Notes on Factor Models and the Hicks Lecture Model with Normal Random Variables

James J. Heckman
The University of Chicago

Econ 312
Spring 2021
This draft, April 23, 2021

Factor Models: Traditionally work with Covariance Information

One Factor Models

$$E(\theta) = 0; \qquad E(\varepsilon_i) = 0; \; i = 1, \ldots, 5$$

$$\begin{aligned} &Y_1 = \alpha_1\theta + \varepsilon_1, \quad Y_2 = \alpha_2\theta + \varepsilon_2, \quad Y_3 = \alpha_3\theta + \varepsilon_3, \\ &Y_4 = \alpha_4\theta + \varepsilon_4, \quad Y_5 = \alpha_5\theta + \varepsilon_5, \quad \varepsilon_i \perp\!\!\!\perp \varepsilon_j \end{aligned}$$

For $T \geqslant 3$, can identify the model with on normalization.

$$\begin{aligned} Cov\left(Y_1, Y_2\right) &= \alpha_1 \alpha_2 \sigma_\theta^2 \\ Cov\left(Y_1, Y_3\right) &= \alpha_1 \alpha_3 \sigma_\theta^2 \\ Cov\left(Y_2, Y_3\right) &= \alpha_2 \alpha_3 \sigma_\theta^2 \end{aligned}$$

Normalize $\alpha_1 = 1$

$$\frac{Cov\left(Y_2, Y_3\right)}{Cov\left(Y_1, Y_2\right)} = \alpha_3$$

$\therefore$ We know σ_θ^2 from $Cov\left(Y_1, Y_2\right)$. From $Cov\left(Y_1, Y_j\right), j = 3, 4, 5,$ we know

$$\alpha_3, \alpha_4, \alpha_5.$$

Can get the variances of the ε_i from variances of the Y_i

$$Var(Y_i) = \alpha_i^2 \sigma_\theta^2 + \sigma_{\varepsilon_i}^2.$$

If $T = 2$, all we can identify is $\alpha_1 \alpha_2 \sigma_\theta^2$, even with the normalization.

If $\alpha_1 = 1$, $\sigma_\theta^2 = 1$, we identify α_2.

2 Factors:

Assume $\theta_1 \perp\!\!\!\perp \theta_2$

$$\varepsilon_i \perp\!\!\!\perp \varepsilon_j \quad \forall i, j$$

Normalize:

$$\begin{aligned}
Y_1 &= \alpha_{11}\theta_1 + (0)\theta_2 + \varepsilon_1 \\
Y_2 &= \alpha_{21}\theta_1 + (0)\theta_2 + \varepsilon_2 \\
Y_3 &= \alpha_{31}\theta_1 + \alpha_{32}\theta_2 + \varepsilon_3 \\
Y_4 &= \alpha_{41}\theta_1 + \alpha_{42}\theta_2 + \varepsilon_4 \\
Y_5 &= \alpha_{51}\theta_1 + \alpha_{52}\theta_2 + \varepsilon_5
\end{aligned}$$

Let $\alpha_{11} = 1$, $\alpha_{32} = 1$.

$$\begin{aligned} Cov\left(Y_1, Y_2\right) &= \alpha_{21}\sigma^2_{\theta_1} \\ Cov\left(Y_1, Y_3\right) &= \alpha_{31}\sigma^2_{\theta_1} \\ Cov\left(Y_2, Y_3\right) &= \alpha_{21}\alpha_{31}\sigma^2_{\theta_1} \end{aligned}$$

$$\text{Form ratio of } \frac{Cov\left(Y_2, Y_3\right)}{Cov\left(Y_1, Y_2\right)} = \alpha_{31},$$

$\therefore$ we identify $\alpha_{31}, \alpha_{21}, \sigma^2_{\theta_1}$, as before.

