Factor Models

Extract from "Notes on 'Econometric Evaluation of Social Programs Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation' Handbook of Econometrics, Vol. 6B, Ch. 72" by Abbring and Heckman

James J. Heckman

THE UNIVERSITY OF CHICAGO

Econ 350, Winter 2023

- Define $P(X, Z) = \Pr(D = 1 \mid X, Z)$.
- If U_C is not perfectly predicted by (U_0, U_1), then we cannot, in general, estimate the joint distribution of (Y_0, Y_1, C) given (X, Z) or the distribution of (U_0, U_1, U_C) from data on Y , D, X and Z.

- However, under the conditions in Appendix B of Heckman and Vytlacil (2007a), we can identify up to an unknown scale for I, $F_{Y_0,I}(y_0, i \mid X, Z)$ and $F_{Y_1,I}(y_1, i \mid X, Z)$.
- The following intuition motivates the conditions under which $F_{Y_0,I}(y_0, i \mid X, Z)$ is identified.
- A parallel argument holds for $F_{Y_1,I}(y_1, i \mid X, Z)$.
- First, under the conditions given in Cosslett (1983), Manski (1988), Matzkin (1992) and Appendix B of Heckman and Vytlacil (2007a), we can identify $\frac{\mu_I(X,Z)}{\sigma_{U_I}}$ from $\Pr(D = 1 \mid X, Z) = \Pr(\mu_I(X, Z) + U_I \geq 0 \mid X, Z)$.
- $\sigma^2_{U_I}$ is the variance of U_I.
- We can also identify the distribution of $\frac{U_I}{\sigma_{U_I}}$.

- Second, from this information and $F_0(y_0 \mid D = 0, X, Z) = \Pr(Y_0 < y_0 \mid \mu_I(X, Z) + U_I < 0, X, Z)$, we can form

$$F_0(y_0 \mid D = 0, X, Z) \Pr(D = 0 \mid X, Z) = \Pr(Y_0 \leq y_0, I < 0 \mid X, Z).$$

- The left hand side of this expression is known (we observe Y_0 when $D = 0$ and we know the probability that $D = 0$ given X, Z).
- The right hand side can be written as

$$\Pr\left(Y_0 \leq y_0, \frac{U_I}{\sigma_{U_I}} < -\frac{\mu_I(X, Z)}{\sigma_{U_I}} \mid X, Z\right).$$

- In particular if $\mu_I(X, Z)$ can be made arbitrarily small ($\mu_I(X, Z) \to -\infty$), for a given X, we can recover the marginal distribution Y_0 from which we can recover $\mu_0(X)$, and hence the distribution of U_0.

- From the definition of Y_0, $U_0 = Y_0 - \mu_0(X)$.
- We may write the preceding probability as

$$\Pr\left(U_0 \leq y_0 - \mu_0(X), \frac{U_I}{\sigma_{U_I}} < \frac{-\mu_I(X,Z)}{\sigma_{U_I}} \mid X, Z\right).$$

- Note that the X and Z can be varied and y_0 is a number.
- Thus, by varying the known y_0 and $\frac{\mu_I(X,Z)}{\sigma_{U_I}}$, we can trace out the joint distribution of $\left(U_0, \frac{U_I}{\sigma_{U_I}}\right)$.
- Thus we can recover the joint distribution of

$$(Y_0, I) = \left(\mu_0(X) + U_0, \frac{\mu_I(X,Z) + U_I}{\sigma_{U_I}}\right).$$

- Notice the three key ingredients required to recover the joint distribution:
  (a) The independence between (U_0, U_I) and (X, Z).
  (b) The assumption that we can make $\frac{\mu_I(X,Z)}{\sigma_{U_I}}$ arbitrarily small for a given X (so we get the marginal distribution of Y_0 and hence $\mu_0(X)$). As noted in Heckman and Vytlacil (2007b), this type of identification-at-infinity assumption plays a key role in the entire selection and evaluation literature for identifying many important evaluation parameters, such as the average treatment effect and treatment on the treated.
  (c) The assumption that $\frac{\mu_I(X,Z)}{\sigma_{U_I}}$ can be varied independently of $\mu_0(X)$. This enables us to trace out the joint distribution of $\left(U_0, \frac{U_I}{\sigma_{U_I}}\right)$.

