Identification of Multivariate Measurement Error Models

Yingyao Hu

The 6th JHU Workshop on
Measurement Errors & Latent Variables
May 9, 2025

Multi-dimensional measurements

- Measurements, $i = 1, 2, 3$

$$X^i = \begin{pmatrix} X_1^i \\ \dots \\ X_{K_i}^i \end{pmatrix} \in \mathbb{R}^{K_i}$$

- Each X^i measures a latent structure described by a vector of random variables

$$X^* = \begin{pmatrix} X_1^* \\ \dots \\ X_L^* \end{pmatrix} \in \mathbb{R}^L$$

- I define X^i as an **incomplete** measurement of X^* if

$$K_i < L.$$

- When incomplete, injectivity of $L_{X^i|X^*}$ doesn't hold under continuity.

A general framework

- $X = \begin{pmatrix} X^1 \\ X^2 \\ X^3 \end{pmatrix}$
- $K = K^1 + K^2 + K^3$
- $$L_{X|X^*} : \mathcal{L}^2\left(\mathcal{R}^L\right) \to \mathcal{L}^2\left(\mathcal{R}^K\right)$$

$$\begin{aligned} p_X &= \int p_{X|X^*} p_{X^*} dX^* \\ &\equiv L_{X|X^*} p_{X^*} \end{aligned} \tag{1}$$

- $$\begin{aligned} X^1 &= M^1 X^* + \varepsilon^1 \\ X^2 &= M^2 X^* + \varepsilon^2 \\ X^3 &= M^3 X^* + \varepsilon^3 \end{aligned} \tag{2}$$

- Each of three error vectors is conditional mean independent of other latent vectors, i.e., for $i \neq j \neq k$

$$E[\varepsilon^i | \varepsilon^j, \varepsilon^k, X^*] = 0. \tag{3}$$

The elements in each vector ε^i are correlated arbitrarily.

- The latent vector X^* satisfy:

$$\begin{aligned} E[X^*] &= 0 \\ E[X^*_{-i}(X^*_{-i})^T | X^*_i] &= E[X^*_{-i}(X^*_{-i})^T] \\ E[(X^*_i)^3] &\neq 0 \end{aligned} \tag{4}$$

where X^*_{-i} denotes the vector X^* with its i-th component removed.

Identification of factor loadings

- Kruskal Rank: $\kappa(M^i)$ defined as

$$\kappa(M^i) = \max\left\{k : \text{every set of } k \text{ columns of } M^i \text{ is linearly independent}\right\}$$

- A key observation is that 3rd order moments satisfy

$$E(X_i^1 \times X_u^2 \times X_v^3) = \sum_{l=1}^{l=L} m_{i,l}^1 \times m_{u,l}^2 \times m_{v,l}^3 \times E(X_l^*)^3$$

- We then assume

$$\kappa(M^1) + \kappa(M^2) + \kappa(M^3) \geq 2L + 2.$$

- Kruskal's Theorem implies that M^1, M^2, M^3 and $E(X_l^*)^3$ are unique (up to permutation and scaling).

- Key assumption:
  Probability function $p(X^*, \varepsilon^1, \varepsilon^2, \varepsilon^3)$ is continuous and satisfies
  1. $p(\varepsilon^1, \varepsilon^2, \varepsilon^3 | X^*) = p(\varepsilon^1) \times p(\varepsilon^2) \times p(\varepsilon^3)$.
  2. $p(X^*) = p(X_1^*) \times ... \times p(X_L^*)$.
  3. $\phi_{X^i} \neq 0$ on $\mathbb{R}^{K_i}$ for $i = 1, 2, 3$, where ϕ_{X^i} is the characteristic function of X^i.
- Key step: To identify the distribution of each X_l^*.
- With ch.f. $\phi_{X_l^*}$, we identify the whole distribution.

$$\phi_{X^*} = \phi_{X_1^*} \times ... \times \phi_{X_L^*}$$

$$\phi_{\varepsilon^i} = \frac{\phi_{X^i}}{\phi_{M^i X^*}}$$

That means $p(X, X^*)$ is identified.