$$\begin{aligned} Cov\left(Y_1, Y_4\right) &= \alpha_{41}\sigma^2_{\theta_1}, \\ &\vdots \\ Cov\left(Y_1, Y_k\right) &= \alpha_{k1}\sigma^2_{\theta_1} \end{aligned}$$

$\therefore$ we identify α_{k1} for all k and $\sigma^2_{\theta_1}$.

$$\begin{aligned}
Cov\,(Y_3, Y_4) - \alpha_{31}\alpha_{41}\sigma^2_{\theta_1} &= \alpha_{42}\sigma^2_{\theta_2} \\
Cov\,(Y_3, Y_5) - \alpha_{31}\alpha_{51}\sigma^2_{\theta_1} &= \alpha_{52}\sigma^2_{\theta_2} \\
Cov\,(Y_4, Y_5) - \alpha_{41}\alpha_{51}\sigma^2_{\theta_1} &= \alpha_{52}\alpha_{42}\sigma^2_{\theta_2},
\end{aligned}$$

By same logic,

$$\frac{Cov\,(Y_4, Y_5) - \alpha_{41}\alpha_{51}\sigma^2_{\theta_1}}{Cov\,(Y_3, Y_4) - \alpha_{31}\alpha_{41}\sigma^2_{\theta_1}} = \alpha_{52}$$

$\therefore$ get $\sigma^2_{\theta_2}$ and the factor "2" loadings.

If we have dedicated measurements of factor, do not need a normalization on Y. They provide a natural scale. Assume $\theta_1 \perp\!\!\!\perp \theta_2$ (testable)

$$\begin{aligned} M_1 &= \theta_1 + \varepsilon_{1M} \\ M_2 &= \theta_2 + \varepsilon_{2M} \end{aligned}$$

$$\begin{aligned} Cov\left(Y_1, M_1\right) &= \alpha_{11}\sigma^2_{\theta_1} \\ Cov\left(Y_2, M_1\right) &= \alpha_{21}\sigma^2_{\theta_1} \\ Cov\left(Y_3, M_1\right) &= \alpha_{31}\sigma^2_{\theta_1} \\ Cov\left(Y_1, Y_2\right) &= \alpha_{11}\alpha_{21}\sigma^2_{\theta_1}, \\ Cov\left(Y_1, Y_3\right) &= \alpha_{11}\alpha_{31}\sigma^2_{\theta_1}, \qquad \therefore \alpha_{21}\sigma^2_{\theta_1}, \end{aligned}$$

$\therefore$ We can get $\alpha_{21}, \sigma^2_{\theta_1}$ and the other factors.

General Case

$$\underset{T\times 1}{Y} = \mu + \underset{T\times K}{\Lambda}\underset{K\times 1}{\theta} + \underset{T\times 1}{\varepsilon}$$

θ are factors, ε uniquenesses

$$E(\varepsilon) = 0$$

$$Var(\varepsilon\varepsilon') = D = \begin{pmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma^2_{\varepsilon_T} \end{pmatrix}$$

$$E(\theta) = 0$$

$$Var(Y) = \Lambda\Sigma_\theta\Lambda' + D \qquad \Sigma_\theta = E(\theta\theta')$$

The only source of information on Λ and Σ_θ is from the covariances.

Associated with each variance of Y_i is a $\sigma^2_{\varepsilon_i}$.

Each variance contributes one new parameter.

How many unique covariance terms do we have?

$\frac{T(T-1)}{2}$ This is the data.

We have T uniquenesses; TK elements of Λ.

$\dfrac{K\left(K-1\right)}{2}$ elements of Σ_{θ}.

$\dfrac{K\left(K-1\right)}{2}+TK$ parameters $(\Sigma_{\theta},\Lambda)$.

Observe that if we multiply Λ by an orthogonal matrix C, $(CC'=I)$, we have

$$Var\left(Y\right)=\Lambda C\left[C'\Sigma_{\theta}C\right]C'\Lambda'+D$$

C is a "rotation". Cannot separate ΛC from Λ.