- A parallel argument establishes identification of the distribution of (Y_1, I) given X and Z.

- Identification of the Roy model follows from this analysis.
- Recall that the model assumes that $U_I = U_1 - U_0$ so $\sigma^2_{U_I} = \mathrm{Var}(U_1 - U_0)$.
- From the distributions of (Y_0, I) and (Y_1, I) given X and Z, we can recover the joint distributions of $(U_0, \frac{U_1 - U_0}{\sigma_{U_I}})$ and $(U_1, \frac{U_1 - U_0}{\sigma_{U_I}})$ and hence the joint distribution of (U_0, U_1).
- We can recover the joint distribution of $U_1 - U_0$ even if $\mu_I(X, Z) \neq \mu_1(X) - \mu_0(X)$ as long as $U_C \equiv 0$.

Using Additional Information

- We have established that data from social experiments or observational data corrected for selection do not in general identify joint distributions of potential outcomes.
- In the special case of the Roy model, choice data supplemented with outcome data will identify the joint distribution.
- But this result is fragile.
- For more general choice criteria, we cannot without further assumptions identify the joint distribution of potential outcomes.
- Recent approaches build on these results to supplement choice models with dependence assumptions to identify the joint distribution of (U_0, U_1).

- Aakvik et al. (2005), Carneiro et al. (2001, 2003), Cunha et al. (2005, 2006), and Cunha and Heckman (2007a, 2008b) use factor models to capture the dependence across the unobservables (U_0, U_1, U_I) and to supplement the information used in order to construct the joint distribution of counterfactuals.
- Their approach is a version of the proxy/replacement function approach developed in Heckman and Robb (1985, 1986) that is discussed in a section of Heckman and Vytlacil (2007b) and in Matzkin (2007).
- It extends factor models developed by Jöreskog and Goldberger (1975) and Jöreskog (1977) to restrict the dependence among the (U_0, U_1, U_I).

- A low dimensional set of random variables generates the dependence across the outcome unobservables.
- Such dimension reduction coupled with the use of choice data and additional measurements that proxy or replace the factors can provide enough information to identify the joint distributions of (Y_0, Y_1) and (Y_0, Y_1, D).

- The factor models are built around a conditional independence assumption.
- Conditional on the factors, outcomes and choice equations are independent.
- Thus the factor models have a close affinity with matching except that they do not assume that the analyst observes the factors and must instead integrate them out and identify their distribution.

- To demonstrate how this approach works, assume separability between observables and unobservables:

$$\begin{aligned} Y_1 &= \mu_1(X) + U_1 \\ Y_0 &= \mu_0(X) + U_0. \end{aligned}$$

- Denote I as the latent variable generating treatment choices:

$$\begin{aligned} I &= \mu_I(Z) + U_I \\ D &= \mathbf{1}\left[I \geq 0\right]. \end{aligned}$$

- Allow any X to be in Z so the notation is general.

- To understand this approach, it is convenient but not essential to assume that (U_0, U_1, U_I) is normally distributed with mean zero and covariance matrix Σ.
- Normality plays no essential role in the analysis of this section.
- The key role is played by the factor structure assumption introduced below.
- Assume access to data on (Y, D, X, Z).
- We can identify $F_0\left(y_0 \mid D = 0, X, Z\right)$, $F_1\left(y_1 \mid D = 1, X, Z\right)$ and $\Pr\left(D = 1 \mid X, Z\right)$.
- Under certain conditions presented in Appendix B, Heckman and Vytlacil (2007a) and the preceding section, we can identify the distributions of $(U_0, \frac{U_I}{\sigma_{U_I}})$ and $(U_1, \frac{U_I}{\sigma_{U_I}})$ nonparametrically.
- We can sometimes identify the scale on U_I.

