Wold Decomposition

Christopher A. Sims
Princeton University
sims@princeton.edu

March 29, 2015
Preliminaries

If we have a set (possibly countably infinite) of random variables \( \{X_i\} \), the set of all finite linear combinations of them forms a linear space.

We can define an inner product, and thus a norm on that space as \( \langle X, Y \rangle = \text{Cov}(X, Y) \). Then defining the distance between \( X \) and \( Y \) as \( \|X - Y\| \), our space is a metric space. We can complete the metric space by extending it to include all limits of Cauchy sequences. That is, if \( \{Z_i, i = 1, \ldots, \infty\} \) has the property that \( \|Z_m - Z_n\| \rightarrow \infty \) as \( m, n \rightarrow \infty \), then \( Z_\infty = \lim_{i \rightarrow \infty} Z_i \) is in the space.


Projections

Suppose $G$ is a complete linear subspace of $H$, with a Hilbert space (i.e., innerproduct and norm defined) structure. We can define the operator $\mathcal{E}$ by

$$\mathcal{E}[X \mid H] = Z \in H \text{ that minimizes } \|X - Z\| \, .$$

It is not hard to prove that such a $Z$ must always exist and is unique.

If $G_1$ and $G_2$ are two subspaces of $H$ such that $\langle X, Y \rangle = 0$ whenever $X \in G_1$ and $Y \in G_2$, we say that $G_1$ and $G_2$ are orthogonal, or $G_1 \perp G_2$. In that case it is not hard to show that $\mathcal{E}[X \mid G_1, G_2] = \mathcal{E}[X \mid G_1] + \mathcal{E}[X \mid G_2]$.

It is always true that $X - \mathcal{E}[X \mid G] \perp G$. 
A finite variance stochastic process and its predictive projections

Now let $Y_t, t = -\infty, \ldots, \infty$ be a vector valued stochastic process. That is, each $Y_t$ is an $n$-dimensional random vector, and the probability law of the stochastic process specifies mutually consistent joint distributions for any finite collection of the $\{Y_t\}$ variables.

Let $H_t$ be the complete metric space generated by $\{Y_s, s \leq t\}$.

We can always project $Y_t$ on $H_{t-1}$ and express the gap between the two as $\varepsilon_t = Y_t - \mathcal{E}[Y_t \mid H_{t-1}]$.

$\varepsilon_t$ is the innovation in $Y_t$. 
Recursive projections, Wold representation

$H_t$ for any $t$ is the same as the space spanned by $\varepsilon_t, H_{t-1}$. (This is obvious if you think about the definitions.) Therefore we can write

$$y_t = \varepsilon_t + \mathcal{E}[y_t \mid \varepsilon_{t-1}] + \mathcal{E}[y_t \mid H_{t-2}] = \varepsilon_t + A_1\varepsilon_{t-1} + \mathcal{E}[y_t \mid H_{t-2}].$$

The $A_1$ is a square matrix of coefficients. Since $\varepsilon_t$ is of dimension $n$ the space it spans is just the space of linear combinations of elements of the $\varepsilon_t$ vector, so each element of the $\mathcal{E}[Y_t \mid \varepsilon_{t-1}]$ vector is a linear combination of elements of $\varepsilon_{t-1}$, given by a row of $A_1$. 
Repeating this $T$ times, we get

$$y_t = \sum_{s=0}^{T-1} A_s \varepsilon_{t-s} + \mathcal{E}[Y_t \mid H_{t-T}] = \tilde{y}_t^T + \bar{y}_t^T.$$
Taking limits

Var($\tilde{y}_t^T$) is increasing in $T$ and is bounded above by Var($y_t$). (These are matrices, so we mean by “increasing” that their differences are positive semi-definite, which implies their diagonal elements are non-negative.) It is therefore a Cauchy sequence and has a limit we call simply $\tilde{y}_t$. This is the linearly regular piece of $y_t$.

Var($\bar{y}_t^T$) is decreasing in $T$ and bounded below by zero, so it too is Cauchy and has a limit, which we call $\bar{y}_t$.

Note that $\bar{y}_t$ is in $H_{t-T}$ for every $T$, so $\mathcal{E}[\bar{y}_t \mid H_{t-T}] = \bar{y}_t$, for every $T$. In other words, $\bar{y}_t$ can be forecast arbitrarily well from data on $y_s$ before time $t - T$, no matter how far back in the past these data are. So we call this part the linearly deterministic part.
Stationarity

If the $y$ process is stationary, meaning the joint distribution of $\{X_1, \ldots, X_m\}$ is the same as that of $\{X_{s+1}, \ldots, X_{s+m}\}$ for any $s$, no matter what $m$ we start with, then our decomposition above produces the same $A_s$ sequence, no matter what $t$ we pick for $y_t$. Furthermore $\tilde{y}_t$, $\bar{y}_t$, and $\varepsilon_t$ will then also be stationary.

A stationary process is called linearly regular if its linearly deterministic component is zero. It is called linearly deterministic if its linearly regular component is zero. (For a non-stationary process, we can do the same decomposition at any $t$, but the component processes could have variance zero for some dates and not for others.)

Examples of LR stationary processes: i.i.d. $N(0, I)$; stationary AR(1) ($\mathcal{E}[y_t | H_{t-1}] = \rho y_{t-1}$), $\text{Var} \varepsilon_t$ constant)
Examples of LD stationary processes: $y_t = \sin(t + \theta)$, $\theta \sim U(0, 2\pi)$; $y_t \sim N(0, 1)$, $y_t \equiv y_{t-1}$; $y_t = \sum_s e^{-(t-s)^2} \nu_{t-s}$, $\nu$ i.i.d. $N(0, 1)$. 