cemmap
centre for microdata methods and practice

Identification of multivariate measurement error models

Yingyao Hu

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP19/25

UKRI Economic and Social Research Council

Identification of Multivariate Measurement Error Models

Yingyao Hu*

December 3, 2025

Abstract

This paper develops new identification results for multidimensional continuous measurement-error models where all observed measurements are contaminated by potentially correlated errors and none provides an injective mapping of the latent distribution. Using third-order cross-moments, the paper constructs a three-way tensor whose unique decomposition, guaranteed by Kruskal's theorem, identifies the factor loading matrices. Starting with a linear structure, the paper recovers the full distribution of latent factors by constructing suitable measurements and applying scalar or multivariate versions of Kotlarski's identity. As a result, the joint distribution of the latent vector and measurement errors is fully identified without requiring injective measurements, showing that multivariate latent structure can be recovered in broader settings than previously believed. Under injectivity, the paper also provides user-friendly testable conditions for identification. Finally, this paper provides general identification results for nonlinear models using a newly-defined generalized Kruskal rank - signal rank - of intergral operators. These results have wide applicability in empirical work involving noisy or indirect measurements, including factor models, survey data with reporting errors, mismeasured regressors in econometrics, and multidimensional latent-trait models in psychology and marketing, potentially enabling more robust estimation and interpretation when clean measurements are unavailable.

Keywords: *identification, measurement error, latent variable, factor model, multivariate, Kruskal rank, signal rank.*

*Contact information: Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. Email: yhu@jhu.edu.

1 Introduction

Multivariate continuous latent variables arise in numerous empirical settings in economics, psychology, marketing, epidemiology, and the social sciences. Examples include multidimensional skills, cognitive factors, latent preferences, health indices, productivity components, and risk attitudes. In practice, such latent constructs are rarely observed directly; instead, researchers rely on multiple imperfect measurements that contain potentially correlated forms of noise. The resulting measurement-error problem is especially severe when the latent variable is multidimensional: each measurement typically captures only a low-dimensional projection of the latent vector, and the noise may exhibit dependence or heterogeneity across measurement channels. As a result, classical approaches to continuous measurement error, which often rely on injectivity, offer limited guidance.

This paper develops an identification strategy tailored specifically to multivariate continuous latent variables measured with noise. The key innovation is to combine tools from multilinear algebra, specifically the uniqueness properties of so-called CP tensor decompositions, with the multivariate extension of Kotlarski's identity, a powerful deconvolution result based on characteristic functions. The starting point is the observation that third-order cross-moments of three separate measurements form a three-way moment tensor whose CP decomposition is governed by the latent factor loading matrices. By invoking Kruskal's theorem (Kruskal, 1977a), I show that these loadings are generically identifiable even when each measurement matrix is rank-deficient and—even more surprisingly—when the stack of all measurement matrices is non-injective. This greatly relaxes the standard requirements for identifying multivariate continuous latent variables, which typically demand injectivity of the measurement operator.

Once the factor loading matrices are recovered through tensor methods, the paper turns to the problem of identifying the entire joint distribution of the continuous latent vector. A crucial step of the identification is that the Kruskal rank condition leads to two conditionally independent noisy measurements of each coordinate of the latent vector, derived from suitable linear projections of the observed measurements. These constructed pairs satisfy the assumptions of Kotlarski's identity, enabling closed-form identification of each coordinate's marginal distribution. For the case of dependent continuous latent factors, a weaker version of the full rank condition allows us to produce two vector-valued measurements whose errors are independent of the latent vector, thereby permitting the use of the multivariate Kotlarski identity. This yields identification of the full joint distribution of the continuous latent vector without assuming independence across its components.

A major contribution of this paper lies in demonstrating that multivariate continuous latent variables can be identified nonparametrically from three non-injective measurements by combining tensor uniqueness with characteristic-function deconvolution. It can be considered as a small step towards the extension of the Kruskal's theorem to continuous settings.