- To restrict the dependence across the unobservables, we adopt a factor structure model for the U_0, U_1, U_I.
- Other restrictions across the unobservables are possible.
- Models for a single factor are extensively developed by Jöreskog and Goldberger (1975).
- Aakvik et al. (2005) and Carneiro et al. (2001, 2003) extend their analysis to generate distributions of counterfactuals.

- Initially assume a one-factor model where θ is a scalar factor (say unmeasured ability) that generates dependence across the unobservables assumed to be independent of (X, Z):

$$\begin{aligned} U_0 &= \alpha_0 \theta + \varepsilon_0, \\ U_1 &= \alpha_1 \theta + \varepsilon_1, \\ U_I &= \alpha_{U_I} \theta + \varepsilon_{U_I}, \end{aligned}$$

$$\theta \perp\!\!\!\perp \left(\varepsilon_0, \varepsilon_1, \varepsilon_{U_I}\right), \quad \left(\varepsilon_0, \varepsilon_1, \varepsilon_{U_I}\right) \text{ are mutually independent.}$$

- We discuss methods for multiple factors in the next section.
- Assume that $E\left(U_0\right) = 0$, $E\left(U_1\right) = 0$ and $E\left(U_I\right) = 0$.
- In addition, $E\left(\theta\right) = 0$.
- Thus $E\left(\varepsilon_0\right) = 0$, $E\left(\varepsilon_1\right) = 0$ and $E\left(\varepsilon_{U_I}\right) = 0$.

- To set the scale of the unobserved factor, normalize one "loading" (coefficient on θ) to 1.
- Note that all the dependence in the unobservables across equations arises from θ.

- From the joint distributions of $\left(U_0, \frac{U_I}{\sigma_{U_I}}\right)$ and $\left(U_1, \frac{U_I}{\sigma_{U_I}}\right)$ we can identify

$$\begin{aligned}\mathrm{Cov}\left(U_0, \frac{U_I}{\sigma_{U_I}}\right) &= \frac{\alpha_0 \alpha_{U_I}}{\sigma_{U_I}} \sigma_\theta^2 \\ \mathrm{Cov}\left(U_1, \frac{U_I}{\sigma_{U_I}}\right) &= \frac{\alpha_1 \alpha_{U_I}}{\sigma_{U_I}} \sigma_\theta^2\end{aligned}$$

  assuming that the covariances on the left hand side exist.
- From the ratio of the second covariance to the first we obtain $\frac{\alpha_1}{\alpha_0}$.
- Thus we obtain the sign of the dependence between U_0, U_1 because

$$\mathrm{Cov}(U_0, U_1) = \alpha_0 \alpha_1 \sigma_\theta^2.$$

- From the ratio, we obtain α_1 if we normalize $\alpha_0 = 1$.

- Without further information, we cannot identify the variance of U_I, $\sigma^2_{U_I}$.
- We normalize it to 1.
- Alternatively, we could normalize the variance of ε_{U_I} to 1.
- Below, we present a condition that sets the scale of U_I.

- With additional information, one can identify the full joint distribution of (U_0, U_1, U_I) and hence can construct the joint distribution of potential outcomes.
- In this section, we show this by a series of examples for a normal model.
- In a normal model, the joint distribution of (Y_0, Y_1) is determined (given X) if one can identify the variances of Y_0 and Y_1 and their covariance.
- We then show that normality plays no essential role in this analysis.
- We first consider what can be identified from access to a proxy M for θ (e.g., a test score).