Model not identified against orthogonal transformations in the general case.

Some common assumptions:

(i) $\theta_i \perp\!\!\!\perp \theta_j, \forall\ i \neq j$

$$\Sigma_\theta = \begin{pmatrix} \sigma^2_{\theta_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\theta_2} & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \sigma^2_{\theta_K} \end{pmatrix}$$

joined with

(ii)

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \alpha_{21} & 0 & 0 & 0 & \cdots & 0 \\ \alpha_{31} & 1 & 0 & 0 & \cdots & 0 \\ \alpha_{41} & \alpha_{42} & 0 & 0 & \cdots & 0 \\ \alpha_{51} & \alpha_{52} & 1 & 0 & \cdots & 0 \\ \alpha_{61} & \alpha_{62} & \alpha_{63} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & 1 & & \vdots \end{pmatrix}$$

We know that we can identify of the Λ, Σ_θ parameters.

$$\underbrace{\frac{K(K-1)}{2} + TK}_{\text{\# of free parameters}} \leq \underbrace{\frac{T(T-1)}{2}}_{\text{data}}$$

"Ledermann Bound"

Generalized Roy Model with Factor Structure

Generalized Roy versions of college choice model:

$$M = \mu(X) + \theta_1 \alpha_{1,M} + \theta_2 \alpha_{2,M} + \varepsilon_M$$

(Measurement: A test score equation)

$$\left.\begin{aligned} Y_1^1 &= \mu_1^1(X) + \theta_1 \alpha_{1,1}^1 + \theta_2 \alpha_{2,1}^1 + \varepsilon_1^1 \\ Y_2^1 &= \mu_2^1(X) + \theta_1 \alpha_{1,2}^1 + \theta_2 \alpha_{2,2}^1 + \varepsilon_2^1 \end{aligned}\right\} \text{College earnings}$$

$$\left.\begin{aligned} Y_1^0 &= \mu_1^0(X) + \theta_1 \alpha_{1,1}^0 + \theta_2 \alpha_{2,1}^0 + \varepsilon_1^0 \\ Y_2^0 &= \mu_2^0(X) + \theta_1 \alpha_{1,2}^0 + \theta_2 \alpha_{2,2}^0 + \varepsilon_2^0 \end{aligned}\right\} \text{High School earnings}$$

Cost

$$C = Z\gamma + \theta_1 \alpha_{1C} + \theta_2 \alpha_{2C} + \varepsilon_C$$

Decision Rule Under Perfect Certainty:
(Assume Interest Rate $r = 0$)

$$
\begin{aligned}
I &= \mu_1^1(X) + \mu_2^1(X) + \theta_1\left(\alpha_{1,1}^1 + \alpha_{1,2}^1\right) \\
&\quad + \theta_2\left(\alpha_{2,1}^1 + \alpha_{2,2}^1\right) + \varepsilon_1^1 + \varepsilon_2^1 \\
&\quad - \left[\begin{array}{c} \mu_1^0(X) + \mu_2^0(X) + \theta_1\left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) \\ + \theta_2\left(\alpha_{2,1}^0 + \alpha_{2,2}^0\right) + \varepsilon_1^0 + \varepsilon_2^0 \end{array}\right] \\
&\quad - Z\gamma - \theta_1\alpha_{1C} - \theta_2\alpha_{2C} - \varepsilon_C \\
&= \mu_1^1(X) + \mu_2^1(X) - \left[\mu_1^0(X) + \mu_2^0(X) + Z\gamma\right] \\
&\quad + \theta_1\left[\left(\alpha_{1,1}^1 + \alpha_{1,2}^1\right) - \left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) - \alpha_{1C}\right] \\
&\quad + \theta_2\left[\left(\alpha_{2,1}^1 + \alpha_{2,2}^1\right) - \left(\alpha_{2,1}^0 + \alpha_{2,2}^0\right) - \alpha_{2C}\right] \\
&\quad + \left(\varepsilon_1^1 + \varepsilon_2^1\right) - \left(\varepsilon_1^0 + \varepsilon_2^0\right) - \varepsilon_C
\end{aligned}
$$

In Reduced Form

$$I = \varphi(X, Z) + \alpha_{I,1}\theta_1 + \alpha_{I,2}\theta_2 + \varepsilon_I.$$

Set $U_I = \alpha_{I,1}\theta_1 + \alpha_{I,2}\theta_2 + \varepsilon_I$.

