^* X_2 + \beta_3^* X_3\end{aligned}$$

where

$$\begin{aligned}\beta_2^* &:= \frac{\sigma_{2Y}\sigma_3^2 - \sigma_{3Y}\sigma_{23}}{\sigma_2^2\sigma_3^2 - \sigma_{23}^2} \\ \beta_3^* &:= \frac{\sigma_{3Y}\sigma_2^2 - \sigma_{2Y}\sigma_{23}}{\sigma_2^2\sigma_3^2 - \sigma_{23}^2} \\ \beta_1^* &:= \mathsf{E}Y - \beta_2^*\mathsf{E}X_2 - \beta_3^*\mathsf{E}X_3\end{aligned}$$

Looks awkward but it is an important result to internalize!

Look what happens when Cov$(X_2, X_3) = 0$:

$$\beta_2^* := \frac{\sigma_{2Y}\sigma_3^2 - \sigma_{3Y}\sigma_{23}}{\sigma_2^2\sigma_3^2 - \sigma_{23}^2} = \frac{\sigma_{2Y}}{\sigma_2^2}$$

$$\beta_3^* := \frac{\sigma_{3Y}\sigma_2^2 - \sigma_{2Y}\sigma_{23}}{\sigma_2^2\sigma_3^2 - \sigma_{23}^2} = \frac{\sigma_{3Y}}{\sigma_3^2}$$

How would you construct an orthonormal basis for $\mathsf{sp}(X_1, X_2, \dots, X_K)$ with $X_1 = 1$ and $X_k \in L_2$ for $k = 2, \dots, K$?

Again, use Gram-Schmidt orthogonalization with the inductive definitions:

1. $\tilde{X}_1 := 1$
2. $\ddot{X}_2 := X_2 - \mathrm{E}(X_2 \tilde{X}_1) \tilde{X}_1$
   $\tilde{X}_2 := \ddot{X}_2 / \|\ddot{X}_2\| = \ddot{X}_2 / \sqrt{\mathrm{Var}\ \ddot{X}_2}$
3. $\ddot{X}_3 := X_3 - \mathrm{E}(X_3 \tilde{X}_1) \tilde{X}_1 - \mathrm{E}(X_3 \tilde{X}_2) \tilde{X}_2$
   $\tilde{X}_3 := \ddot{X}_3 / \|\ddot{X}_3\| = \ddot{X}_3 / \sqrt{\mathrm{Var}\ \ddot{X}_3}$
4. $\ddot{X}_4 := X_4 - \mathrm{E}(X_4 \tilde{X}_1) \tilde{X}_1 - \mathrm{E}(X_4 \tilde{X}_2) \tilde{X}_2 - \mathrm{E}(X_4 \tilde{X}_3) \tilde{X}_3$
   $\tilde{X}_4 := \ddot{X}_4 / \|\ddot{X}_4\| = \ddot{X}_4 / \sqrt{\mathrm{Var}\ \ddot{X}_4}$
5. and so forth

The resulting projection will have the form

$$\widehat{Y} = \mathbb{P}_X Y = \sum_{i=1}^{K} \mathrm{E}(\tilde{X}_i \cdot Y)\tilde{X}_i,$$

where, for simplicity, we write $\mathbb{P}_X$ for $\mathbb{P}_{\mathrm{sp}(X_1,\ldots,X_K)}$

Another symbol we use a lot is $\hat{Y}$ which is defined to be $\mathbb{P}_X Y$

The projection can be summarized neatly in matrix notation:

Theorem

Let $X := (X_1, X_2, \ldots, X_K)'$ be a $K \times 1$ vector. Then

$$\mathbb{P}_X Y = X'\beta^*,$$

where $\beta^* := \left(E(XX')\right)^{-1} E(XY)$.

For simplicity we will write $\mathrm{E}(XX')^{-1}$ for $\left(\mathrm{E}(XX')\right)^{-1}$