- Consider To identify the distribution of X_1^*, we find two measurements of X_l^* satisfying the conditions to use the Kotlarski's identity.
- Let $\kappa^i = \kappa(M^i)$. That means one can generate a $\kappa^1 \times \kappa^1$ identity matrix from M^i by applying the Gram-Schmidt algorithm to the rows. Therefore, there exist Q^1 s.t.

$$\begin{aligned} W^1 \equiv Q^1 X^1 &= Q^1 M^1 X^* + Q^1 \varepsilon^1 \qquad (5) \\ &= \begin{pmatrix} I_{\kappa^1\times\kappa^1} & R^1_{\kappa^1\times(L-\kappa^1)} \\ 0 & R^1_{(K^1-\kappa^1)\times(L-\kappa^1)} \end{pmatrix} X^* + Q^1\varepsilon^1 \\ &\equiv \begin{pmatrix} 1 & 0 & R^1_{1\times(L-\kappa^1)} \\ 0 & I_{(\kappa^1-1)\times(\kappa^1-1)} & R^1_{(\kappa^1-1)\times(L-\kappa^1)} \\ 0 & 0 & R^1_{(K^1-\kappa^1)\times(L-\kappa^1)} \end{pmatrix} X^* + Q^1\varepsilon^1 \end{aligned}$$

- we consider the first element in X^*

$$\begin{aligned} W_1^1 &= X_1^* + R^1_{1\times(L-\kappa^1)}(X^*_{\kappa^1+1}, ..., X_L^*)^T + [Q^1]_1\varepsilon^1 \\ &\equiv X_1^* + e_1 \qquad (6) \end{aligned}$$

Key step: Revealing latent distributions

- Similarly, there exist Q^2 s.t.

$$
\begin{aligned}
W^2 \equiv Q^2 X^2 &= Q^2 M^2 X^* + Q^2 \varepsilon^2 \qquad (7) \\
&= \begin{pmatrix} R^2_{(K^2-\kappa^2)\times(L-\kappa^2)} & 0 \\ R^2_{(\kappa^2)\times(L-\kappa^2)} & I_{\kappa^2\times\kappa^2} \end{pmatrix} X^* + Q^2\varepsilon^2 \\
&\equiv \begin{pmatrix} R^2_{(K^2-\kappa^2)\times(L-\kappa^2)} & 0 & 0 \\ R^2_{1\times(L-\kappa^2)} & 1 & 0 \\ R^2_{(\kappa^2-1)\times(L-\kappa^2)} & 0 & I_{(\kappa^2-1)\times(\kappa^2-1)} \end{pmatrix} X^* + Q^2\varepsilon^2 \\
&\equiv \begin{pmatrix} R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)} & 0 \\ R^2_{(\kappa^2-1)\times(L-\kappa^2+1)} & I_{(\kappa^2-1)\times(\kappa^2-1)} \end{pmatrix} X^* + Q^2\varepsilon^2
\end{aligned}
$$

- We then pick an element in W^2 which does not contain $(X^*_{\kappa^1+1}, \ldots, X^*_L)$, which are in W^1_1, but does contain X^*_1.