As Allman et al. (2009) showed, the Kruskal's theorem can identify distribution of discrete latent variables. Extending that to continuous variables would require the extension of the Kruskal's theorem to continuous setting, which, to the best of my knowledge, is still an open question. The theorem in Hu and Schennach (2008) can be considered as such an extension to the continuous case under injectivity because a full rank condition in the discrete case naturally extends to the injectivity condition in the continuous case. Furthermore, another important contribution of this paper is to provide a generalized Kruskal rank - signal rank - for integral operators, and show that identification can hold for general nonlinear models under a signal rank condition similar to the well-known Kruskal rank condition. The newly-defined rank is consistent with the existing Kruskal rank for matrices and the injectivity of integral operators, and can be used to describe how informative a measurement is in the continuous multivariate case. Based on the signal rank condition, I provide identification of nonlinear models under high-level assumptions. In summary, this paper provides a constructive identification solution to a linear multivariate measurement error model, and builds a foundation for a better solution to the identification of general nonlinear measurement error models.

This work is closely related to several strands of literature. Nonparametric identification of measurement-error models with continuous latent variables has been advanced by e.g., Kotlarski (1967), and Hu and Schennach (2008). Tensor decomposition methods have been widely used to handle discrete latent variables in psychometrics, e.g., Carroll and Chang (1970); Harshman (1970), signal processing, e.g., Sidiropoulos et al. (2000), and in econometrics and statistics, e.g., Hu (2008), Allman et al. (2009) s,Bonhomme et al. (2016). This paper contributes a novel bridge between these literatures by demonstrating how tensor uniqueness can be used to identify both the measurement system and the full distribution of a multivariate continuous latent variable without injectivity or with less injectivity restrictions.

The limitation of this paper lies in the high-level assumptions in the nonlinear cases. The linear specification of the multivariate measurement structure may be rich enough for most empirical applications, but not complicated enough for theoretical explorations. Relaxaion of the linearity and additivity requires new technical tools. For that purpose, this paper provides a promising starting point towards continuous or operator-valued generalizations of Kruskal's theorem. Such continuous tensor identifiability results would naturally complement the multivariate Kotlarski identity and could relax the need for discretized third-order moments. With the results and the new concept in this paper, we are a small step closer toward a unified theory of identification for high-dimensional continuous latent-variable models in the presence of complex, nonclassical measurement error.

This paper is organized as follows: Section 2 specifies the linear model; Section 3 shows the identification of the factor loadings using the Kruskal's theorem; Section 4 presents the identification of continuous distributions without injectivity, but also provides a user friendly

identification results under injectivity; Section 5 uses a newly-defned signal rank to present the identification of nonlinear cases. Section 6 concludes. All the proofs are in the Appendix.

2 A 3-measurement model and injectivity

We are interested in a latent structure described by a vector of latent variables

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

Instead of X^*, we observe in a sample of three multi-dimensional measurements, for $i = 1, 2, 3$

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

The relationship between the latent vector X^* and its measurements can be described by the integral operator $L_{X^i|X^*} : \mathcal{L}^2\left(\mathcal{R}^L\right) \to \mathcal{L}^2\left(\mathcal{R}^{K^i}\right)$

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

where p_{X^*} stands for the probability density function of X^*. It is known that nonparametric identification of the distribution of the latent variable X^* requires that the integral operator $L_{X^i|X^*}$ is injective. This paper presents the identification of the joint distribution of X^* and $X = \left((X^1)^T, (X^2)^T, (X^3)^T\right)^T$ without such injectivity or with fewer injectivity restrictions than in the existing literature.

