paper shows that the present value budget balance restriction leads to an exact linear rational expectations model. In a separate paper by Roberds, such a model is estimated and the restriction is tested for U.S. data on government expenditures and taxes. Roberds finds evidence against the restriction for these data.

Notes

1. Chapter 2 was written in 1981–82. Chapter 3 was written in 1980–81 (Hansen and Sargent 1981e), and revised in 1990. Chapter 4 was written in 1982, with minor revisions being made in 1984 (Hansen and Sargent 1984) and 1989. Most of chapter 5 was written in 1987, with the empirical work being completed in 1990. Chapter 6 was written in 1988. Chapters 7 and 8 were written in 1988–90. They amount to revisions and extensions of ideas that initially appeared in working papers by Hansen and Sargent that appeared in 1980 and 1981 (Hansen and Sargent (1980b, 1981d). Chapter 9 was completed in 1980–81 (Hansen and Sargent 1981c). Chapter 10 was written as part of Albert Marcet's Ph.D. dissertation, which was completed in 1987.
2. See Muth (1960, 1961), Nerlove (1967), Griliches (1967), Sims (1974), Lucas (1972), and Sargent (1971).
3. This way of introducing errors in econometric models was used to great advantage by Shiller (1972).
4. The paper also describes how its partial equilibrium model is to be interpreted as a special case of the class of general equilibrium models studied by Hansen and Sargent (1990).
5. See Sargent (1987b, chapter XIII) for a discussion of the relationship between Hall's (1978) model and Barro's (1979) tax smoothing model. Evidently, there are tax smoothing models that are similarly related to the general class of consumption smoothing models described by Hansen, Roberds, and Sargent or by Hansen and Sargent (1990).

2

Lecture Notes on Least Squares Prediction Theory

by Lars Peter HANSEN and Thomas J. SARGENT

1. Introduction

In these notes we establish some basic results for least squares prediction theory. These results are useful in a variety of contexts. For instance, they are valuable for solving linear rational expectations models, representing covariance stationary time series processes, and obtaining martingale difference decompositions of strictly stationary processes.

The basic mathematical construct used in these notes is an inner product defined between two random variables. This inner product is calculated by taking the expectation of the product of the two random variables. Many of the results obtained using this particular inner product are analogous to results obtained using the standard inner product on multi-dimensional Euclidean spaces. Hence intuition obtained for Euclidean spaces can be quite valuable in this context as well.

The formal mathematical machinery that is exploited in these notes is the Hilbert space theory. There is a variety of references on Hilbert spaces that should provide good complementary reading, e.g. Halmos (1957) and Luenberger (1969).

2. Prediction Problem

In this section we specify formally the problem of forecasting a random variable y given a collection of random variables H. This problem is sufficiently general to include conditional expectations and best linear predictors as special cases. We also consider a second problem that is closely related to the prediction problem. This second problem is termed the orthogonality problem and can be interpreted as providing a set of necessary and sufficient first-order conditions for the prediction problem. In particular, we will show that these two problems have the same solution when some auxiliary assumptions are imposed on the set H.

A formal statement of the least squares prediction problem is:

PROBLEM 2.1 (Least Squares Forecasting Problem): Find the random variable h_o in H for which $E[(y-h_o)^2] \leq E[(y-h)^2]$ for all h in H.

There is a second problem that often has the same solution as Problem 2.1. This second problem uses the following definition in its statement.

Definition 2.1: The random variable y is orthogonal to the set H if $E(hy) = 0$ for all h in H.

The second problem is:

PROBLEM 2.2 (Orthogonality Problem): Find the random variable h_o in H such that $y - h_o$ is orthogonal to H.

To ensure that the statements of these two problems are well-defined, we make the assumption that the random variable being forecast and the random variables that are used in forecasting have finite second moments. With this in mind, we let L^2 denote the set of all random variables defined on the underlying probability space that have finite second moments. We assume that both y and the elements in H are in L^2. There are two important inequalities that apply to random variables in L^2.

Lemma 2.1: (Cauchy-Schwarz Inequality): For any y_1 and y_2 in L^2, $|E(y_1 y_2)| \leq E(y_1^2)^{\frac{1}{2}} E(y_2^2)^{\frac{1}{2}}$.

Proof: If either y_1 or y_2 is zero, then the inequality is satisfied trivially. Suppose that y_1 and y_2 are both different from zero. Then $E(y_1^2)$ and $E(y_2^2)$ are both positive. Notice that

$$\begin{aligned} 0 &\leq \left[\frac{|y_1|}{(Ey_1^2)^{\frac{1}{2}}} - \frac{|y_2|}{(Ey_2^2)^{\frac{1}{2}}}\right]^2 \\ &= \left[\frac{y_1^2}{E(y_1^2)} - \frac{2|y_1 y_2|}{E(y_1^2)^{\frac{1}{2}} E(y_2^2)^{\frac{1}{2}}} + \frac{y_2^2}{E(y_2^2)}\right]. \end{aligned} \tag{2.1}$$

Consequently,

$$\frac{|y_1 y_2|}{E(y_1^2)^{\frac{1}{2}} E(y_2^2)^{\frac{1}{2}}} \leq \frac{1}{2}\left[\frac{y_1^2}{E(y_1^2)} + \frac{y_2^2}{E(y_2^2)}\right]. \tag{2.2}$$

Taking expectations of both sides of (2.2) gives

$$\frac{E|y_1 y_2|}{E(y_1^2)^{\frac{1}{2}} E(y_2^2)^{\frac{1}{2}}} \leq 1. \tag{2.3}$$

Multiplying by $E(y_1^2)^{\frac{1}{2}}$ and $E(y_2^2)^{\frac{1}{2}}$ gives

$$E(|y_1 y_2|) \leq E(y_1^2)^{\frac{1}{2}} E(y_2^2)^{\frac{1}{2}}. \tag{2.4}$$

The Cauchy-Schwarz Inequality then follows from the fact that

$$|E(y_1 y_2)| \leq E(|y_1 y_2|). \blacksquare \tag{2.5}$$

Lemma 2.2: (Triangle Inequality): For any y_1 and y_2 in L^2, $E[(y_1 + y_2)^2]^{\frac{1}{2}} \leq E(y_1^2)^{\frac{1}{2}} + E(y_2^2)^{\frac{1}{2}}$.

Proof: Notice that

$$\begin{aligned} (y_1 + y_2)^2 &= y_1^2 + 2y_1 y_2 + y_2^2 \\ &\leq y_1^2 + 2|y_1 y_2| + y_2^2. \end{aligned} \tag{2.6}$$

Taking expectations and using inequality (2.4) gives

$$\begin{aligned} E[(y_1 + y_2)^2] &\leq E(y_1^2) + 2E(y_1^2)^{\frac{1}{2}} E(y_2^2)^{\frac{1}{2}} + E(y_2^2) \\ &= [E(y_1^2)^{\frac{1}{2}} + E(y_2^2)^{\frac{1}{2}}]^2. \end{aligned} \tag{2.7}$$

Finally, taking square roots of each side of (2.7) gives the Triangle Inequality. $\blacksquare$

Among other things, the Cauchy-Schwarz and Triangle Inequalities imply that the statements of Problems 2.1 and 2.2 are well-defined as long as y and the elements of H are in L^2. In addition, the Triangle Inequality implies that L^2 is a linear space in the sense that linear combinations of elements in L^2 are in L^2. When h_o solves Problem 2.2 for a subset of G of L^2, then h_o also solves Problem 2.2 for the set

$$\begin{aligned} H = \Big\{h : h = c_1 g_1 + c_2 g_2 + \cdots + c_n g_n &\text{ for some integer } n, \\ \text{some real numbers } c_1, c_2, \ldots, c_n, &\text{ and some elements} \\ g_1, g_2, \ldots, g_n \text{ in } G\Big\}. \end{aligned} \tag{2.8}$$

Thus H is constructed to be a linear subspace of L^2.

Definition 2.2: The set H is a linear subspace of L^2 if H is a subset of L^2 and if for any h_1 and h_2 in H and any real numbers c_1 and c_2, the random variable $c_1h_1 + c_2h_2$ is in H.

Consequently, when h_o solves Problem 2.2 on an arbitrary subset of L^2, this subset can always be expanded to be a linear subspace of L^2 via (2.8). The random variable h_o will continue to solve Problem 2.2 on the expanded set. In what follows we will focus on linear subspaces of H.

Our next two lemmas show that Problems 2.1 and 2.2 have the same solution when H is a linear subspace of L^2.

Lemma 2.3: Suppose that y is in L^2 and H is a linear subspace of L^2. If h_o solves problem 2.2, then h_o is the unique solution to Problems 2.1 and 2.2.[1]

Proof: Let h_o be the solution to Problem 2.2. For any h in H,

$$\begin{aligned} E[(y-h)^2] &= E[(y-h_o+h_o-h)^2] = E[(y-h_o)^2]+ \\ &\quad 2E[(y-h_o)(h_o-h)] + E[(h_o-h)^2] \\ &= E[(y-h_o)^2] + E[(h_o-h)^2]\ . \end{aligned} \tag{2.9}$$

Hence,

$$E[(y-h_o)^2] \le E[(y-h)^2] \tag{2.10}$$

for any h in H implying that h_o solves Problem 2.1. Furthermore, if $E[(h-h_o)^2]$ is greater than zero, then

$$E[(y-h_o)^2] < E[(y-h)^2]\ . \tag{2.11}$$

Since $E[(h_o-h)^2] = 0$ only if $h = h_o$, h_o is the unique solution to Problem 2.1. It follows that h_o is also the unique solution to Problem 2.2 since we have shown that any solution to Problem 2.2 is a solution to Problem 2.1. ▮

Lemma 2.4: Suppose y is in L^2 and H is a linear subspace of L^2. If h_o solves Problem 2.1, then h_o also solves Problem 2.2.