- There must be a nonzero entry $[R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)}]_{i,1} \neq 0$ in the first column (corresponding to X_1^*) in $R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)}$. Otherwise, the Kruskal rank has to be smaller than κ^2. We then consider

$$
\begin{aligned}
W_i^2 &= [R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)}]_{i,1} X_1^* + \\
&\quad + \sum_{j=2}^{L-\kappa^2+1} [R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)}]_{i,j} X_j^* + [Q^2]_i \varepsilon^2 \qquad (8) \\
&\equiv r X_1^* + e_2 \qquad (9)
\end{aligned}
$$

Key step: Revealing latent distributions

- In summary, we have for a known r

$$\begin{aligned} W_1^1 &\equiv X_1^* + e_1 \\ W_i^2 &\equiv rX_1^* + e_2 \end{aligned}$$

- $$\phi_{X_1^*}(t) = \exp\left(\int_0^t \frac{\left[\frac{\partial}{\partial t_2} \phi_{W_1^1, W_i^2}\left(s, \frac{t_2}{\gamma}\right) \right]_{t_2=0}}{\phi_{W_1^1}(s)} ds \right)$$

- $$\begin{aligned} \phi_{X^*} &= \phi_{X_1^*} \times \dots \times \phi_{X_L^*} \\ \phi_{\varepsilon^i} &= \frac{\phi_{X^i}}{\phi_{M^i X^*}} \end{aligned}$$

- $$\begin{aligned} & p(X^1, X^2, X^3, X^*) \\ = \; & p_{\varepsilon^1}(X^1 - M^1 X^*) p_{\varepsilon^2}(X^2 - M^2 X^*) p_{\varepsilon^3}(X^3 - M^3 X^*) p(X_1^*) \dots p(X_L^*) \end{aligned}$$

Identification of distributions without injectivity

Theorem

Suppose that Assumptions above hold. Then, factor loading matrices M^i for $i = 1, 2, 3$ are unique up to permutation and scaling and $p(X^i|X^*)$ for $i = 1, 2, 3$ and $p(X^*)$ are unique in

$$p(X^1, X^2, X^3) = \int p(X^1|X^*)p(X^2|X^*)p(X^3|X^*)p(X^*)dX^* \tag{10}$$

if

$$\kappa(M^1) + \kappa(M^2) + \kappa(M^3) \geq 2L + 2.$$

- Notice that $L_{X|X^*}$ and $L_{X^i|X^*}$ may be non-injective.

Implications

- Extending the Kruskal's Theorem to the continuous case using a linear structure
- All the measurements can be incomplete. e.g. $K_i = 4$ and $L = 5$, under the full rank condition $4 + 4 + 4 = 2 \times 5 + 2$.
  - Three four-dimensional measurements can reveal five latent dimensions.
  - No injectivity is needed directly.
- Constructive identification with closed-form solution
- f_{X^1,X^2,X^3} uniquely determines f_{X^1,X^2,X^3,X^*}
- Some assumptions can be relaxed. e.g. independence of ε^3 can be relaxed to
$$E[\varepsilon^3|X^*, \varepsilon^1, \varepsilon^2] = 0$$
- Elements inside measurement error ε^i can be correlated.

Extension: Identification with correlated factors

- Assumption: Probability function $p(X^*, \varepsilon^1, \varepsilon^2, \varepsilon^3)$ is continuous and satisfies
  1. $p(\varepsilon^1, \varepsilon^2, \varepsilon^3 | X^*) = p(\varepsilon^1) \times p(\varepsilon^2) \times p(\varepsilon^3)$.
  2. $\phi_{X^1} \neq 0$ on $\mathbb{R}^{K_1}$

Theorem

Suppose that Assumptions above hold. Then, factor loading matrices M^i for $i = 1, 2, 3$ are unique up to permutation and scaling and $p(X^i|X^*)$ for $i = 1, 2, 3$ and $p(X^*)$ are unique in

$$p(X^1, X^2, X^3) = \int p(X^1|X^*)p(X^2|X^*)p(X^3|X^*)p(X^*)dX^* \quad (11)$$

if both M^1 and $\begin{pmatrix} M^2 \\ M^3 \end{pmatrix}$ have a full column rank L and

$$\kappa(M^2) + \kappa(M^3) \geq L + 2. \quad (12)$$

- For vectors of random variables $W \in \mathbb{R}^K$ and $X^* \in \mathbb{R}^L$, we define

$$\mathrm{J}(p_{W|X^*}) = \left[\mathrm{I} \left(\frac{\partial \, p_{W_i|X^*}(\cdot|x^*)}{\partial x_j^*} \neq 0 \right) \right]_{i=1,\ldots,K;j=1,\ldots,L} \tag{13}$$

  where $\mathrm{J}(p_{W|X^*})$ is a matrix of indicators of whether X_j^* enters the distribution $p_{W_i|X^*}$ for each W_i in vector W.
- An indictor matrix or sparsity matrix