We start with a linear specification of the relationship between the latent vector X^* and the three measurements as follows:

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

where M^i are factor loading matrices. In the case where the dimension of X^i is smaller than that of X^*, integral operator $L_{X^i|X^*}$ can't be injective if $p(X^i|X^*)$ is continuous in all its arguments. In particular, when the factor loading matrix M^i doesn't have a full column rank, the injectivity of $L_{X^i|X^*}$ fails. Furthermore, the injectivity may still fail even if we combine

the three measurements. For example, we consider $L_{X|X^*} : \mathcal{L}^2\left(\mathcal{R}^L\right) \to \mathcal{L}^2\left(\mathcal{R}^{K^1+K^2+K^3}\right)$

$$\begin{aligned} p_X &= \int p_{X|X^*} p_{X^*} dX^* \\ &\equiv L_{X|X^*} p_{X^*} \end{aligned} \tag{3}$$

A necessary condition for the injectivity of $L_{X|X^*}$ is that matrix M has a full column rank, where

$$M = \begin{pmatrix} M^1 \\ M^2 \\ M^3 \end{pmatrix}.$$

However, it is possible that $M^1 = M^2 = M^3$ with $\text{Rank}[M^i] < L$ so that $\text{Rank}[M] < L$. That means the rank of the stacked matrix is the same as that of M^i and the integral operator $L_{X|X^*}$ is still not injective. Such a special case is important for the situation where X^i are repeated measurements, which may have the same factor loading matrix M^i. In such a case one can't generate an injective measurement by combining the measurements.

This paper provides a set of sufficient conditions under which the model is fully identified, even when there are no injective measurements, single or combined. We show identification in two steps. First, we provide sufficient conditions to identify the factor loadings. Then, we provide stronger assumptions to show identification of distribution of latent variables without injectivity. In addition, we also show what we can achieve with injectivity under some testable assumptions.

3 Identification of factor loadings

We first identify the factor loadings under assumptions as follows:

Assumption 1. *The following conditions hold:*

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

$$E[\varepsilon^i | \varepsilon^j, \varepsilon^k, X^*] = 0. \tag{4}$$

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

2. *The latent vector X^* satisfy:*

$$\begin{aligned} E[X^*] &= 0 \\ E[X^*_{-i}(X^*_{-i})^T | X^*_i] &= E[X^*_{-i}(X^*_{-i})^T] \\ E[(X^*_i)^3] &\neq 0 \end{aligned} \tag{5}$$

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

In order to identify the factor loadings, we only need the measurement error vectors to be conditional mean independent of each other and the common factors X^*. The condition that unconditional mean is equal to zero holds without loss of generality in this linear setting. Notice that the measurement error in different dimension in one multivariate measurement can be correlated with each other arbitrarily. Therefore, we can't split a multivariate measurement to generate more measurements satisfying the same conditions.

Assumption 1.2 impose restrictions on the first three moments of the latent vector X^*. Assumption 1.2 guarantees that for $i \neq j$ and $i \neq k$

$$E[X^*_i \cdot X^*_j \cdot X^*_k] = 0. \tag{6}$$

which allows us to focus on nonzero third moments of X^* for further identification of factor loadings.

A key observation is that Assumption 1 implies

$$E(X^1_i \times X^2_u \times X^3_v) \;=\; \sum_{l=1}^{l=L} m^1_{i,l} \times m^2_{u,l} \times m^3_{v,l} \times E[(X^*_l)^3] \tag{7}$$

which presents a decomposition of a 3-way tensor on the left hand side. The Kruskal's Theorem (Kruskal, 1977b) implies that the elements on the right hand side, i.e., M^1, M^2, M^3, are unique up to permutation and scaling if the Kruskal ranks of M^1, M^2, M^3 satisfy

$$\kappa(M^1) + \kappa(M^2) + \kappa(M^3) \geq 2L + 2 \tag{8}$$

where $\kappa(M^i)$ defined as

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

The Kruskal rank of M^i describes how much information the measurement X^i has even when M^i is not injective or doesn't have a full column rank. Kruskal's Theorem implies that we can identify the factor loadings from the decomposition of the 3-way tensor. For example, identification is possible with $K_i = 4$ and $L = 5$ when M^i has a full column rank $\kappa(M^i) = K^i = 4$. That means three 4-dimensional measurements may recover a 5-dimensional latent vector. In summary, we have

Lemma 1. *Under Assumption 1, factor loading matrices M^1, M^2, M^3 in Equation 2 are unique up to permutation and scaling if*

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

Proof: See the Appendix.