$\therefore$ we can write

$$\begin{aligned}
Y_1^1 &= \mu_1^1(X) + U_1^1 \\
Y_2^1 &= \mu_2^1(X) + U_2^1 \\
Y_1^0 &= \mu_1^0(X) + U_1^0 \\
Y_2^0 &= \mu_2^0(X) + U_2^0
\end{aligned}$$

U_1^1, U_2^1 etc. match the error terms previously shown.

$$\begin{aligned}
U_1^1 &= \theta_1\alpha_{1,1}^1 + \theta_2\alpha_{2,1}^1 + \varepsilon_1^1 \text{ etc.} \\
U_M &= \theta_1\alpha_{1,M} + \theta_2\alpha_{2,M} + \varepsilon_M
\end{aligned}$$

$$E\left(Y_1^1 \mid X, Z, I > 0\right) = \mu_1^1(X) + \frac{Cov\left(U_1^1, I\right)}{Var(I)}\lambda()$$

Using notes on the Roy model, we can identify beside the means,

$\mu_1^1(X), \mu_2^1(X), \mu_2^0(X), \mu_2^0(X)$, the following parameters:

$$\begin{array}{l}
Cov\left(U_1^1, U_2^1\right), Var\left(U_1^1\right), Var\left(U_2^1\right) \\
Cov\left(U_1^1, U_M\right), Cov\left(U_2^1, U_M\right), Var\left(U_M\right) \\
Cov\left(U_1^0, U_2^0\right), Var\left(U_1^0\right), Var\left(U_2^0\right) \\
Cov\left(U_1^0, U_M\right), Cov\left(U_2^0, U_M\right)
\end{array}$$

Normal Case: (θ, ε) normal.

$$(\theta, \varepsilon) \perp\!\!\!\perp (X, Z)$$

$$\begin{aligned} & \Pr\left(S = 1 \mid X, Z, \theta_1, \theta_2\right) \\ = \; & \Phi\left[\frac{1}{\sigma_{\varepsilon_I}}\left[\begin{array}{c} \mu_1^1\left(X\right) + \mu_2^1\left(X\right) - \left[\mu_1^0\left(X\right) + \mu_2^0\left(X\right)\right] \\ -\left[Z\gamma + \theta_1\alpha_{I,1} + \theta_2\alpha_{I,2}\right] \end{array}\right]\right] \end{aligned}$$

Fact:

If $S = \mathbf{1}\left[X\beta + \theta > V\right]$, $X \perp\!\!\!\perp (\theta, V)$
θ, V are normal, $\theta \perp\!\!\!\perp V$, $E(\theta) = 0, E(V) = 0$

$$\begin{aligned} \Pr\left(S = 1 \mid X, \theta\right) &= \Phi\left(\frac{X\beta + \theta}{\sigma_V}\right) \\ \Pr\left(S = 1 \mid X\right) &= \Phi\left(\frac{X\beta}{\left(\sigma_V^2 + \sigma_\theta^2\right)^{\frac{1}{2}}}\right) \end{aligned}$$

Why?
$S = \mathbf{1}\left[X\beta > V - \theta\right]$.
Rest follows from independence (between $V - \theta$, and X, and normality).