Proof: Let h_o be the solution to Problem 2.1 and let h be any element in H. If h is zero, it follows immediately that $E[(y-h_o)h]$ is zero. If h is different from zero, let $\delta = E[(y-h_o)h]/E(h^2)$. Notice that

$$\begin{aligned} E[(y-h_o-\delta h)^2] &= E[(y-h_o)^2] + \delta^2E[h^2] - 2\delta E[(y-h_o)h] \\ &= E[(y-h_o)^2] - \delta^2E[h^2]\ . \end{aligned} \tag{2.12}$$

Since h_o solves Problem 2.1 and $h_o + \delta h$ is in H, δ must be zero. However, δ is zero only when $E[(y-h_o)h]$ is zero. Since our choice of h in H was arbitrary, $y - h_o$ is orthogonal to H. ▮

Lemmas 2.3 and 2.4 showed that Problems 2.1 and 2.2 have the same solution as long as H is a linear subspace of L^2. However, we have not shown that a solution exists for either problem. An additional restriction on the set H is sufficient to guarantee that a solution exists. This restriction uses the following definitions:

Definition 2.3: A sequence $\{y_n : n \ge 1\}$ in L^2 is Cauchy if $\lim_{n\to\infty} \sup_{m\ge n} E[(y_n-y_m)^2] = 0$.

Definition 2.4: A sequence $\{y_n : n \ge 1\}$ in L^2 is convergent if there exists a random variable y_o in L^2 for which $\lim_{n\to\infty} E[(y_n-y_o)^2] = 0$.[2]

The linear space L^2 is complete in the sense that any Cauchy sequence in L^2 is convergent (see the Appendix). If the analogous property holds for a subset of L^2, we say that the subset is closed.

Definition 2.5: A subset H of L^2 is closed if any Cauchy sequence in H converges to an element in H.

If a given set H is not closed, we can form its closure by adding all limit points of Cauchy sequences in H to H. The resulting augmented set will be closed.

Lemma 2.5: Suppose H is a closed linear subspace of L^2. Then, for any y in L^2, Problem 2.1 has a solution.

Proof: Let δ be the nonnegative real number that satisfies

$$\delta^2 = \inf_{h\in H} E[(y-h)^2]\ , \tag{2.15}$$

and let $\{h_n : n = 1, 2, \ldots\}$ be a sequence of random variables in H for which

$$\lim_{n\to\infty} E[(y-h_n)^2] = \delta^2\ . \tag{2.16}$$

First, we show that when H is a linear space, $\{h_n : n \ge 1\}$ is a Cauchy sequence. Notice that

$$\begin{aligned} &E[(h_m-h_n)^2] + 4E\big\{[(1/2)(h_m+h_n)-y]^2\big\} \\ &\quad = 2E[(h_m-y)^2] + 2E[(h_n-y)^2]\ . \end{aligned} \tag{2.17}$$

Therefore,

$$E[(h_m - h_n)^2] \le 2E[(h_m - y)^2] + 2E[(h_n - y)^2] - 4\delta^2 \tag{2.18}$$

since $(1/2)(h_m + h_n)$ is in H and $\delta^2 < E[(y-h)^2]$ for any h in H. Taking limits of both sides of (2.18) gives

$$\begin{aligned} 0 \le \lim_{m\to\infty} \sup_{m \ge n} E[(h_m - h_n)^2] &\le 2 \lim_{m\to\infty} E[(h_m - y)^2] + \\ &\quad 2 \lim_{n\to\infty} E[(h_n - y)^2] - 4\delta^2 \\ &= 2\delta^2 + 2\delta^2 - 4\delta^2 \\ &= 0 \, . \end{aligned} \tag{2.19}$$

Hence, $\{h_n : n \ge 1\}$ is Cauchy as long as H is a linear space. If in addition H is closed, this sequence converges to an element, h_o, in H. We now show that h_o is the solution to the least squares forecasting problem. By the Triangle Inequality (Lemma 2.2),

$$E[(y - h_o)^2]^{\frac{1}{2}} \le E[(y - h_n)^2]^{\frac{1}{2}} + E[(h_n - h_o)^2]^{\frac{1}{2}} \, . \tag{2.20}$$

Taking limits of both sides of (2.20) gives

$$\left\{E[(y - h_o)^2]\right\}^{\frac{1}{2}} \le \delta \tag{2.21}$$

which proves that h_o solves Problem 2.1. ∎

We have shown that if H is a closed linear space then there exists a unique solution to the least squares forecasting problem (Problem 2.1). In these circumstances we define the projection operator as follows.

Definition 2.6: For any H that is a closed linear subspace of L^2, the least squares projection operator, $P(y|H)$ is the solution to Problem 2.1 of forecasting y given H.

There are two trivial examples of projection operators. For the first example let H be L^2. The projection operator $P(\cdot|L^2)$ is just the identity operator since $P(y|L^2) = y$. For the second example let H be the set Z that contains only the zero random variable. The projection operator $P(\cdot|Z)$ maps all random variables into the zero random variable since the zero random variable is the only random variable in Z. In sections five and six we study two nontrivial projection operators that are used in studying time series prediction problems.

One can be interpreted as a best linear predictor and the other as a conditional expectation.

Our final result in this section establishes that $P(\cdot|H)$ satisfies a linearity property.

Lemma 2.6: Suppose that H is a closed linear subspace of L^2. Then for any y_1 and y_2 in L^2 and any real numbers c_1 and c_2

$$P(c_1 y_1 + c_2 y_2 | H) = c_1 P(y_1|H) + c_2 P(y_2|H) \, . \tag{2.22}$$

Proof: To prove this lemma, we will show that the right-hand side of (2.22) solves the orthogonality problem (Problem 2.2). Notice that $y_1 - P(y_1|H)$ and $y_2 - P(y_2|H)$ are orthogonal to H. Consequently,

$$c_1 y_1 + c_2 y_2 - c_1 P(y_1|H) - c_2 P(y_2|H) \tag{2.23}$$

is orthogonal to H. Furthermore, $c_1 P(y_1|H) + c_2 P(y_2|H)$ is in H. Therefore, $c_1 P(y_1|H) + c_2 P(y_2|H)$ solves the orthogonality problem. ∎

3. Useful Results Involving Projections

In this section we present some results that are useful in calculating projections. In preparation for these results, we introduce some new concepts.

Definition 3.1: The sets G and U are orthogonal if $E(gu) = 0$ for all g in G and all u in U.

Definition 3.2: The sum of the sets G and U is $G + U = \{g + u : g$ is in G and u is in $U\}$.

The sum of two closed linear subspaces of L^2 that are orthogonal is also a closed linear subspace of L^2.

Lemma 3.1: Suppose that G and U are orthogonal closed linear subspaces of L^2. Then $G + U$ is a closed linear subspace of L^2.

Proof: It is straightforward to show that if G and U are linear spaces, then their sum is also a linear space. We leave this as an exercise for the reader. To show that $G + U$ is closed, let $\{(g_n + u_n) : n \ge 1\}$ be a Cauchy sequence in $G + U$ where $\{g_n : n \ge 1\}$ is a sequence in G and $\{u_n : n \ge 1\}$ is a sequence in U. We will show that these latter two sequences are Cauchy. Since $g_n - g_m$ is in G, $u_n - u_m$ is in U, and G and U are orthogonal,

$$E[(g_n + u_n - g_m - u_m)^2] = E[(g_n - g_m)^2] + E[(u_n - u_m)^2] \, . \tag{3.1}$$

Therefore, the following two inequalities hold:

$$E[(g_n + u_n - g_m - u_m)^2] \geq E[(u_n - u_m)^2] \; , \tag{3.2}$$

and

$$E[(g_n + u_n - g_m - u_m)^2] \geq E[(g_n - g_m)^2] \; . \tag{3.3}$$

Since $\{g_n + u_n : n \geq 1\}$ is Cauchy, it follows that $\{g_n : n \geq 1\}$ and $\{u_n : n \geq 1\}$ are Cauchy and hence convergent. Recall that both G and U are assumed to be closed so that $\{g_n : n \geq 1\}$ converges to an element g_o in G and $\{u_n : n \geq 1\}$ converges to an element u_o in U. Since

$$E[(g_n + u_n - g_o - u_o)^2] = E[(u_n - u_o)^2] + E[(g_n - g_o)^2] \; , \tag{3.4}$$

$\{g_n + u_n : n \geq 1\}$ converges to $g_o + u_o$ in $G + U$. ∎

Among other things Lemma 3.1 guarantees that the projection operator is well-defined on $G + U$ when G and U are orthogonal closed linear spaces. In this case the projection onto $G + U$ is the sum of the projections onto G and U.

Lemma 3.2: Suppose that G and U are orthogonal closed linear subspaces of L^2. Then $P(\cdot|G + U) = P(\cdot|G) + P(\cdot|U)$.