4 Identification of distributions without injectivity

In this section, we show identification of distributions of latent variables without injectivity under assumptions as follows:

Assumption 2. *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.

This assumption suggests that the latent variable X^* and the measurement errors ε^1, ε^2, and ε^3 are jointly independent and that the multi-dimensional latent vector contains jointly independent factors. Such a independence allows us to focus on these factors one by one. First, we find two scalar measurements of one factor. Then, we show that the Kruskal rank condition in Equation (8) guarantees that the two measurements satisfy the conditions to use the Kotlarski's identity.

Given the definition of the Kruskal rank, this identification procedure applies to all the latent factors with any permutation of the columns or X_l^* for $l = 1, ..., L$. The distributions of measurement errors can then be identified by deconvolution under that Assumption 2.3, which implies $\phi_{\varepsilon^i} \neq 0$ due to $\phi_{X^i} = \phi_{\varepsilon^i} \phi_{M^i X^*}$. As shown in Hu and Shiu (2022), this condition is testable. Then, distributions of X^*, ε^1, ε^2 are nonparametrically identified with a closed-form solution as follows:

$$\phi_{X^*} = \phi_{X_1^*} \times ... \times \phi_{X_L^*} \tag{9}$$

$$\phi_{\varepsilon^i} = \frac{\phi_{X^i}}{\phi_{M^i X^*}} \tag{10}$$

In summary, we have

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

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

if

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

Proof: See the Appendix.

The Kruskal rank condition may be abstract to practitioners. In the case where all the factor loading matrices M^i have a full rank, we have

$$\kappa(M^i) = \begin{cases} K_i \text{ if } K_i < L \\ L \text{ if } K_i \geq L \end{cases} \tag{12}$$

That means practitioners can easily detect whether the rank condition holds. In the meantime, we can also interpret the rank condition as how many latent factors the data can identify. We have

Corollary 1. *Suppose that Assumptions 1 and 2 hold and M^i for $i = 1, 2, 3$ have a full rank. Then, the results in Theorem 1 hold if the number of latent factors satisfies*

$$L \leq \frac{K_1 + K_2 + K_3}{2} - 1 \tag{13}$$

where K_i is the dimension of X^i.

4.1 An illustration

In this section, we present an illustrative example to provide more details on how the Kruskal rank condition leads to two classical measurements. Suppose each of the three measurements X^i has $K_i = 7$ dimensions with $L = 8$ latent common factors X^* with the Kruskal ranks equal to $\kappa_i = 6$. Therefore, the Kruskal rank condition holds.

In this case, X^1 is a 7-dimentional vector

$$\begin{aligned} X^1 &= \left(m^1_{k,l}\right)_{k=1,...,7;l=1,...,8} X^* + \varepsilon^1 \\ &\equiv \left(m^1_{.,1} \quad m^1_{.,2} \quad ... \quad m^1_{.,8}\right) X^* + \varepsilon^1 \\ &\equiv \left(M^1_{7\times 6} \quad m^1_{.,7} \quad m^1_{.,8}\right) X^* + \varepsilon^1 \end{aligned} \tag{14}$$

where $M^1_{7\times 6}$ has a full column rank for any permutation of the columns in M^1. Therefore, we can find Q^1 so that

$$W^1 = Q^1 X^1 \equiv \left(\begin{pmatrix} I_{6\times 6} \\ 0 \end{pmatrix} \quad q^1_{.,7} \quad q^1_{.,8}\right) X^* + e^1 \tag{15}$$