Unconditional Probability: (Not conditional on Factors)

$$\Pr(S = 1 \mid X, Z) = \Phi\left[\frac{\mu_1^1(X) + \mu_2^1(X) - [\mu_1^0(X) + \mu_2^0(X)] - Z\gamma}{\left(\sigma_{\varepsilon_I}^2 + \alpha_{I,1}^2\sigma_{\theta_1}^2 + \alpha_{I,2}^2\sigma_{\theta_2}^2\right)^{1/2}}\right]$$

Observe that if we know $\mu_1^1(X), \mu_2^1(X), \mu_1^0(X), \mu_2^0(X)$ we know

$$\left[\mu_1^1(X) + \mu_2^1(X)\right] - \left[\mu_1^0(X) + \mu_2^0(X)\right].$$

If $Z\gamma$ not perfectly collinear with this term (e.g. one X or more not in Z) we can identify

$$\left(\sigma_{\varepsilon_I}^2 + \alpha_{I,1}^2\sigma_{\theta_1}^2 + \alpha_{I,2}^2\sigma_{\theta_2}^2\right)^{\frac{1}{2}}$$

$\therefore$ we also identify γ (get absolute scale on costs).

Suppose agents do not know θ_2 or the future $\varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^0, \varepsilon_2^0$ but know ε_c and θ_1.

Then if what they know is set at mean zero, (they use rational expectations in a linear decision rule) and their mean forecast is the population mean,

$$\sigma^2_{\varepsilon_I} = \sigma^2_{\varepsilon_c}$$

and $\alpha_{I,2} = 0$, what can we identify?

What information do we have about covariances?

Suppose we have two dedicated measurement systems for θ_1 and θ_2. We normalize the First loading as a convention.

$$\left.\begin{aligned} M_1^1 &= \theta_1 &+&\quad \varepsilon_{1,M}^1 \\ M_2^1 &= \alpha_{2,M}^1\theta_1 &+&\quad \varepsilon_{2,M}^1 \\ M_3^1 &= \alpha_{3,M}^1\theta_1 &+&\quad \varepsilon_{3,M}^1 \end{aligned}\right\} \text{ Cognitive Ability}$$

$$\left.\begin{aligned} M_1^2 &= \theta_2 &+&\quad \varepsilon_{1,M}^2 \\ M_2^2 &= \alpha_{2,M}^2\theta_2 &+&\quad \varepsilon_{2,M}^2 \\ M_3^2 &= \alpha_{3,M}^2\theta_2 &+&\quad \varepsilon_{3,M}^2 \end{aligned}\right\} \text{ Noncognitive Ability}$$

Observe from M^1 system we get

$$Var\left(\theta_1\right), \alpha^1_{2,M}, \alpha^1_{3,M}$$

From M^2 system we get

$$Var\left(\theta_2\right), \alpha^2_{2,M}, \alpha^2_{3,M}$$

Then

$$
\begin{aligned}
Cov\left(U_1^1, M_1^1\right) &= \alpha_{1,1}^1 \sigma_{\theta_1}^2 \\
Cov\left(U_2^1, M_1^1\right) &= \alpha_{1,2}^1 \sigma_{\theta_1}^2
\end{aligned}
$$

$\therefore$ we get all of the factor loadings in Y^1 on θ_1.

Using M_1^2 we get $\alpha_{2,1}^1, \alpha_{2,2}^1$ and we get variances of uniquenesses $Var\left(\varepsilon_1^1\right), Var\left(\varepsilon_2^1\right)$.

By similar reasoning, we get

$$
\begin{aligned}
&\alpha_{1,1}^0, \alpha_{2,1}^0, \alpha_{1,2}^0, \alpha_{2,2}^0 \\
&Var\left(\varepsilon_1^0\right), Var\left(\varepsilon_2^1\right)
\end{aligned}
$$

Observe that from

$$Cov\left(I, M_1^1\right) = \sigma_{\theta_1}^2 \left[\alpha_{1,1}^1 + \alpha_{1,2}^1 - \left(\alpha_{1,1}^0 + \alpha_{1,2}^0\right) - \alpha_{1,C}\right]$$

$\therefore$ We can get α_{1C}, since we know all other terms on the right hand side by the previous reasoning.