Some Examples

Example 1

- **Access to a single proxy measure (e.g., a test score)**
  Assume access to data on Y_0 given $D = 0$, X, Z; to data on Y_1 given $D = 1$, X, Z; and data on D given X, Z.
- Suppose that the analyst also has access to a proxy for θ.
- Denote the proxy measure by M.

- In a schooling example, it could be a test score:

$$M = \mu_M(X) + U_M,$$

where

$$U_M = \alpha_M \theta + \varepsilon_M,$$

so

$$M = \mu_M(X) + \alpha_M \theta + \varepsilon_M,$$

where ε_M is independent of ε_0, ε_1, ε_{U_I} and θ, as well as (X, Z) $(\varepsilon_M \perp\!\!\!\perp (\varepsilon_0, \varepsilon_1, \varepsilon_{U_I}, \theta, X, Z))$.

- We can identify the mean $\mu_M(X)$ from observations on M and X.

- From this additional information, we acquire three additional covariance terms, conditional on X, Z, where we keep the conditioning implicit and define I as normalized by σ_{U_I}:

$$\begin{aligned} \text{Cov}(Y_1, M) &= \alpha_1 \alpha_M \sigma_\theta^2, \\ \text{Cov}(Y_0, M) &= \alpha_0 \alpha_M \sigma_\theta^2, \\ \text{Cov}(I, M) &= \frac{\alpha_{U_I}}{\sigma_{U_I}} \alpha_M \sigma_\theta^2. \end{aligned}$$

- Suppose that we normalize the loading on the proxy (or test score) to one ($\alpha_M = 1$).
- It is no longer necessary to normalize $\alpha_0 = 1$ as in the preceding section.

- From the ratio of the covariance of Y_1 with I with the covariance of I with M, we obtain the right-hand side of

$$\frac{\mathrm{Cov}\left(Y_1, I\right)}{\mathrm{Cov}\left(I, M\right)} = \frac{\alpha_1 \alpha_{U_I} \sigma_\theta^2}{\alpha_{U_I} \alpha_M \sigma_\theta^2} = \alpha_1,$$

because $\alpha_M = 1$ (normalization).

- From the discussion in the preceding section where no proxy is assumed, we obtain α_0 since

$$\frac{\mathrm{Cov}\left(Y_1, I\right)}{\mathrm{Cov}\left(Y_0, I\right)} = \frac{\alpha_1 \alpha_{U_I} \sigma_\theta^2}{\alpha_0 \alpha_{U_I} \sigma_\theta^2} = \frac{\alpha_1}{\alpha_0}.$$

- From knowledge of α_1 and α_0 and the normalization for α_M, we obtain σ_θ^2 from $\mathrm{Cov}\left(Y_1, M\right)$ or $\mathrm{Cov}\left(Y_0, M\right)$.

- We obtain α_{U_I} (up to scale σ_{U_I}) from $\mathrm{Cov}\left(I, M\right) = \frac{\alpha_{U_I} \alpha_M \sigma_\theta^2}{\sigma_{U_I}}$ since we know α_M $(= 1)$ and σ_θ^2.

- The model is overidentified.
- We can identify the scale of σ_{U_I} by a standard argument from the discrete choice literature.
- We review this argument below.
- Observe that if we write out the decision rule in terms of costs, we can characterize the latent variable determining choices as:

$$I = Y_1 - Y_0 - C,$$

where $C = \mu_C(Z) + U_C$ and $U_C = \alpha_C \theta + \varepsilon_C$, where ε_C is independent of θ and the other ε's.