Proof: Let y be any element in L^2. To prove this lemma, we will show that $y - P(y|G) - P(y|U)$ is orthogonal to $G + U$. Since $P(y|G)$ solves the orthogonality problem (Problem 2.2) for the subspace G, $y - P(y|G)$ is orthogonal to G. Since U is assumed to be orthogonal to G, $P(y|U)$ is orthogonal to G. Hence, $y - P(y|G) - P(y|U)$ is orthogonal to G. By reversing the roles of G and U, it follows that $y - P(y|U) - P(y|G)$ is also orthogonal to U. Therefore, $y - P(y|G) - P(y|U)$ is orthogonal to $U + G$. ∎

Lemma 3.2 gives a convenient way for revising forecasts when new information is received. For instance, suppose that G contains information available in the past and U contains new information (orthogonal to G) that becomes available today. Then Lemma 3.2 shows how to update forecasts after the arrival of the new information. This raises the following question. When can new information be viewed as a closed linear space that is orthogonal to the set of information available previously? The following result will be used in answering this question.

Lemma 3.3: Suppose the sequences $\{y_n : n \geq 1\}$ and $\{y_n^* : n \geq 1\}$ in L^2 converge to y_o and y_o^*, respectively. Then the sequence of real numbers $\{E(y_n y_n^*) : n \geq 1\}$ converges to $E(y_o y_o^*)$.

Proof: Notice that

$$\begin{aligned} E(y_n y_n^*) - E(y_o y_o^*) = {} & E[y_o(y_n^* - y_o^*)] \\ & + E[(y_n - y_o)y_o^*] \\ & + E[(y_n - y_o)(y_n^* - y_o^*)] \; . \end{aligned} \tag{3.5}$$

From the Cauchy-Schwarz Inequality and the Triangle Inequality for real numbers,

$$\begin{aligned} \left| E(y_n y_n^*) - E(y_o y_o^*) \right| \leq {} & E(y_o^2)^{\frac{1}{2}} E[(y_n^* - y_o^*)^2]^{\frac{1}{2}} \\ & + E[(y_n - y_o)^2]^{\frac{1}{2}} E(y_o^{*2})^{\frac{1}{2}} \\ & + E[(y_n - y_o)^2]^{\frac{1}{2}} E[(y_n^* - y_o^*)^2]^{\frac{1}{2}} \; . \end{aligned} \tag{3.6}$$

The conclusion follows from noting that the right-hand side of (3.6) converges to zero since $\{y_n - y_o : n \geq 1\}$ and $\{y_n^* - y_o^* : n \geq 1\}$ converge to zero. ∎

Our next lemma gives a construction for the set U of new information.

Lemma 3.4: Suppose that G and H are closed linear subspaces of L^2 and G is a subset of H. Let

$$U = \{u : u = h - P(h|G) \text{ for some } h \text{ in } H\} \; . \tag{3.7}$$

Then U is a closed linear subspace of L^2 and $H = U + G$.

Proof: Since G is a subset of H, $h - P(h|G)$ is in H for any h in H. Therefore, U is a subset of H that is orthogonal to G. Hence, $U + G$ is a subset of H. However, any h in H can be represented as $h = P(h|G) + [h - P(h|G)]$ which shows that h is in $U + G$. Consequently, $H = G + U$ where G and U are orthogonal. Next we show that U is a closed linear subspace of L^2. Let u be any element in H that is orthogonal to G. Then $P(u|G) = 0$ so that u is in U. Since any element of U is orthogonal to G,

$$U = \{u \text{ in } H : u \text{ is orthogonal to } G\} \; . \tag{3.8}$$

Recall that H is a linear subspace of L^2. Also, any linear combination of elements in H that are orthogonal to G is also orthogonal to G. Hence,

U is a linear space. From Lemma 3.3 and the closure of H, it follows that any Cauchy sequence in H that is orthogonal to G converges to an element in H that is also orthogonal to G. Therefore, U is closed. ∎

Given a specification of G and using (3.7) as the definition of new information, one can always decompose a closed linear subspace of L^2. The Law of Iterated Projections can be proved using such a decomposition.

Lemma 3.5: Suppose that G and H are closed linear subspaces of L^2 and G is a subset of H. Then for any y in L^2, $P(y|G) = P[P(y|H)|G]$.

Proof: Using Lemma 3.4 we obtain an orthogonal decomposition of H into the sum of U given by (3.7) and G. Using Lemma 3.2 it follows that $P(y|H) = P(y|G) + P(y|U)$. Since $P(y|U)$ is in U and U is orthogonal to G, $P[P(y|U)|G] = 0$. Therefore, $P[P(y|H)|G] = P[P(y|G)|G] = P(y|G)$. ∎

Lemmas 3.2 and 3.5 offer an interesting comparison. Lemma 3.2 shows how to update a projection from a smaller information set G to larger information set H. In contrast, Lemma 3.5 shows how to calculate a projection onto a smaller information set G in terms of a projection onto a larger information set H.

As noted previously, one interpretation of the set G is that it contains past information. An alternative interpretation will also be used in some of our analysis. This interpretation involves a continuous linear functional π mapping H into the set of real numbers $\mathbf{R}$.

Definition 3.3: π is a linear functional on a linear subspace H of L^2 if π maps H into $\mathbf{R}$, and for any h_1 and h_2 in H and c_1 and c_2 in $\mathbf{R}$, $\pi(c_1h_1 + c_2h_2) = c_1\pi(h_1) + c_2\pi(h_2)$.

Definition 3.4: A linear functional π on a linear subspace H of L^2 is continuous if for any sequence $\{h_n : n \geq 1\}$ in H that converges to zero, the sequence of real numbers $\{\pi(h_n) : n \geq 1\}$ converges to zero.[3]

Linear functionals are important in studying competitive equilibrium pricing in environments with uncertainty. For instance, we can think of H as being a set of possible portfolio payoffs and π as a pricing function that assigns a price to each payoff. The pricing function can be used to construct G.

Lemma 3.6: Suppose H is a closed linear subspace of L^2 and π is a continuous linear functional on H. Let

$$G = \{g \in H : \pi(g) = 0\} . \tag{3.9}$$

Then G is a closed linear subspace of H.

Proof: The linearity of G follows from the linearity of π since for any g_1 and g_2 in G and any real numbers c_1 and c_2,

$$\begin{aligned}\pi(c_1g_1 + c_2g_2) &= c_1\pi(g_1) + c_2\pi(g_2)\\ &= 0 .\end{aligned} \tag{3.10}$$

The closure of G follows from the continuity of π. To see this, let $\{g_n : n \geq 1\}$ be a Cauchy sequence in G. Then $\{g_n : n \geq 1\}$ is a Cauchy sequence in H and hence converges to some element g_o in H. To see that g_o is also in G, notice that $\{g_n - g_0 : n \geq 1\}$ converges to zero. Since π is linear and continuous, $\{\pi(g_n) - \pi(g_o) \ : n \geq 1\}$ converges to zero. However, $\pi(g_n) = 0$ for all n so that $\pi(g_o) = 0$. Therefore, g_o is in G implying that G is closed. ∎

The set G given in (3.9) can be viewed as the set of all portfolio payoffs with zero prices. Taken together, Lemmas 3.4 and 3.6 provide an orthogonal decomposition of a closed linear subspace H of L^2. The next lemma provides a characterization of U as given by (3.7) when G is given by (3.9).

Lemma 3.7: Suppose H is a closed linear subspace of L^2 and π is a continuous linear functional on H. Then for G given by (3.9), the corresponding set U given by (3.7) satisfies

$$u = \pi(u)u^* \tag{3.11}$$

for all u in U and some u^* in U.

Proof: Consider two cases. First, suppose that U contains only the zero random variable. Then u^* can be chosen to be zero. Second, suppose that U contains some element u_o other than zero. Then $\pi(u_o)$ is different from zero since u_o is not in G. Let

$$u^* = u_o/\pi(u_o) . \tag{3.12}$$

For any u in U, $u - \pi(u)u^*$ is in G since $\pi(u^*)$ is one. Thus,

$$u = \pi(u)u^* \tag{3.13}$$

since $u - \pi(u)u^*$ is both in U and orthogonal to U and hence is orthogonal to itself. ∎

Lemma 3.7 shows that the set U is at most one-dimensional when a linear functional π is used to construct G. It turns out that an element in this one-dimensional set can be used to represent π.

Lemma 3.8: Suppose that H is a closed linear subspace of L^2 and that π is a continuous linear functional on H. Then there exists a unique element h^* in H such that $\pi(h) = E(hh^*)$.[4]

Proof: First, we study existence of h^*, and then we consider uniqueness of h^*. From Lemma 3.7 we know that there exists a u^* satisfying (3.11). Suppose u^* is zero. Then h^* can be chosen to be zero since $\pi(h)$ is zero for all h. Next suppose u^* is different from zero. Let $h^* = u^*/E(u^{*2})$. Notice that h^* is in U and

$$\pi(h^*) = 1/E(u^{*2}) \ . \tag{3.14}$$

Since $\pi(u^*)$ is one, for any h in H, $h - \pi(h)u^*$ is in G and hence orthogonal to h^*. Thus,

$$\begin{aligned} E(hh^*) &= \pi(h)E(u^*h^*) \\ &= \pi(h)E(u^{*2})/E(u^{*2}) \\ &= \pi(h) \ . \end{aligned} \tag{3.15}$$

Next, we show that h^* is unique. Let h^* and h^+ be any two elements in H for which $\pi(h) = E(hh^*) = E(hh^+)$ for all h in H. Then $h^* - h^+$ is orthogonal to H and in particular is orthogonal to itself. This can happen only if $h^* - h^+$ is zero. ∎

4. Infinite Sequences of Subspaces

In this section we study the limiting behavior of projections onto infinite sequences of closed linear subspaces of L^2. We present two types of results corresponding to whether the sequences are increasing or decreasing.