- We introduce a generalized Kruskal rank for integral operators, called signal rank, as follows:

Definition

For a measurement $X \in \mathbb{R}^K$ of a latent variable $X^* \in \mathbb{R}^L$, the **signal rank** of integral operator $L_{X|X^*}$ is defined as the largest integer $\kappa^s = \kappa^s(L_{X|X^*})$ such that: For any permutation of X_l^* for $l \in \{1, 2, ..., L\}$ in X^*, there exists a continuous one-to-one mapping $Q : \mathbb{R}^K \to \mathbb{R}^K$, which satisfies, for $W = Q(X)$

1. $\mathrm{J}(p_{W|X^*}) = \begin{pmatrix} I_{\kappa^s \times \kappa^s} & R_{\kappa^s \times (L-\kappa^s)} \\ R_{(K-\kappa^s)\times\kappa^s} & R_{(K-\kappa^s)\times(L-\kappa^s)} \end{pmatrix}$;
2. $L_{W_{1:k}|X^*_{1:k}}$ is invertible for all $1 \leq k \leq \kappa^s$.

- Note $W_{1:k}$ is the sub-vector of the first k entries in vector W.

Nonlinear cases: A generalization of Kruskal rank

- Assumption:
  Probability function $p(X^1, X^2, X^3, X^*)$ is continuous and satisfies
  1. $p(X^1, X^2, X^3 | X^*) = p(X^1 | X^*) \times p(X^2 | X^*) \times p(X^3 | X^*)$.
  2. $p(X^*) = p(X_1^*) \times ... \times p(X_L^*)$.
- This assumption leads to

$$
\begin{aligned}
& p(X^1, X^2, X^3) \\
= & \int p(X^1 | X^*) p(X^2 | X^*) p(X^3 | X^*) p(X_1^*) ... p(X_L^*) dX^*
\end{aligned}
$$

Nonlinear cases: Nonparametric identification

- Given the original permutation of vector $X^* = (X_1^*, ..., X_L^*)^T$, we apply its corresponding mapping to obtain $W^1 = Q_1(X^1)$.
- For X^2, we consider a permutation of X^* as follows:

$$X^* = (X_1^*; X_{L-\kappa_2^s+2}^*, X_{L-\kappa_2^s+3}^*, ..., X_L^*; X_2^*, X_3^*, ..., X_{L-\kappa_2^s+1}^*)^T$$

with $W^2 = Q_2(X^2)$.

- Consider the first elements

$$\begin{aligned}
& p(W_1^1, W_1^2) \\
= & \int p(W_1^1|X^*)p(W_1^2|X^*)p(X_1^*)...p(X_L^*)dX^* \\
= & \int p(W_1^1|X_1^*, ..., X_L^*) \times \\
& \times p(W^2|X_1^*; X_{L-\kappa_2^s+2}^*, X_{L-\kappa_2^s+3}^*, ..., X_L^*; X_2^*, X_3^*, ..., X_{L-\kappa_2^s+1}^*) \\
& \times p(X_1^*) \times ... \times p(X_L^*)dX^*
\end{aligned}$$