- $E(U_C) = 0$ and U_C is independent of (X, Z).
- Then, $U_I = U_1 - U_0 - U_C$ and

$$\begin{aligned} \alpha_{U_I} &= \alpha_1 - \alpha_0 - \alpha_C, \\ \varepsilon_{U_I} &= \varepsilon_1 - \varepsilon_0 - \varepsilon_C, \\ \mathrm{Var}\left(\varepsilon_{U_I}\right) &= \mathrm{Var}(\varepsilon_1) + \mathrm{Var}(\varepsilon_0) + \mathrm{Var}(\varepsilon_C). \end{aligned}$$

- Identification of α_0, α_1 and α_{U_I} implies identification of α_C.
- Identification of the variance of ε_{U_I} implies identification of the variance of ε_C since the variances of ε_1 and ε_0 are known.
- Observe further that the scale σ_{U_I} is identified if there are variables in X but not in Z (see Heckman, 1976, 1979; Heckman and Robb, 1985, 1986; Willis and Rosen, 1979).

- From the variance of M given X, we obtain $\text{Var}(\varepsilon_M)$ since we know $\text{Var}(M)$ (conditional on X) and we know $\alpha_M^2\sigma_\theta^2$:

$$\text{Var}(M) - \alpha_M^2\sigma_\theta^2 = \sigma_{\varepsilon_M}^2.$$

- Recall that we keep the conditioning on X implicit.
- By similar reasoning, it is possible to identify $\text{Var}(\varepsilon_0)$, $\text{Var}(\varepsilon_1)$ and the fraction of $\text{Var}(U_I)$ due to ε_{U_I}.
- We can thus construct the joint distribution of (Y_0, Y_1, C) and hence the joint distribution of (Y_0, Y_1) since we identified $\mu_C(Z)$ and all of the factor loadings.
- Thus we can identify the objective outcome distribution for $(Y_{0,} Y_1)$ and the subjective distribution for C as well as their joint distribution (Y_0, Y_1, C).
- We have assumed normality because it is convenient to do so.

- Carneiro et al. (2003), Cunha et al. (2005, 2006) and Cunha and Heckman (2008b) show that it is possible to nonparametrically identify the distributions of θ, ε_0, ε_1, ε_{U_I} and ε_M so our results do not hinge on arbitrary distributional assumptions as we establish in the next section.
- We next show by way of example that choice data are not strictly required to secure identification of the joint distributions of counterfactuals.
- It is the extra information joined with the factor restriction on the dependence that allows us to identify the joint distribution of outcomes.

Example 2

- **Identification Without Choice Data** This example builds on Example 1.
- Let M be two dimensional so $M = (M_1, M_2)$, and M_1, M_2 are indicators that depend on θ and assume that they are both observed.
- In place of I from choice theory as in the preceding section, we can work with a second indicator of θ, i.e., a second measurement M_2.
- Suppose that either by limit operations ($P(X, Z) \to 0$ or $P(X, Z) \to 1$ along certain sequences in its support) or some randomization we observe triplets (Y_0, M_1, M_2), (Y_1, M_1, M_2) but not Y_0 and Y_1 together.

- We can still identify the joint distribution of (Y_0, Y_1).
- Example 1 applies to this case with only trivial modifications.
- We can identify all of the variances and covariances of the factor model as well as the factor loadings up to one normalization.
- Thus we can identify the joint distribution of (Y_0, Y_1).
- Since the (M_1, M_2) are assumed to be observed and their scale is known, we can identify the variances of M_1 and M_2 directly.

- In this example, we do not need to use any of the apparatus of discrete choice theory except to govern the limit operations that control for selection.
- There are other ways to construct the joint distributions that do not require a proxy M that may be extended to the model.
- Access to panel data on earnings affords identification.
- One way, that motivates our analysis of *ex ante* vs. *ex post* returns developed later, is given next.

Example 3

- **Two (or more) periods of panel data on outcomes**
  Suppose that for each person we have two periods of outcome data in one counterfactual state or the other.
- Thus we observe $(Y_{0,1}, Y_{0,2})$ or $(Y_{1,1}, Y_{1,2})$ but never both pairs of vectors together for the same person.
- We also observe choices.