Definition 4.1: A sequence $\{H_n : n \geq 1\}$ of subsets of L^2 is increasing if $H_{n+1} \supset H_n$ for all $n \geq 1$.

Definition 4.2: A sequence $\{H_n : n \geq 1\}$ of subsets of L^2 is decreasing if $H_n \supset H_{n+1}$ for all $n \geq 1$.

First, we consider projections onto an increasing sequence $\{H_n : n \geq 1\}$ of closed linear subspaces of L^2. The set

$$\bigcup_{n\geq 1} H_n \tag{4.1}$$

is a linear subspace of L^2 that is not necessarily closed. We leave it as an exercise for the reader to verify that the set given in (4.1) is a linear subspace. Let H_o denote the closure of the linear subspace given in (4.1). The following result shows how to approximate projections onto H_o.

Lemma 4.1: Suppose $\{H_n : n \geq 1\}$ is an increasing sequence of closed linear subspaces of L^2. Then for any y in L^2, $\{P(y|H_n) : n \geq 1\}$ converges to $P(y|H_o)$.

Proof: Let y be any element of L^2. For $m \geq n$, H_m contains H_n. Consequently, Lemma 3.5 implies that $P(y|H_n) = P[P(y|H_m)|H_n]$ which in turn implies that $P(y|H_m) - P(y|H_n)$ is orthogonal to H_n. Therefore,

$$E\Big\{[P(y|H_m) - P(y|H_n)]^2\Big\} = E[P(y|H_m)^2] - E[P(y|H_n)^2] \ . \tag{4.2}$$

Since the left hand side of (4.2) is nonnegative, the sequence of real numbers $\{E[P(y|H_n)^2] : n \geq 1\}$ is nondecreasing. Furthermore,

$$E(y^2) = E[P(y|H_n)^2] + E\Big\{[y - (y|H_n)]^2\Big\} \geq E[P(y|H_n)^2] \ . \tag{4.3}$$

Thus, $\{E[P(y|H_n)^2] : n \geq 1\}$ is a bounded monotone sequence. Therefore, there exists a real number δ such that

$$\lim_{n\to\infty} E[P(y|H_n)^2] = \delta \ . \tag{4.4}$$

Limit (4.4) and equality (4.2) imply that

$$\begin{aligned} &\lim_{n\to\infty} \sup_{m\geq n} E\Big\{[P(y|H_m) - P(y|H_n)]^2\Big\} \\ &= \lim_{n\to\infty} \sup_{m\geq n} \Big\{E[P(y|H_m)^2] - E[P(y|H_m)^2]\Big\} \\ &= 0 \ . \end{aligned} \tag{4.5}$$

because $\{E[P(y|H_n)^2] : n \geq 1\}$ is a convergent and hence Cauchy sequence of real numbers. Thus, $\{P(y|H_n) : n \geq 1\}$ is Cauchy and hence converges to some element h_o in H_o.

To prove that $h_o = P(y|H_o)$, let h be any element in H_o. Then there exists a sequence $\{h_n : n \geq 1\}$ in H_o that converges to h for which h_n is in H_n. Thus by Lemma 3.3,

$$E[(y - h_o)h] = \lim_{n\to\infty} E\Big\{[y - P(y|H_n)]h_n\Big\} \ . \tag{4.6}$$

However, $E\{[y - P(y|H_n)]h_n\}$ is zero for all $n \geq 1$. Consequently,

$$E[(y - h_o)h] = 0 \tag{4.7}$$

for any h in H_o which proves that h_o solves the orthogonality problem (Problem 2.2). ∎

One way to construct an increasing sequence of closed linear subspaces of L^2 is as follows. Let $\{y_n : n \geq 1\}$ be a sequence in L^2, and let

$$(4.8)\quad H_n = \Big\{h : h = c_1 y_1 + c_1 y_2 + \cdots + c_n y_n \ \text{ for some real numbers } c_1,\, c_2,\, \ldots,\, c_n\Big\} .$$

Lemma 4.2: If y_1, y_2, ..., y_n are in L^2, then H_n given by (4.8) is a closed linear subspace of L^2.

Proof: It is straightforward to prove that H_n is a linear subspace of L^2. We leave this as an exercise for the reader. To prove that H_n is closed, first suppose that $n = 1$. Let $\{h_m : m \geq 1\}$ be any Cauchy sequence in H_1. We consider two cases. First suppose y_1 is zero. Then h_m is zero for each m implying that $\{h_m : m \geq 1\}$ converges to zero. Hence H_1 is closed in this case. Next suppose y_1 is different from zero. Then for each $n \geq 1$, $h_m = c_m y_1$ for some real number c_m. Hence for $\ell \geq m$,

$$(4.9)\quad E[(h_\ell - h_m)^2] = (c_\ell - c_m)^2\, E(y_1^2) .$$

Since $\{h_n : n \geq 1\}$ is Cauchy in L^2 and $E(y_1^2) > 0$, $\{c_n : n \geq 1\}$ is a Cauchy sequence of real numbers. Thus $\{c_n : n \geq 1\}$ converges to some real number c_o. Let $h_o = c_o y_1$. Then

$$(4.10)\quad \begin{aligned} \lim_{n\to\infty} E[(h_n - h_o)^2] &= \lim_{n\to\infty} (c_n - c_o)^2\, E(y_1^2) \\ &= 0 \end{aligned}$$

which proves that H_1 is closed since $c_o y_1$ is in H_1.

Next, suppose that H_{n-1} is closed. Let

$$(4.11)\quad U_n = \Big\{u : u = c[y_n - P(y_n|H_{n-1})] \ \text{ for some } c \text{ in } \mathbf{R}\Big\} .$$

Notice that U_n is a subset of H_n since elements of U_n are linear combinations of y_n and an element in H_{n-1}. By mimicking the proof that H_1 is closed, it can be shown that U_n is closed. Hence, Lemma 3.1 implies that $U_n + H_{n-1}$ is closed. Since H_{n-1} and U_n are subsets of H_n and H_n is a subspace of L^2, $U_n + H_{n-1}$ is a subset of H_n. However, any element in H_n can be represented as

$$(4.12)\quad h = c_1 y_1 + c_2 y_2 + \cdots + c_n P(y_n|H_{n-1}) + c_n[y_n - P(y_n|H_{n-1})] .$$

Therefore, $U_n + H_{n-1} = H_n$ which proves that H_n is closed. ∎

Taken together, Lemmas 4.1 and 4.2 characterize the limiting behavior of projections as additional variables are included in the space onto which we are projecting. In particular, Lemma 4.1 shows that the sequence of projections converges to a random variable that can be interpreted as the projection onto the limiting space.

In some cases the increasing sequence of subspaces of L^2 is constructed from an underlying sequence of orthogonal spaces.

Definition 4.3: The sequence of subsets $\{U_n : n \geq 1\}$ of L^2 is orthogonal if U_n is orthogonal to U_m for all $n \geq 1$ and $m \geq 1$ for which $n \neq m$.

Definition 4.4: The sum of a sequence of orthogonal subsets of L^2 is

$$\sum_{n=1}^{\infty} U_n = \Big\{u : u = \sum_{n=1}^{\infty} u_n \ \text{ where } u_n \text{ is in } U_n \text{ and } \sum_{n=1}^{\infty} E(u_n)^2 < \infty\Big\} .$$

A sequence $\{U_n : n \geq 1\}$ of orthogonal closed linear subspaces of L^2 can be used to construct an increasing sequence of closed linear subspaces of L^2. Let

$$(4.13)\quad H_1 = U_1 ,$$

and

$$(4.14)\quad H_n = U_n + H_{n-1} \ \text{ for } n \geq 2 .$$

Then Lemma 3.1 guarantees that H_n is a closed linear subspace of L^2 for $n \geq 1$. Furthermore, since the zero random variable is in U_n, H_n contains H_{n-1}. The following result is the infinite dimensional counterpart to Lemmas 3.1 and 3.2.

Lemma 4.3: Suppose $\{U_n : n \geq 1\}$ is a sequence of orthogonal closed linear subspaces of L^2 and H_n is given by (4.13) and (4.14). Then

$$(4.15)\quad H_o = \sum_{n=1}^{\infty} U_n$$

and

$$P(\cdot|H_0) = \sum_{n=1}^{\infty} P(\cdot|U_n) . \tag{4.16}$$

Proof: Let y be any element in L^2. By repeated application of Lemma 3.2, it follows that

$$P(y|H_m) = \sum_{n=1}^{m} P(y|U_n) . \tag{4.17}$$

Lemma 4.1 implies that $\{P(y|H_m) : m \geq 1\}$ converges to $P(y|H_o)$ so that $\{P(y|H_m) : m \geq 1\}$ is a Cauchy sequence in H_o. Thus,

$$\lim_{m\to\infty} \sup_{\ell\geq m+1} E\Big\{\Big[\sum_{n=m+1}^{\ell} P(y|U_n)\Big]^2\Big\} = 0 . \tag{4.18}$$

Since the sequence $\{U_n : n \geq 1\}$ is orthogonal,

$$E\Big\{\Big[\sum_{n=m+1}^{\ell} P(y|U_n)\Big]^2\Big\} = \sum_{n=m+1}^{\ell} E[P(y|U_n)^2] \tag{4.19}$$

for any $\ell \geq m+1$. Thus (4.18) implies that

$$\lim_{m\to\infty} \sum_{n=m+1}^{\infty} E[P(y|U_n)^2] = 0 \tag{4.20}$$

which in turn implies that

$$\sum_{n=1}^{\infty} E[P(y|U_n)^2] < \infty . \tag{4.21}$$

Consequently, $\sum_{n=1}^{\infty} P(y|U_n)$ is in $\sum_{n=1}^{\infty} U_n$ and is the limit point of $\{P(y|H_n) : n \geq 1\}$. This proves (4.16).