Given the same permutation of the columns in M^1, we find a full column rank matrix from

the right side as follows:

$$\begin{aligned} X^2 &= \left(m^2_{k,l}\right)_{k=1,\ldots,7;l=1,\ldots,8} X^* + \varepsilon^2 \\ &\equiv \begin{pmatrix} m^2_{\cdot,1} & m^2_{\cdot,2} & \ldots & m^2_{\cdot,8} \end{pmatrix} X^* + \varepsilon^1 \\ &\equiv \begin{pmatrix} m^2_{\cdot,1} & m^2_{\cdot,2} & M^2_{7\times 6} \end{pmatrix} X^* + \varepsilon^2 \end{aligned} \tag{16}$$

and with a Q^2

$$\begin{aligned} W^2 = Q^2 X^2 &\equiv \begin{pmatrix} q^2_{\cdot,1} & q^2_{\cdot,2} & \begin{pmatrix} 0 \\ I_{6\times 6} \end{pmatrix} \end{pmatrix} X^* + e^2 \\ &\equiv \begin{pmatrix} q^2_{1,1} & q^2_{1,2} & 0 & 0 \\ q^2_{2,1} & q^2_{2,2} & 1 & 0 \\ * & * & 0 & I_{5\times 5} \end{pmatrix} X^* + e^2 \end{aligned} \tag{17}$$

Notice that $\kappa(M^2) = 6$ implies that

$$\begin{pmatrix} q^2_{1,1} \\ q^2_{2,1} \end{pmatrix} \neq 0.$$

Otherwise, there exist 6 columns, i.e., the 1st one and the last 5, which are linearly dependent so that the Kruskal rank of M^2 can't be 6. Without loss of generality, we let $q^2_{1,1} \neq 0$ We then consider the first element in each of the two transformed measurements.

$$\begin{aligned} W^1_1 &= X^*_1 + q^1_{1,7} X^*_7 + q^1_{1,8} X^*_8 + e^1_1 \qquad (18) \\ &\equiv X^*_1 + e_1 \\ W^2_1 &= q^2_{1,1} X^*_1 + q^2_{1,2} X^*_2 + e^2_1 \qquad (19) \\ &\equiv \gamma X^*_1 + e_2 \end{aligned}$$

The only common factor in these two scalar measurements is the first latent factor X^*_1. Given γ, we then identified the distribution of X^*_1 from the joint distribution of $(W^1_1, W^2_1/\gamma)$ by Kotlarski's Identity.

In order to identify the distribution of other factors, we may simply swap the columns and then apply the same procedure. For example, for factor X^*_3, we just apply the same procedure to

$$\begin{aligned} X^1 &= \begin{pmatrix} m^1_{\cdot,3} & m^1_{\cdot,2} & m^1_{\cdot,1} & m^1_{\cdot,4} & \ldots & m^1_{\cdot,8} \end{pmatrix} X^* + \varepsilon^1 \\ X^2 &= \begin{pmatrix} m^2_{\cdot,3} & m^2_{\cdot,2} & m^2_{\cdot,1} & m^2_{\cdot,4} & \ldots & m^2_{\cdot,8} \end{pmatrix} X^* + \varepsilon^2 \end{aligned} \tag{20}$$

In summary, we may identify 8 latent factors from three 7-dimensional measurements, while the Kruskal rank corresponding to each measurement is 6, i.e., without injectivity.

4.2 Extention: Identification with correlated factors

In many empirical applications, especially in the estimation of structural models, it is important to allow the unobserved variables in X^* to be correlated. In fact, we may extend our results to this case with modified assumptions as follows:

Assumption 3. *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* $\mathbb{R}^{K_1}$

Because the latent factors may be correlated, we are not able to identify the distribution of the factors one by one. Therefore, we impose full rank restrictions on the loading matrix as follows:

$$\text{Rank}(M^1) = \text{Rank} \begin{pmatrix} M^2 \\ M^3 \end{pmatrix} = L \tag{21}$$

The Kruskal rank codition in Equation (8) then becomes

$$\kappa(M^2) + \kappa(M^3) \geq L + 2 \tag{22}$$

These restrictions in Equation 21 implies that we may generate two classical measurements of the whole latent vector X^*, to which the multivariate Kotlarski's identity applies. That means the distribution of the latent vector is identified with a closed form. Notice that these rank condition and Assumptions 3.2 are testable from the data given the first step identification of the factor loadings.