- We assume that $Y_{j,t} = \mu_{j,t}(X) + U_{j,t}$, $j = 0, 1$, $t = 1, 2$, and write

$$U_{1,t} = \alpha_{1,t}\theta + \varepsilon_{1,t} \quad \text{and} \quad U_{0,t} = \alpha_{0,t}\theta + \varepsilon_{0,t}$$

to obtain

$$\begin{aligned} Y_{1,t} &= \mu_{1,t}(X) + \alpha_{1,t}\theta + \varepsilon_{1,t}, \qquad t = 1, 2, \\ Y_{0,t} &= \mu_{0,t}(X) + \alpha_{0,t}\theta + \varepsilon_{0,t}, \qquad t = 1, 2. \end{aligned}$$

- In the context of a schooling choice model as analyzed by Carneiro et al. (2001, 2003) and Cunha et al. (2005, 2006), if we assume that the interest rate is zero and that agents maximize the present value of their income, the index generating choices is

$$I = (Y_{1,2} + Y_{1,1}) - (Y_{0,2} + Y_{0,1}) - C$$

$$D = \mathbf{1}[I \geq 0],$$

where C was defined previously, and

$$\begin{aligned} I = {} & \mu_{1,1}(X) + \mu_{1,2}(X) - \mu_{0,1}(X) - \mu_{0,2}(X) - \mu_C(Z) \\ & + U_{1,1} + U_{1,2} - U_{0,1} - U_{0,2} - U_C. \end{aligned}$$

- We assume no proxy — just two periods of panel data.
- The multiple periods of earnings serve as the proxy.
- Under normality, application of the standard normal selection model allows us to identify $\mu_{1,t}(X)$ for $t = 1, 2$; $\mu_{0,t}(X)$ for $t = 1, 2$ and $\mu_{1,1}(X) + \mu_{1,2}(X) - \mu_{0,1}(X) - \mu_{0,2}(X) - \mu_C(Z)$, the latter up to a scalar σ_{U_I} where

$$U_I = U_{1,1} + U_{1,2} - U_{0,1} - U_{0,2} - U_C.$$

- Following our discussion of Example 1, we can recover the scale σ_{U_I} if there are variables in X that are not in Z such that $(\mu_{1,1}(X) + \mu_{1,2}(X) - (\mu_{0,1}(X) + \mu_{0,2}(X)))$ can be varied independently from $\mu_C(Z)$.

- To simplify the analysis, we assume that this condition holds.
- From normality, we can recover the joint distributions of $(I, Y_{1,1}, Y_{1,2})$ and $(I, Y_{0,1}, Y_{0,2})$ but not directly the joint distribution of $(I, Y_{1,1}, Y_{1,2}, Y_{0,1}, Y_{0,2})$.
- Thus, conditioning on X and Z, we can recover the joint distribution of $(U_I, U_{0,1}, U_{0,2})$ and $(U_I, U_{1,1}, U_{1,2})$ but apparently not that of $(U_I, U_{0,1}, U_{0,2}, U_{1,1}, U_{1,2})$.
- However, under our factor structure assumptions, this joint distribution can be recovered as we next show.

- From the available data, we can identify the following covariances:

$$
\begin{aligned}
\text{Cov}(U_I, U_{1,2}) &= (\alpha_{1,2} + \alpha_{1,1} - \alpha_{0,2} - \alpha_{0,1} - \alpha_C)\alpha_{1,2}\sigma_\theta^2 \\
\text{Cov}(U_I, U_{1,1}) &= (\alpha_{1,2} + \alpha_{1,1} - \alpha_{0,2} - \alpha_{0,1} - \alpha_C)\alpha_{1,1}\sigma_\theta^2 \\
\text{Cov}(U_I, U_{0,1}) &= (\alpha_{1,2} + \alpha_{1,1} - \alpha_{0,2} - \alpha_{0,1} - \alpha_C)\alpha_{0,1}\sigma_\theta^2 \\
\text{Cov}(U_I, U_{0,2}) &= (\alpha_{1,2} + \alpha_{1,1} - \alpha_{0,2} - \alpha_{0,1} - \alpha_C)\alpha_{0,2}\sigma_\theta^2 \\
\text{Cov}(U_{1,1}, U_{1,2}) &= \alpha_{1,1}\alpha_{1,2}\sigma_\theta^2 \\
\text{Cov}(U_{0,1}, U_{0,2}) &= \alpha_{0,1}\alpha_{0,2}\sigma_\theta^2.
\end{aligned}
$$