To prove (4.15), let h be any element of H_o. Then $h = P(h|H_o)$. Applying (4.16), it follows that

$$h = \sum_{n=1}^{\infty} P(h|U_n) . \tag{4.22}$$

Hence, h is in $\sum_{n=1}^{\infty} U_n$. Conversely, suppose that u is in $\sum_{n=1}^{\infty} U_n$. Then u can be represented as

$$u = \sum_{n=1}^{\infty} u_n \tag{4.23}$$

where u_n is in U_n for $n \geq 1$. Notice that u_m is orthogonal to H_n for $m > n$. Hence, Lemma 3.3 guarantees that $\sum_{m=n+1}^{\infty} u_m$ is orthogonal to H_n. This in turn implies that

$$P(u|H_n) = \sum_{m=1}^{n} u_m . \tag{4.24}$$

Applying Lemma 4.1 then gives

$$\begin{aligned} P(u|H_o) &= \sum_{m=1}^{\infty} u_m \\ &= u . \end{aligned} \tag{4.25}$$

Therefore, u is in H_o. ▮

For the remainder of this section, we study decreasing sequences of subspaces. Let $\{H_n : n \geq 1\}$ be a decreasing sequences of subspaces of L^2. Define the limiting set to be

$$H_\infty = \bigcap_{n=1}^{\infty} H_n . \tag{4.26}$$

Lemma 4.4: Suppose $\{H_n : n \geq 1\}$ is a decreasing sequence of closed linear subspaces of L^2. Then H_∞ is a closed linear subspace of L^2.

Proof: To prove that H_∞ is a linear subspace of L^2, let h_1 and h_2 be any two elements in H_∞ and let c_1 and c_2 be any two real numbers. Then h_1 and h_2 are in H_n for all $n \geq 1$. Since H_n is a closed linear subspace of L^2, $c_1h_1 + c_2h_2$ is in H_n for all $n \geq 1$. Thus $c_1h_1 + c_2h_2$ is in H_∞ which proves that H_∞ is linear.

To prove that H_∞ is closed, let $\{h_n : n \geq 1\}$ be any Cauchy sequence in H_∞. Then $\{h_n : n \geq 1\}$ is in H_m for all $m \geq 1$. Since H_m is closed, $\{h_n : n \geq 1\}$ converges to some element h_∞ in H_m. Consequently, h_∞ is in H_∞ which proves that H_∞ is closed. ▮

Lemma 4.5: Suppose $\{H_n : n \geq 1\}$ is a decreasing sequence of closed linear subspaces of L^2. Then for any y in L^2, $\{P(y|H_n) : n \geq 1\}$ converges to $P(y|H_\infty)$.

Proof: Let y be any element of L^2. The first part of this proof mimics the proof of Lemma 4.1 except that for $m \geq n$, $P(y|H_m) = P[P(y|H_n)|H_m]$. Hence $P(y|H_n) - P(y|H_m)$ is orthogonal to $P(y|H_m)$. Consequently,

$$E\Big\{[P(y|H_m) - P(y|H_n)]^2\Big\} = E[P(y|H_n)^2] - E[P(y|H_m)^2] \, . \tag{4.27}$$

and $\{E[P(y|H_n)^2] : n \geq 1\}$ is a decreasing sequence with a lower bound of zero. Thus $\{P(y|H_m) : m \geq n\}$ is a Cauchy sequence in H_n for each $n \geq 1$. Since all of the H_n spaces are closed, the limit point h_∞ is in each of these spaces. Hence, h_∞ is also in H_∞.

To prove that $h_\infty = P(y|H_\infty)$, let h be any element of H_∞. Then for all $n \geq 1$, $y - P(y|H_n)$ is orthogonal to h since H_∞ is contained in H_n. Lemma 3.3 implies that

$$\begin{aligned} E[(y - h_\infty)h] &= \lim_{n\to\infty} E\Big\{[y - P(y|H_n)]h\Big\} \\ &= 0 \, . \end{aligned} \tag{4.28}$$

Therefore, h_∞ solves the orthogonality problem (Problem 2.2). ▮

When $\{H_n : n \geq 1\}$ is a decreasing sequence of closed linear subspaces of L^2, it is always possible to obtain an infinite dimensional orthogonal decomposition of any space in the sequence. To see this, let

$$U_n = \Big\{u : u = h - P(h|H_{n+1}) \text{ for some } h \text{ in } H_n\Big\} \, . \tag{4.29}$$

Lemma 3.4 shows that $H_n = U_n + H_{n+1}$ where U_n is a closed linear subspace of L^2 that is orthogonal to H_{n+1}. Also, for $m \geq n+1$, U_m is contained in H_{n+1} so that U_m is orthogonal to U_n. Consequently, $\{U_n : n \geq 1\}$ is an orthogonal sequence of closed linear subspaces of L^2.

Lemma 4.6: Suppose $\{H_n : n \geq 1\}$ is a decreasing sequence of closed linear subspaces of L^2. Then for any $n \geq 1$,

$$H_n = \sum_{m=0}^{\infty} U_{n+m} + H_\infty \tag{4.30}$$

where U_n is given by (4.29).

Proof: Given $n \geq 1$ and $\ell \geq 1$,

$$\begin{aligned} H_n &= U_n + H_{n+1} \\ &= U_n + U_{n+1} + \ldots + U_{n+\ell-1} + H_{n+\ell} \, . \end{aligned} \tag{4.31}$$

Notice that the sets on the right-hand side of (4.31) are mutually orthogonal. Let h be any element in H_n. Then applying Lemma 3.2,

$$h = P(h|H_n) = \sum_{m=0}^{\ell-1} P(h|U_{n+m}) + P(h|H_{n+\ell}) \, . \tag{4.32}$$

Lemma 4.5 implies that $\{h - P(h|H_{n+\ell}) : \ell \geq 1\}$ converges to $h - P(h|H_\infty)$. Consequently,

$$\Big\{\sum_{m=1}^{\ell-1} P(h|U_{n+m}) : \ell \geq 1\Big\} \tag{4.33}$$

converges. Lemma 4.3 guarantees that

$$\sum_{m=0}^{\infty} U_{n+m} \tag{4.34}$$

is a closed linear subspace of L^2. Hence, the random variable

$$\sum_{j=o}^{\infty} P(h|U_{n+j}) = h - P(h|H_\infty) \tag{4.35}$$

is in that closed linear space. Therefore,

$$h = \sum_{m=0}^{\infty} P(h|U_{n+m}) + P(h|H_\infty) \tag{4.36}$$

where $P(h|H_\infty)$ is in H_∞ and $\sum_{m=0}^{\infty} P(h|U_{n+m})$ is in $\sum_{m=o}^{\infty} U_{n+m}$. Relation (4.36) implies that the set on the left-hand side of (4.30) is contained in the set on the right-hand side. The equality follows from the fact that H_n is closed and U_{n+m} is contained in H_n for all m as is H_∞. ▮

Lemma 4.6 gives an important decomposition of decreasing sequences of closed linear subspaces of L^2. In the next two sections we will use this decomposition with two different notions of prediction. One of these notions of corresponds to calculating the conditional expectation and the other to finding the best linear prediction.

5. Conditional Expectations

In this section we show that conditional expectations can be viewed as a special case of projections. Our treatment of this topic presumes some familiarity with measure theory. Since this discussion is not essential to the subsequent analysis, readers may choose to proceed to section six.

Let (Ω, A, Pr) denote the underlying probability space, let x be a p-dimensional random vector defined on this space, let $\mathbf{R}^p$ denote the p-dimensional Euclidean space, and let

$$G = \{y \text{ in } L^2 : y = f(x) \text{ for some Borel measurable function } f \text{ mapping } \mathbf{R}^p \text{ into } \mathbf{R}\} \; . \tag{5.1}$$

Lemma 5.1: G given by (5.1) is a closed linear subspace of L^2.

Proof: Let B be the collection of subsets of Ω given by

$$B = \{b : b = \{\omega \text{ in } \Omega \text{ such that } x(\omega) \text{ is in } b^*\} \text{ for some Borel set } b^* \text{ in } \mathbf{R}^p\} \; . \tag{5.2}$$

We leave it as an exercise for the reader to verify that B is a subsigma algebra of A. In the appendix we prove that the L^2 space given in section two is complete. This space is defined using the sigma algebra A. By replacing A with B it follows that

$$G^* = \{y : y \text{ is measurable with respect to } B \text{ and } E(y^2) < \infty\} \tag{5.3}$$

is a closed linear subspace of L^2.

We will prove that $G = G^*$. First, suppose y is in G. Then y is in L^2 and $y = f(x)$ for some Borel measurable function f. Let b^+ be any Borel set in $\mathbf{R}$. Then

$$b^* = \{r : f(r) \text{ is in } b^+\} \tag{5.4}$$

is a Borel set in $\mathbf{R}^p$. Hence

$$\{\omega : f[x(\omega)] \text{ is in } b^*\} = \{\omega : x(\omega) \text{ is in } b^*\} \tag{5.5}$$

is in B. Therefore, y is in G^*.