Finally, We can then identify the distribution of all the latent random variables under additivity and independence. We have

$$\phi_{\varepsilon^i} \quad = \quad \frac{\phi_{X^i}}{\phi_{M^i X^*}} \tag{23}$$

where the denominsator is nonzero because $\phi_{X^1} = \phi_{\varepsilon^1} \phi_{M^1 X^*} \neq 0$ implies that $\phi_{X^*} \neq 0$ with a full rank M^1. In summary, we have

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

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

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

$$\kappa(M^2) + \kappa(M^3) \quad \geq \quad L + 2. \tag{25}$$

Proof: See the Appendix.

Here we assume that all the density functions are continuous. Therefore, it is not possible to map from a vector to a lower dimension vector without losing information. Technically, there is a one-to-one mapping from $\mathbb{R}^K$ to $\mathbb{R}$, but such a mapping can't be continuous for $K > 1$, which makes it less useful for practitioners. Also for this reason, the theorem in Hu and Schennach (2008) applies to the current model under three injectivity restrictions, i.e., for $i = 1, 2, 3$, M^i has a full column rank equal to L, i.e., $\text{Rank}(M^i) = L$, in order to identify all the distributions. The advantage of their results is that they can handle general nonlinear measurement functions, such as $X^i = g(X^*, \varepsilon^i)$.

Nevertheless, Our Theorem 1 achieves identification of all the distributions without injectivity and Proposition 1 identifies the model under weaker injectivity conditions. In particular, although it has stronger assumptions, Proposition 1 is more friendly to practitioner because the full rank condition and the nonzero ch.f. condition are both testable and easy to interpret.

5 Nonlinear cases: A generalization of Kruskal rank

In this section, we show identification without imposing linearity by introducing a newly defined generalized Kruskal rank for integral operators. Our analysis focuses on the effectiveness of this new rank concept, while maintaining other high-level assumptions.

We start with an indicator/sparsity matrix similar to a Jacobian matrix. Let $\mathrm{I}(\cdot)$ be the indicator function, which equals 1 if the statement is true and 0 otherwise. for For vectors of random variables $W \in \mathbb{R}^K$ and $X^* \in \mathbb{R}^L$, we define

$$\mathrm{J}(p_{W|X^*}) = \left[\mathrm{I}\left(\frac{\partial\, p_{W_i|X^*}(\cdot|x^*)}{\partial x_j^*} \neq 0 \right) \right]_{i=1,\ldots,K; j=1,\ldots,L} \tag{26}$$

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. Let $W_{1:k}$ be the sub-vector of the first k entries in vector W, $I_{K\times K}$ be a $K \times K$ identity matrix, and $R_{K\times L}$ be a $K \times L$ generic matrix of unspecified remainders.

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

Definition 1. *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* $1 \leq k \leq \kappa^s$.

In the previous linear case, the signal rank is equal to the Kruskal rank of the factor loading matrix and a mapping Q can be generated by the Gram–Schmidt process based on the loading matrix. Note that $\kappa^{\mathrm{s}}(L_{X|X^*}) \in \{0, 1, 2, ..., L\}$. We show that a signal rank condition, similar to a Kruskal rank condition for matrices, is crucial for the identification of distributions of latent variables, while keeping other assumptions at the high level for simplicity.