- If we normalize $\alpha_{0,1} = 1$ (recall that one normalization is needed to set the scale of θ), we can form the ratios

$$\frac{\mathrm{Cov}(U_I, U_{1,2})}{\mathrm{Cov}(U_I, U_{0,1})} = \alpha_{1,2}, \qquad \frac{\mathrm{Cov}(U_I, U_{1,1})}{\mathrm{Cov}(U_I, U_{0,1})} = \alpha_{1,1},$$

$$\frac{\mathrm{Cov}(U_I, U_{0,2})}{\mathrm{Cov}(U_I, U_{0,1})} = \alpha_{0,2}.$$

- From these coefficients and the remaining covariances, we identify σ_θ^2 using $\mathrm{Cov}(U_{1,1}, U_{1,2})$ and/or $\mathrm{Cov}(U_{0,1}, U_{0,2})$.

- Thus if the factor loadings are nonzero, we can identify σ_θ^2 from two relationships, both of which are identified:

$$\frac{\mathrm{Cov}(U_{1,1}, U_{1,2})}{\alpha_{1,1}\alpha_{1,2}} = \sigma_\theta^2$$

and

$$\frac{\mathrm{Cov}(U_{0,1}, U_{0,2})}{\alpha_{0,1}\alpha_{0,2}} = \sigma_\theta^2.$$

- Since we know $\alpha_{1,1}\alpha_{2,2}$ and $\alpha_{0,1}\alpha_{0,2}$, we can recover σ_θ^2 from $\mathrm{Cov}(U_{1,1}, U_{1,2})$ and $\mathrm{Cov}(U_{0,1}, U_{0,2})$.
- We can also recover α_C since we know σ_θ^2, $\alpha_{1,2} + \alpha_{1,1} - \alpha_{0,2} - \alpha_{0,1} - \alpha_C$, and $\alpha_{1,1}$, $\alpha_{1,2}$, $\alpha_{0,1}$, $\alpha_{0,2}$.

- We can form (conditional on X) $\text{Cov}(Y_{1,1}, Y_{0,1}) = \alpha_{1,1}\alpha_{0,1}\sigma_\theta^2$; $\text{Cov}(Y_{1,2}, Y_{0,1}) = \alpha_{1,2}\alpha_{0,1}\sigma_\theta^2$; $\text{Cov}(Y_{1,1}, Y_{0,2}) = \alpha_{1,1}\alpha_{0,2}\sigma_\theta^2$ and $\text{Cov}(Y_{1,2}, Y_{0,2}) = \alpha_{1,2}\alpha_{0,2}\sigma_\theta^2$.
- We can identify $\mu_C(Z)$ from the schooling choice equation since we know $\mu_{0,1}(X)$, $\mu_{0,2}(X)$, $\mu_{1,1}(X)$, $\mu_{1,2}(X)$ and we have assumed that there are some Z not in X so that σ_{U_I} is identified.
- Thus we can identify the joint distribution of $(Y_{0,1}, Y_{0,2}, Y_{1,1}, Y_{1,2}, C)$.

- These examples extend to nonnormal and nonparametric models.
- The key idea to constructing joint distributions of counterfactuals using the analysis of Cunha and Heckman (2008b) and Cunha et al. (2005, 2006) is *not* the factor structure for unobservables although it is convenient.
- The crucial idea is the assumption that a low dimensional set of random variables generates the dependence across outcomes.
- Other low dimensional representations such as the ARMA model or the dynamic factor structure model (see Sargent and Sims, 1977) can also be used.