Next, suppose y is in G^*. Then we can write $y = y^+ - y^-$ where

$$y^+ = \begin{cases} y & \text{if } y \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

and

$$y^- = \begin{cases} -y & \text{if } y \leq 0 \\ 0 & \text{otherwise} \end{cases} \; . \tag{5.7}$$

Both y^+ and y^- are in G^*. Indicator functions for sets in B can be expressed as Borel measurable functions of x. Hence, simple functions (linear combinations of indicator functions) can also be expressed as Borel measurable functions of x. Since y^+ can always be approximated by an increasing sequence of simple functions, there exists an increasing sequence $\{f_n(x) : n \geq 1\}$ that converges almost surely to y^+ where f_n is a Borel measurable function for each n. Let

$$f^* = \lim_{n\to\infty} \sup f_n \tag{5.8}$$

and

$$f^+ = \begin{cases} f^* & \text{if } f^* < \infty \\ 0 & \text{otherwise} \end{cases} \; . \tag{5.9}$$

Then both f^* and f^+ are Borel measurable and $\{f_n(x) : n \geq 1\}$ converges almost surely to $f^+(x) = y^+$. A similar argument can be used to construct a Borel measurable function f^- such that $y^- = f^-(x)$. Let $f = f^+ - f^-$. Then f is Borel measurable and $y = f(x)$. ▮

Next we show that projections onto G are equal to expectations conditioned on x.

Lemma 5.2: For G given in (5.1), $P(z|G) = E(z|x)$.

Proof: Notice that B given in (5.2) is the smallest sigma algebra for which x is measurable. Let b be any set in B, and let b^* be the corresponding Borel set in $\mathbf{R}^p$ for which

$$b = \{\omega : x(\omega) \text{ is in } b^*\} \; . \tag{5.10}$$

We define

$$1_b = \begin{cases} 1 & \text{if } x \text{ is in } b^* \\ 0 & \text{otherwise} \end{cases} \; . \tag{5.11}$$

Then 1_b is in G since it is a Borel measurable function of x and $E(1_b^2) < \infty$. For any y in L^2, $P(y|G)$ solves the orthogonality problem (Problem 2.2). Hence

$$E\{[P(y|G) - y]1_b\} = 0 \; , \tag{5.12}$$

or equivalently

$$\int_b P(y|G)\, dPr = \int_b y\, dPr \; . \tag{5.13}$$

Since $P(y|G)$ is measurable with respect to B, and (5.13) holds for any b in B, it follows that $P(y|G) = E(y|x)$.[5] ∎

In light of this Lemma 5.2, we see that conditional expectations solve the least squares forecasting problem (Problem 2.1) as long as arbitrary (Borel measurable) nonlinear forecasting rules are permitted in constructing elements of G. Thus conditional expectations are special cases of projections.

Next, we shift from conditioning on a random vector to conditioning on current and past values of a vector stochastic process. Let $\{x(t) : -\infty < t < +\infty\}$ be a p-dimensional stochastic process, and let

(5.14)
$$H_n(t) = \Big\{ y \text{ in } L^2 : y = f[x(t),\ x(t-1),\ \ldots,\ x(t-n)] \text{ for some Borel measurable function } f \text{ mapping } \mathbf{R}^{p(n+1)} \text{ into } \mathbf{R} \Big\} .$$

Then Lemma 5.1 guarantees that $H_n(t)$ is a closed linear subspace of L^2, and Lemma 5.2 guarantees that projections onto $H_n(t)$ are the same as expectations conditioned on $x(t),\ x(t-1),\ \ldots,\ x(t-n)$. We define $H(t)$ to be the closure of

$$\bigcup_{n=1}^{\infty} H_n(t) . \tag{5.15}$$

Lemma 4.1 shows that projections onto $H(t)$ can be approximated by projections onto $H_n(t)$ for sufficiently large n. This raises the question of whether projections onto $H(t)$ can be interpreted as expectations conditioned on $\{x(t-n) : \text{ for } n \geq 0\}$.

<u>Lemma 5.3:</u> $P[z|H(t)] = E[z|x(t),\ x(t-1), \ldots]$.

<u>Proof:</u> For each $n \geq 1$, let

$$B_n = \Big\{ b : b = \big\{[x(t)',\ x(t-1)',\ \ldots,\ x(t-n)']' \text{ is in } b^* \big\} \text{ for some Borel set in } \mathbf{R}^{p(n+1)} \Big\} . \tag{5.16}$$

Then $\{B_n : n \geq 1\}$ is an increasing sequence of subsigma algebras of A. Let

$$B^* = \bigcup_{n \geq 1} B_n . \tag{5.17}$$

The collection of events B^* is an algebra since for any b_1 and b_2 in B^* the union of b_1 and b_2 is in B^*, and for any b in B^* the complement of b is in B^*. However, B^* is not necessarily a sigma algebra. For any b in B^*, the random variable

$$1_b = \begin{cases} 1 & \text{on } b \\ 0 & \text{otherwise} \end{cases} \tag{5.18}$$

is in $H_m(t)$ for some positive integer m. Hence, for any y in L^2, $y - P[y|H_n(t)]$ is orthogonal to 1_b for all $n \geq m$. It follows from Lemmas 3.3 and 4.1 that

$$E\Big\{\big\{y - P[y|H(t)]\big\}1_b\Big\} = \lim_{n\to\infty} E\Big\{\big\{y - P[y|H_n(t)]\big\}1_b\Big\} = 0 . \tag{5.19}$$

Therefore, $y - P[y|H(t)]$ is orthogonal to all indicator functions of sets in B^* or equivalently,

$$\int_b y\, dPr = \int_b P[y|H(t)] dPr \tag{5.20}$$

for all b in B^*.

As in the proof of Lemma 5.1, we write $y = y^+ - y^-$ where y^+ and y^- are given by (5.6) and (5.7) respectively. Since (5.20) holds for any y in L^2 it must hold for y^+ as well. For a given y^+, we define a measure μ on B^* to be

$$\mu(b) = \int_b y^+\, dPr = \int_b P[y^+|H(t)]\, dPr . \tag{5.21}$$

Let B_o be the smallest sigma algebra containing B^*. This sigma algebra is also the smallest sigma algebra for which $x(t),\ x(t-1),\ \ldots,$ are measurable. There exists a unique extension of the measure μ from B^* to B_o.[6] Thus

$$\mu(b) = \int_b y^+\, dPr = \int_b P[y^+|H(t)]\, dPr \tag{5.22}$$

for all b in B_o. Recall that $\{P[y^+|H_n(t)] : n \geq 1\}$ converges to $P[y^+|H(t)]$. The Riesz-Fischer Theorem (see the appendix) implies that the space

$$G = \{y^* \text{ in } L^2 : y^* \text{ is measurable with respect to } B_o\} \tag{5.23}$$

is complete. Therefore, $P[y^+|H(t)]$ is measurable with respect to B_o and satisfies the defining characteristics for $E[y^+|x(t), x(t-1), \ldots]$.[7]

A similar argument can be applied to y^-. Combining these results and applying Lemma 2.6 gives

(5.24)
$$\begin{aligned} P[y|H(t)] &= P[y^+|H(t)] - P[y^-|H(t)] \\ &= E[y^+|x(t), x(t-1), \ldots] - E[y^-|x(t), x(t-1), \ldots] \\ &= E[y|x(t), x(t-1), \ldots]\ . \end{aligned}$$ ∎

A convenient way to calculate projections is as follows. First obtain an orthogonal decomposition of the linear subspace being projected onto. Then calculate projections onto these smaller orthogonal subspaces and add them together. This approach can be used in calculating projections onto $H(t)$ since the sequence of subspaces $\{H(t-n) : n \geq 0\}$ is decreasing. Lemma 4.6 shows how to obtain an infinite-dimensional decomposition of this decreasing sequence. This decomposition uses the subspaces

(5.25)
$$U(t-n) = \left\{u : u = h - P[h|H(t-1-n)] \text{ for some } h \text{ in } H(t-n)\right\}$$

for $n \geq 0$. In general, these subspaces will be infinite-dimensional. Also, projections onto $U(t)$ cannot be interpreted as conditional expectations although they can be interpreted as revisions in conditional expectations. In the next section, we consider an alternative notion of prediction that has the computational advantage that the space corresponding to $U(t-n)$ in (5.25) is finite-dimensional.

6. Linear Prediction Theory

In this section we consider an alternative notion of prediction to conditional expectations. This notion has the advantage that the associated orthogonal decompositions are easier to characterize. In order to obtain this simplification we restrict forecasting rules to be linear functions of an underlying set of random variables. As in section five, a sequence of closed linear spaces is generated using a p-dimensional stochastic process $\{x(t) : -\infty < t < +\infty\}$. We assume that each component of $x(t)$ is in L^2 for each time period t, and we define

(6.1)
$$\begin{aligned} H_n(t) = \{h : h = \alpha_1 \cdot x(t) + \alpha_2 \cdot x(t-1) + \ldots + \alpha_n \cdot x(t-n+1) \\ \text{for some vectors } \alpha_1, \alpha_2, \ldots, \alpha_n \text{ in } \mathbf{R}^p\}\ . \end{aligned}$$

Lemma 4.2 guarantees that $H_n(t)$ is a closed linear subspace of L^2 for each t and each $n \geq 1$. Also, $\{H_n(t) : n \geq 1\}$ is an increasing sequence of subspaces for each t. We define $H(t)$ to be the closure of

$$\bigcap_{n=1}^{\infty} H_n(t)\ . \tag{6.2}$$

Lemma 4.1 shows that projections onto $H(t)$ can be approximated by projections onto $H_n(t)$ for large values of n. We interpret the projection operator onto $H(t)$ as the best linear predictor given current and past values of $x(t)$.