From

$$Cov\left(I, M_1^2\right) = \sigma_{\theta_2}^2 \left[\alpha_{2,1}^1 + \alpha_{2,2}^1 - \left(\alpha_{2,1}^0 + \alpha_{2,2}^0\right) - \alpha_{2,C}\right]$$

we can get α_{2C}.

From $\Pr\left(S = 1 \mid X, Z\right)$, we can identify $\sigma_{\varepsilon_I}^2$ using previous reasoning

Therefore we can identify everything in the model if there is one X not in Z since we can identify the terms in the numerator.

Can we test the model?

In the notation of the Hicks lecture notes, we have for a test of whether θ_2 belongs in the model

$$
\begin{aligned}
&\Pr\left(S = 1 \mid X, Z\right) \\
&\qquad = \Phi\left[\frac{\mu_1^1\left(X\right) + \mu_2^1\left(X\right) - \left[\mu_1^0\left(X\right) + \mu_2^0\left(X\right)\right] - Z\gamma}{\left(\sigma_{\varepsilon_I}^2 + \alpha_{I,1}^2\sigma_{\theta_1}^2 + \alpha_{I,2}^2\sigma_{\theta_2}^2\Delta_{\theta_2}\right)^{\frac{1}{2}}}\right]
\end{aligned}
$$

Apparently, we can test the null

$$H_0 : \Delta_{\theta_2} = 0$$

$\therefore$ we can test if θ_2 components enter or not.

The problem with this test is that if $\sigma^2_{\varepsilon_c} \neq 0$, we can always adjust its value to fit the model perfectly well.
(This problem vanishes if we assume a pure Roy model (so $\sigma^2_{\varepsilon_c} = 0$).)

Notice, however, that we can also tolerate $\gamma \neq 0$ so long as $\sigma^2_{\varepsilon_c} = 0$.

Correct idea of the correct test:
Form

$$Cov\left(\frac{I}{\sigma_I}, U_1^1\right) = \frac{\sigma^2_{\theta_1}}{\sigma_I}\alpha^1_{1,1}\left[\alpha^1_{1,1} + \alpha^1_{1,2} - \left(\alpha^0_{1,1} + \alpha^0_{1,2}\right) - \alpha_{1,C}\right]$$
$$+ \Delta_{\theta_2}\sigma^2_{\theta_2}\alpha^1_{1,2}\left[\alpha^1_{1,1} + \alpha^1_{1,2} - \left(\alpha^0_{1,1} + \alpha^0_{1,2}\right) - \alpha_{1,C}\right]$$

$\therefore$ we can compute the test under the null.

Under the null that $\Delta_{\theta_2} = 0$, we can identify $\sigma^2_{\varepsilon_c}$

$\therefore$ we construct a test under null:

$$Cov\left(\frac{I}{\sigma_I}, U_1^1\right) - \frac{\sigma^2_{\theta_1}\alpha^1_{1,1}\left[\alpha^1_{1,1} + \alpha^1_{1,2} - \left(\alpha^0_{1,1} + \alpha^0_{1,2}\right) - \alpha_{1,c}\right]}{\sigma_I} = 0$$

We know both terms under the null. Departures are evidence that agents know θ_2.

If the agent knows θ_1 but not θ_2 and sets

$$E\left(\theta_2\right) = 0.$$

Justified by linearity of the criterion and rational expectations, assuming $E\left(\theta_2 \mid \mathcal{I}_0\right) = 0$.

Then we have that the test amounts to deciding

- Which model fits the data better?

Average effect (we estimate the average probability):

$$\int \Pr\left(S = 1 \mid X, Z, \theta_1, \Delta_{\theta_2}, \theta_2\right) f\left(\theta_1\right) f\left(\theta_2\right) d\theta.$$

(we test $\Delta_{\theta_2} = 0$)
This is what is done in the Hicks lecture.