- Cunha and Heckman (2008a) and Cunha et al. (2006) extend factor models to more general frameworks where the θ evolve over time as in state space models.
- The factor structure model presented in this section is easy to exposit and has been used to estimate joint distributions of counterfactuals.

- We present some examples in a later subsection.
- That subsection reviews recent work that generalizes the analysis of this section to derive *ex ante* and *ex post* outcome distributions, and measure the fundamental uncertainty facing agents in the labor market.
- With these methods it is possible to compute the distributions of both *ex ante* and *ex post* returns to treatments.
- Before presenting a more general analysis, we relate factor models to matching models.

Relationship to Matching

If the analyst knew θ and could condition on it, the analyst would obtain the conditional independence assumption of matching, (M-1):

U-1

$(Y_0, Y_1) \perp\!\!\!\perp D \mid X, Z, \theta.$

This is also the general control function assumption (U-1) in Heckman and Vytlacil (2007b).

- The approach developed by Aakvik et al. (2005), Carneiro et al. (2001, 2003), Cunha et al. (2005, 2006), and Cunha and Heckman (2007a,b, 2008b) extends matching and treats θ as an unobservable.
- It uses proxies for θ and identifies the distribution of θ under the following assumption:

U-2

$\theta \perp\!\!\!\perp X, Z.$

- Thus the factor approach is a version of matching on unobservables, where the unobserved match variables are integrated out.

Nonparametric Extensions

- The analysis of the generalized Roy model developed in Appendix B of Heckman and Vytlacil (2007a) establishes conditions under which it is possible to nonparametrically identify the joint distribution of (Y_0, I, M) given X, Z and the joint distribution of (Y_1, I, M) given X, Z, where we also allow the functions determining M to be nonparametrically determined.

- These conditions can be extended to provide identification of the distributions of (Y_0, I, M) and (Y_1, I, M) where M is observed for all persons treated or not whereas Y_0 and Y_1 are observed only if $D = 0$ or $D = 1$, respectively.
- The identification conditions are also easily extended to account for vector Y_0 and Y_1 (e.g., $Y_0 = (Y_{0,1}, Y_{0,2})$ and $Y_1 = (Y_{1,1}, Y_{1,2})$) as our third example.

- We present a general theorem for the identification of state-contingent outcomes free of selection bias in the next section and in the appendix of this presentation.
- With the state-contingent distributions nonparametrically identified, we can apply factor analysis to identify the factor loadings because we identify the required covariances as a by-product of our nonparametric analysis.

- With the α_j (or $\alpha_{i,j}$) in hand, we can nonparametrically identify the distribution of θ and the ε_j (or $\varepsilon_{i,j}$) for the different models assuming mutual independence between θ and all of the components of ε_j (or $\varepsilon_{i,j}$) using Kotlarski's Theorem (Kotlarski, 1967; Prakasa Rao, 1992).
- That theorem states that, for any pair of random variables T_1, T_2 generated by a common random variable θ, we can nonparametrically identify the distribution of θ and the associated components of errors: ε_1 and ε_2.

Stated precisely:

Theorem 1

If

$$T_1 = \theta + \varepsilon_1$$

and

$$T_2 = \theta + \varepsilon_2$$

and $(\theta, \varepsilon_1, \varepsilon_2)$ are mutually independent, the means of all three generating random variables are finite and are normalized to $E(\varepsilon_1) = E(\varepsilon_2) = 0$, and the random variables possess nonvanishing ($a.e.$) characteristic functions, then the densities of $(\theta, \varepsilon_1, \varepsilon_2)$, $g_\theta(\theta), g_1(\varepsilon_1), g_2(\varepsilon_2)$, respectively, are identified.

Proof

Kotlarski (1967). See also Prakasa Rao (1992).