One is tempted to think that any element in $H(t)$ can be represented as

$$\sum_{\tau=0}^{\infty} \alpha_\tau \cdot x(t-\tau) \tag{6.3}$$

for some sequence $\{\alpha_\tau : \tau \geq 0\}$ in $\mathbf{R}^p$. It turns out that this is not always true. For example, suppose that p is one, and let

$$x(t) = v(t) - v(t-1) \tag{6.4}$$

where

$$\begin{aligned} &E[v(t)^2] = 1 && \text{for all } t \\ &E[v(t)\, v(t-\tau)] = 0 && \text{for all } t \text{ and for all } \tau \neq 0\ . \end{aligned} \tag{6.5}$$

First, we will show that $v(t)$ is in $H(t)$. To see this, let

$$y_n(t) = (n/n)\, x(t) + [(n-1)/n]\, x(t-1) + \ldots + (1/n)\, x(t-n+1)\ . \tag{6.6}$$

Then $y_n(t)$ is in $H(t)$ and

$$y_n(t) = v(t) - (1/n) \sum_{\tau=1}^{n} v(t-\tau)\ . \tag{6.7}$$

Notice that

$$E\left\{[(1/n) \sum_{\tau=1}^{n} v(t-\tau)]^2\right\} = (1/n)\ . \tag{6.8}$$

Therefore, $\{y_n(t) : n \geq 1\}$ converges to $v(t)$ proving that $v(t)$ is in $H(t)$. It turns out that $v(t)$ cannot be represented as

$$v(t) = \sum_{\tau=0}^{\infty} \alpha_\tau \cdot x(t-\tau)\ . \tag{6.9}$$

To see this, suppose to the contrary that (6.9) is true. Multiplying both sides of (6.9) by $v(t-j)$ and taking expectations gives

$$\begin{aligned} 1 &= \alpha_0 && \text{for } j = 0 \\ 0 &= \alpha_j - \alpha_{j-1} && \text{for } j \geq 1 . \end{aligned} \tag{6.10}$$

These calculations use Lemma 3.3. Relation (6.10) implies that $\alpha_j = 1$ for all $j \geq 1$. However,

$$\sum_{\tau=0}^{n} x(t-\tau) = v(t) - v(t-n-1) . \tag{6.11}$$

Thus,

$$v(t) - \sum_{\tau=0}^{n} x(t-\tau) = v(t-n-1) \tag{6.12}$$

Clearly, the right-hand side of (6.12) does not converge to zero as n goes to infinity. Therefore, a representation of the form (6.3) does not exist for $v(t)$. Consequently, we cannot necessarily characterize the elements of $H(t)$ as infinite linear combinations of current and past values of $x(t)$.

It is often fruitful to characterize $H(t)$ in terms of an orthogonal decomposition. As in section five, we define

$$U(t) = \left\{ u : u = h - P[h|H(t-1)] \text{ for some } h \text{ in } H(t) \right\} . \tag{6.13}$$

The following result illustrates an advantage of using linear predictors.

Lemma 6.1: The set $U(t)$ given in (6.13) satisfies

$$U(t) = \{u : u = \alpha \cdot u(t) \text{ for some } \alpha \text{ in } \mathbf{R}^p\} \tag{6.14}$$

where $u(t) = x(t) - P[x(t)|H(t-1)]$.[8]

Proof: Let $U^*(t)$ be the space defined by the right side of (6.14). Suppose h is in $H(t)$. We must prove that

$$h - P[h|H(t-1)] = \alpha \cdot u(t) \tag{6.15}$$

for some α in $\mathbf{R}^p$. Since h is in $H(t)$, Lemma 4.1 implies that there exists a sequence $\{h_n : n \geq 1\}$ converging to h such that

$$h_n = \alpha_0^n \cdot x(t) + \alpha_1^n \cdot x(t-1) + \ldots + \alpha_n^n \cdot x(t-n) \tag{6.16}$$

for some vectors $\alpha_0^n, \alpha_1^n, \ldots, \alpha_n^n$ in $\mathbf{R}^p$. Then

$$\begin{aligned} h_n - P[h_n|H(t-1)] &= \alpha_0^n \cdot \left\{x(t) - P[x(t)|H(t-1)]\right\} \\ &= \alpha_0^n \cdot u(t) . \end{aligned} \tag{6.17}$$

For $m \geq n$

$$\begin{aligned} E[(h_n - h_m)^2] = &E\left\{[(\alpha_0^m - \alpha_0^n) \cdot u(t)]^2\right\} \\ &+ E\left\{P[(h_n - h_m)|H(t-1)]^2\right\} . \end{aligned} \tag{6.18}$$

Therefore, $\{\alpha_0^n \cdot u(t) : n \geq 1\}$ is a Cauchy sequence in $U^*(t)$ where $U^*(t)$ is given by the right-hand side of (6.14). Lemma 4.3 guarantees that $U^*(t)$ is closed so that there exists α in $\mathbf{R}^p$ such that $\{\alpha_0^n \cdot u(t) : n \geq 1\}$ converges to $\alpha \cdot u(t)$. Similarly, (6.18) implies that $\{P[h_n|H(t-1)] : n \geq 1\}$ converges to some random variable h_o in $H(t-1)$. For any h^* in $H(t-1)$,

$$\begin{aligned} E[(h - h_o)h^*] &= \lim_{n\to\infty} E\left\{\{h_n - P[h_n|H(t-1)]\}h^*\right\} \\ &= 0 \end{aligned} \tag{6.19}$$

so that $h_o = P[h|H(t-1)]$. Thus $\{h_n - P[h_n|H(t-1)] : n \geq 1\}$ converges to $h - P[h|H(t-1)] = \alpha \cdot u(t)$. ▮

Lemma 6.1 shows that $U(t)$ is a finite dimensional subspace of $H(t)$ containing linear combinations of the one-step ahead forecast errors in forecasting $x(t)$ using $H(t-1)$. This characterization underlies the Wold decomposition of a time series process. Let

$$H(-\infty) = \bigcap_{n=1}^{\infty} H(t-n) . \tag{6.20}$$

Notice that this definition of $H(-\infty)$ does not depend on the choice of time t. Lemma 4.6 shows that

$$H(t) = \sum_{n=0}^{\infty} U(t-n) + H(-\infty) . \tag{6.21}$$

Projecting $x(t)$ onto $H(t)$ and applying Lemma 4.6 gives

$$x(t) = \sum_{n=o}^{\infty} P[x(t)|U(t-n)] + P[x(t)|H(-\infty)] . \tag{6.22}$$

In light of Lemma 6.1, we obtain

$$x(t) = \sum_{n=0}^{\infty} \beta_n^t \, u(t-n) + P[x(t)|H(-\infty)] \tag{6.23}$$

where $\{\beta_n^t : n \geq 0\}$ is a sequence of $p \times p$ matrices of real numbers satisfying

$$\sum_{n=o}^{\infty} \text{trace}\left\{\beta_n^t \, E[u(t-n)\,u(t-n)']\beta_n^{t\prime}\right\} < \infty \,. \tag{6.24}$$

Representation (6.23) is referred to as the Wold decomposition of a stochastic process. Notice the coefficients of this representation depend on the time period t. In the special case that $\{x(t)\}$ is a covariance stationary stochastic process, the coefficients turn out to be time invariant. In this case, (6.23) is known as a "Wold representation for $x(t)$."

It is convenient to mention two (not exhaustive) categories of stochastic processes:

Definition 6.1: The stochastic process $\{x(t) : -\infty < t < +\infty\}$ is linearly regular if $H(-\infty) = \{0\}$.

Definition 6.2: The stochastic process $\{x(t) : -\infty < t < +\infty\}$ is linearly deterministic if $H(t) = H(-\infty)$ for all t.

When a stochastic process is linearly regular, the term $P[x(t)|H(-\infty)]$ drops out of the Wold decomposition. When a stochastic process is linearly deterministic,

$$x(t) = P[x(t)|H(t-n)] \tag{6.25}$$

for all n so that $x(t)$ can be forecast perfectly given past information. These two types of stochastic processes represent two extremes from the standpoint of linear prediction theory. It turns out that any stochastic process can be represented as the sum of a linearly regular process and a linearly deterministic process. Let

$$x^*(t) = x(t) - P[x(t)|H(-\infty)] \,, \tag{6.26}$$

and

$$x^+(t) = P[x(t)|H(-\infty)] \,. \tag{6.27}$$

Clearly,

$$x(t) = x^*(t) + x^+(t) \,. \tag{6.28}$$

We shall show that $\{x^*(t) : -\infty < t < +\infty\}$ is linearly regular and $\{x^+(t) : -\infty < t < +\infty\}$ is linearly deterministic.

Lemma 6.2: The process $\{x^*(t) : -\infty < t < +\infty\}$ is linearly regular.