Nonlinear cases: Nonparametric identification

- By definition of signal rank

$$\mathrm{J}(p_{W^1|X^*}) = \begin{pmatrix} I_{\kappa_1^s \times \kappa_1^s} & R_{\kappa_1^s \times (L-\kappa_1^s)} \\ R_{(K_1-\kappa_1^s)\times \kappa_1^s} & R_{(K_1-\kappa_1^s)\times(L-\kappa_1^s)} \end{pmatrix} \quad (14)$$

and

$$\mathrm{J}(p_{W^2|X^*}) = \begin{pmatrix} I_{\kappa_2^s \times \kappa_2^s} & R_{\kappa_2^s \times (L-\kappa_2^s)} \\ R_{(K_2-\kappa_2^s)\times \kappa_2^s} & R_{(K_2-\kappa_2^s)\times(L-\kappa_2^s)} \end{pmatrix} \quad (15)$$

- We have

$$\begin{aligned} & p(W_1^1, W_1^2) \quad (16) \\ = & \int p(W_1^1|X^*)p(W_1^2|X^*)p(X_1^*)...p(X_L^*)dX^* \\ = & \int p(W_1^1|X_1^*, X_{\kappa_1^s+1}^*, ..., X_L^*)p(W^2|X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*) \\ & \times p(X_1^*)...p(X_L^*)dX^* \end{aligned}$$

- Note that $\kappa_3^s \leq L$. Then, we have

$$\kappa_1^s + \kappa_2^s \geq L + 2 \tag{17}$$

and

$$\kappa_1^s + 1 > L - \kappa_2^s + 1 \tag{18}$$

which implies that X_1^* is the only common element between W_1^1, and W_1^2.

- We have

$$\begin{aligned} & p(W_1^1, W_1^2) \\ = & \int p(W_1^1 | X_1^*, X_{\kappa_1^s+1}^*, ..., X_L^*) p(W^2 | X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*) \\ & \times p(X_1^*)...p(X_L^*) dX^* \\ = & \int p(W_1^1 | X_1^*) p(W_1^2 | X_1^*) p(X_1^*) dX_1^* \end{aligned} \tag{19}$$

Nonlinear cases: Nonparametric identification

- In operator form, we have

$$L_{W_1^1,W_1^2} = L_{W_1^1|X_1^*} D_{X_1^*} (L_{W_1^2|X_1^*})^T \tag{20}$$

- Assumption: $L_{X^3|X^*}$ has a full signal rank with $W^3 = Q_3(X^3)$
- we can show

$$\begin{aligned} L_{W_1^1,W_1^3} &= L_{W_1^1|X_1^*} D_{X_1^*} (L_{W_1^3|X_1^*})^T \\ L_{W_1^3,W_1^2} &= L_{W_1^3|X_1^*} D_{X_1^*} (L_{W_1^2|X_1^*})^T \end{aligned}$$

- Given that these operators are invertable, we have

$$\begin{aligned} & L_{W_1^3,W_1^2} (L_{W_1^1,W_1^2})^{-1} L_{W_1^1,W_1^3} \\ = \; & L_{W_1^3|X_1^*} D_{X_1^*} (L_{W_1^3|X_1^*})^T \end{aligned} \tag{21}$$

- Assumption (normalization): $L_{W_1^3|X_1^*} D_{X_1^*} (L_{W_1^3|X_1^*})^T$ uniquely determines $L_{W_1^3|X_1^*}$ and $D_{X_1^*}$.

Discussion on additional assumptions

- Assumption: $L_{X^3|X^*}$ has a full signal rank.
  1. This assumption may not be required under some specifications similar to the linear case, e.g.,

$$X^i = M^i \begin{pmatrix} g_1^i(X_1^*) \\ \dots \\ g_L^i(X_L^*) \end{pmatrix} + \varepsilon^i \tag{22}$$

  2. My conjecture: this full rank condition can be avoided under mild conditions, just as in the linear case. But the proof may be as complicated as the main proof of Kruskal' theorem.
- Assumption (normalization): $L_{W_1^3|X_1^*} D_{X_1^*} (L_{W_1^3|X_1^*})^T$ uniquely determines $L_{W_1^3|X_1^*}$ and $D_{X_1^*}$.
  1. This is a high level normalization assumption.
  2. A special case could be

$$\begin{aligned} W_1^3 &= X_1^* + e^1 \\ \tilde{W}_1^3 &= X_1^* + \tilde{e}^1 \end{aligned} \tag{23}$$