Proof: Let $H^*(t)$ be defined in the same manner as $H(t)$ except with $x(t)$ replaced by $x^*(t)$. Also, let

$$H^*(-\infty) = \bigcap_{n=1}^{\infty} H^*(t-n) \,. \tag{6.29}$$

Now $x^*(t)$ is orthogonal to $H(-\infty)$ for all t which implies that $H^*(-\infty)$ is orthogonal to $H(-\infty)$. However, $x^*(t-n)$ is contained in $H(t)$ for all $n \geq 0$ so that $H^*(t)$ is contained in $H(t)$. Consequently, $H^*(-\infty)$ is contained in $H(-\infty)$. Therefore, $H^*(-\infty)$ is orthogonal to itself. The only subspace that is orthogonal to itself is $H^*(-\infty) = \{0\}$. ∎

Lemma 6.3: The process $\{x^+(t) : -\infty < t < +\infty\}$ is linearly deterministic.

Proof: Let $H^+(t)$ be defined in the same manner as $H(t)$ except with $x(t)$ replaced by $x^+(t)$. Also, let

$$H^+(-\infty) = \bigcap_{n=1}^{\infty} H^+(t-n) \,. \tag{6.30}$$

Now $x^+(t)$ is in $H(-\infty)$ for all t which implies that $H^+(t)$ is contained in $H(-\infty)$ for all t. Let h be any element of $H(-\infty)$. Then there exists a sequence $\{h_n : n \geq 1\}$ converging to h such that

$$h_n = \alpha_0^n \cdot x(t) + \alpha_1^n \cdot x(t-1) + \ldots + \alpha_n^n \cdot x(t-n) \tag{6.31}$$

for some vectors $\alpha_0^n, \alpha_1^n, \ldots, \alpha_n^n$ in $\mathbf{R}^p$. Note that

$$E[(h_n - h)^2] \geq E\left\{P[h_n - h|H(-\infty)]^2\right\} \tag{6.32}$$

so that $\{P[h_n|H(-\infty)] : n \geq 1\}$ converges to $P[h|H(-\infty)] = h$. However, $P[h_n|H(-\infty)]$ is in $H^+(t)$ which proves that $H^+(t)$ contains $H(-\infty)$. Therefore $H^+(t) = H(-\infty) = H^+(-\infty)$ for all t. ∎

In many applications in economics, including all of those contained in this book attention is restricted to processes that are linearly regular. In making such a restriction, we have a decomposition such as (6.28) in mind. Then our attention is focused on the portion of the forecasting problems that can be related to the one-step ahead forecast errors in current and past values of $x(t)$. This portion of the prediction problems is dynamic in the sense that information is accumulating over time.

Appendix

In this appendix we prove that the space L^2 is complete. This is a special case of the Riesz-Fischer Theorem. Our proof of this theorem uses the Monotone and Dominated Convergence Theorems.

Theorem A.1: Suppose $\{y_n : n \geq 1\}$ is a Cauchy sequence in L^2. Then there exists a random variable y_o in L^2 such that $\{y_n : n \geq 1\}$ converges to y_o.

Proof: Since $\{y_n : n \geq 1\}$ is Cauchy, for each $k \geq 1$ there is an n_k such that

$$(A.1) \qquad E[(y_m - y_{n_k})^2]^{\frac{1}{2}} \leq 1/2^k \quad \text{for all } m \geq n_k \ .$$

Therefore,

$$(A.2) \qquad \sum_{j=1}^{\infty} E[(y_{n_{j+1}} - y_{n_j})^2]^{\frac{1}{2}} \leq 1 \ .$$

By the Triangle Inequality,

$$(A.3) \qquad E\Big\{\Big[\sum_{j=1}^{k} |y_{n_{j+1}} - y_{n_j}|\Big]^2\Big\}^{\frac{1}{2}} \leq 1 \qquad \text{for all } k \geq 1 \ .$$

For $k \geq 1$, let

$$(A.4) \qquad y_k^* = |y_{n_1}| + \sum_{j=1}^{k} |y_{n_{j+1}} - y_{n_j}| \ .$$

The sequence $\{y_k^* : k \geq 1\}$ is nondecreasing. Consequently, y_o^* given by

$$(A.5) \qquad y_o^* = \lim_{k\to\infty} y_k^*$$

is a well-defined random variable that is possibly infinite-valued. By the Triangle Inequality and (A.3),

$$(A.6) \qquad [E(y_k^*)^2]^{\frac{1}{2}} \leq E[(y_{n_1})^2]^{\frac{1}{2}} + 1 \ ,$$

or equivalently,

$$(A.7) \qquad E[(y_k^*)^2] \leq \left\{E[(y_{n_1})^2]^{\frac{1}{2}} + 1\right\}^2 \quad \text{for all } k \geq 1 \ .$$

Thus, by the Monotone Convergence Theorem,

$$(A.8) \qquad \begin{aligned} E[(y_o^{*2})] &= \lim_{k\to\infty} E[(y_k^*)^2] \\ &\leq \left\{E[(y_{n_1})^2]^{\frac{1}{2}} + 1\right\}^2 \end{aligned}$$

so that y_o^* is finite with probability one and is in L^2.

Now,

$$(A.9) \qquad y_{n_k} = y_{n_1} + \sum_{j=1}^{k} (y_{n_{j+1}} - y_{n_j}) \ .$$

Also, absolutely summable sequences of real numbers are summable. Thus, y_o given by

$$(A.10) \qquad y_o = \begin{cases} \lim_{k\to\infty} y_{n_k} & \text{if } y_o^* \text{ is finite} \\ 0 & \text{otherwise} \ . \end{cases}$$

is a well-defined random variable since y_o^* is a random variable. Equalities (A.4) and (A.5) imply that

$$(A.11) \qquad |y_{n_k}| \leq y_o^* \quad \text{for all } k \geq 1 \ ,$$

which in turn implies that

$$(A.12) \qquad |y_o| \leq y_o^* \ .$$

Consequently, y_o is in L^2. Inequalities (A.11) and (A.12) and the Triangle Inequality guarantee that

$$(A.13) \qquad (y_o - y_{n_k})^2 \leq 4(y_o^*)^2 \ .$$

It follows from the Dominated Convergence Theorem that

$$(A.14) \qquad \lim_{k\to\infty} E[(y_o - y_{n_k})^2] = 0$$

since

$$(A.15) \qquad \lim_{k\to\infty} y_{n_k} = y_o$$

on a set of probability measure one. Thus, we have proved that $\{y_{n_k} : k \geq 1\}$ converges to y_o in L^2.

To complete this proof, we must show that $\{y_n : n \geq 1\}$ converges to y_o in L^2. Given $\delta > 0$, there is k such that

$$E[(y_m - y_{n_k})^2]^{\frac{1}{2}} < \delta/2 \tag{A.16}$$

for all $m \geq n_k$ and

$$E[(y_{n_k} - y_o)^2]^{\frac{1}{2}} < \delta/2 \ . \tag{A.17}$$

By the Triangle Inequality,

$$E[(y_m - y_o)^2]^{\frac{1}{2}} < \delta \quad \text{for } m \geq n_k \ . \quad \blacksquare \tag{A.18}$$

Notes

1. We follow the usual convention of viewing the equivalence class of random variables that are equal almost surely as a unique random variable.
2. This notion of convergence is induced by the norm, $||y|| = E(y^2)^{\frac{1}{2}}$ on the linear space L^2.
3. The definition of continuity given here only considers the behavior of the function π in the neighborhood of zero. However, since c is linear, continuity at a single point immediately translates into continuity at all points.
4. This lemma is a special case of the Riesz Representation Theorem.
5. We take the notation $E(z|x)$ to mean the expectation conditioned on the sigma algebra generated by x.
6. The unique extension of measures from algebras to sigma algebras follows from the Carathéodory Theorem. See Halmos (1950) or Royden (1968) for a discussion of this result.
7. Recall that a projection is defined only up to an equivalence class of random variables that are almost surely equal. Some (but not necessarily all) members of this equivalence class will be measurable with respect to B_o.
8. By $P[x(t)|H(t-1)]$ we mean the vector containing the projections of components $x(t)$ onto $H(t-1)$.

3

Exact Linear Rational Expectations Models: Specification and Estimation

by Lars Peter HANSEN and Thomas J. SARGENT

Introduction

A distinguishing characteristic of econometric models that incorporate rational expectations is the presence of restrictions across the parameters of different equations. These restrictions emerge because people's decisions are supposed to depend on the stochastic environment which they confront. Consequently, equations describing variables affected by people's decisions inherit parameters from the equations that describe the environment. As it turns out, even for models that are linear in the variables, these cross-equation restrictions on the parameters can be complicated and often highly nonlinear.

This paper proposes a method for conveniently characterizing cross-equation restrictions in a class of linear rational expectations models, and also indicates how to estimate statistical representations satisfying these restrictions. For most of the paper, we restrict ourselves to models in which there is an *exact* linear relation across forecasts of future values of one set of variables and current and past values of some other set of variables. The key requirement is that all of the variables entering this relation must be observed by the econometrician. While probably only a minority of rational expectations models belong to this class, it does contain interesting models that have been advanced to study the term structure of interest rates, stock prices, consumption and permanent income, the dynamic demand for factors of production, and many other subjects.

It is useful to compare the class of exact models with the class studied by Hansen and Sargent (1980a). The differences lie entirely in the interpretations of the *error terms* in the equations that are permitted. In Hansen and Sargent (1980a), random processes which the econometrician treats as disturbances in decision rules can have a variety of