  3. In matrix form, it is to assume HH^T uniquely determines H. A common normalization in factor analysis.

- Identification of $L_{W_1^3|X_1^*}$ implies that we can identify $L_{W_l^3|X_l^*}$ and $D_{X_l^*}$, i.e., $p(W_l^3, X_l^*)$, for all l are identified.
- Finally, we show the whole distribution $p(X^1, X^2, X^3, X^*)$ is identified. We consider

$$\begin{aligned} p(X^1, X^2, W^3) &= \int p(W^3|X^*)p(X^1, X^2, X^*)dX^* \\ &= \int p(W_1^3|X_1^*)...p(W_L^3|X_L^*)p(X^1, X^2, X^*)dX^* \end{aligned}$$

- Given the mapping Q^3 is one-to-one and continuous, we also identify the distribution of $p(X^3|X^*)$ from $p(W^3|X^*)$. That means we have identified

$$p(X^1, X^2, X^3, X^*) = p(X^3|X^*)p(X^1, X^2, X^*)$$

- Finally, we have

Theorem

Suppose that Assumptions above hold. Then, $p(X^i|X^*)$ for $i = 1, 2, 3$ and $p(X^*)$ are unique, up to the permutation of X_l^* for $l = 1, 2, ..., L$, in

$$p(X^1, X^2, X^3) = \int p(X^1|X^*)p(X^2|X^*)p(X^3|X^*)p(X^*)dX^* \tag{24}$$

if

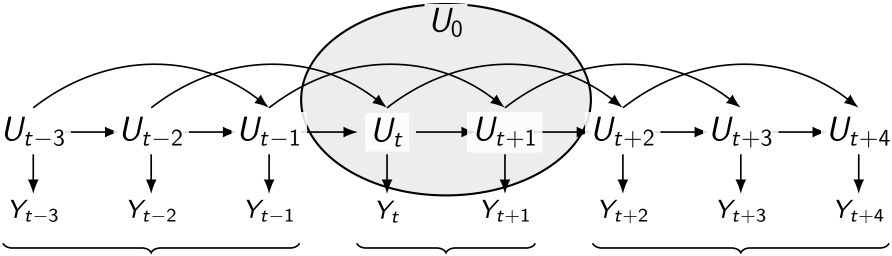
$$\kappa^s(L_{X^1|X^*}) + \kappa^s(L_{X^2|X^*}) + \kappa^s(L_{X^3|X^*}) \geq 2L + 2. \tag{25}$$

where $\kappa^s(L_{X^i|X^*})$ is the signal rank of $L_{X^i|X^*}$.

- Mapping Q^i needs to be found out in applications, just as factor loadings in the linear case.

Application: Earning dynamics with permanent distributional effects

$$Y_t = g_t\left(U_t, U_0, \varepsilon_t\right),$$
$$U_t = h_t\left(U_{t-1}, U_{t-2}, U_0, e_t\right).$$



Conclusion

- Discovering latent information with incomplete measurements.
- Extending the Kruskal's Theorem to the multivariate continuous case
- f_{X^1,X^2,X^3} uniquely determines f_{X^1,X^2,X^3,X^*}. Widely applicable.
- The linear case is very intuitive with nice features
  1. All the measurements can be incomplete. e.g. three four-dimensional measurements can reveal five latent dimensions.
     - No injectivity is needed.
  2. Constructive identification with closed-form solution
- Nonlinear cases can be very general. My conjecture is that identification is possible to without injectivity under certain restrictions on the nonlinear structure.
- General implication: Machine can dig out latent information using this procedure without human. Maybe that is how machine gains intelligence!?