We assume

Assumption 4. *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

$$p(X^1, X^2, X^3) \;=\; \int p(X^1|X^*)p(X^2|X^*)p(X^3|X^*)p(X_1^*) \times ... \times p(X_L^*)dX^* \tag{27}$$

Our goal is to identify every probability function within the integral on the right hand side. 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:

$$X^* = (X_1^*; X^*_{L-\kappa_2^s+2}, X^*_{L-\kappa_2^s+3}, ..., X_L^*; X_2^*, X_3^*, ..., X^*_{L-\kappa_2^s+1})^T \tag{28}$$

with $W^2 = Q_2(X^2)$. We show that a sufficient condition for identification is that

$$\kappa_1^s + \kappa_2^s + \kappa_3^s \geq 2L + 2. \tag{29}$$

where κ_i^s is the signal rank of $L_{X^i|X^*}$. This condition guarantees that X_1^* is the only comment factor in the first elements in W^1 and W^2. Therefore, we can show that idiosyncratic elements in W_1^1, and W_1^2 are integrated out as follows:

$$p(W_1^1, W_1^2) \;=\; \int p(W_1^1|X_1^*)p(W_1^2|X_1^*)p(X_1^*)dX_1^* \tag{30}$$

In order to identify the whole distribution with a relatively simple procedure, we assume

Assumption 5. $\kappa_3^s = L$.

Kruskal's theorem does not require Assumption 5, that is, it does not rely on any of the matrices having full rank, yet it is supported by a more than ten-page proof involving highly sophisticated mathematics. The proof becomes far less complicated with one of the three matrices being full rank. That is why we adopt this assumption. It is possible that Assumption 5 is not necessary for our results as well, as we show in the linear case before. Given that this section focuses on the effectiveness of the newly defined signal rank, we mimic this relatively simple case in the proof of the Kruskal's theorem.

Assumption 5 implies that $L_{X^3|X^*}$ has a full signal rank with $W^3 = Q_3(X^3)$ satisfying

$$p(W_{1:L}^3|X^*) = \Pi_{l=1}^L p(W_l^3|X_l^*). \tag{31}$$

and that for $i = 1, 2$

$$p(W_1^3, W_1^i) \;=\; \int p(W_1^3|X_1^*)p(W_1^i|X_1^*)p(X_1^*)dX_1^* \tag{32}$$

We can show that three joint distributions $p(W_1^i, W_1^j)$ in Equations (30) and (32) identifies $L_{W_1^3|X_1^*}D_{X_1^*}(L_{W_1^3|X_1^*})^T$, which contains only one unknown - $p(W_1^3, X_1^*)$. We put this as a high-level normalization assumption as follows:

Assumption 6. *There exist $W^3 = Q_3(X^3)$ s.t. $L_{W_l^3|X_l^*}D_{X_l^*}(L_{W_l^3|X_l^*})^T$ uniquely determines $L_{W_l^3|X_l^*}$ and $D_{X_l^*}$, i.e., $p(W_1^3, X_1^*)$, for all l.*

Notice that the operator $L_{W_1^3|X_1^*}D_{X_1^*}(L_{W_1^3|X_1^*})^T$ can be considered as an operator based on the joint distribution of $(W_1^3, \tilde{W}_1^3)$ with

$$\begin{aligned} W_1^3 &= X_1^* + e^1 \\ \tilde{W}_1^3 &= X_1^* + \tilde{e}^1 \end{aligned} \tag{33}$$

where e^1 and $\tilde{e}^1$ have the same distribution conditional on X_1^*. If X_1^*, e^1, and $\tilde{e}^1$ on the right hand side are jointly independent and have a nonvarnishing ch.f., the Kortlarski's identity implies that the distributions of $(X_1^*, e^1, \tilde{e}^1)$, equivalently $p(W_1^3, X_1^*)$, are all identified with a closed form. Note that these derivations hold for $L_{W_l^3|X_l^*}$ as well. That means that Assumption 6 implies that $L_{W_l^3|X_l^*}$ for all l are identified.

Finally, we show the whole distribution $p(X^1, X^2, X^3, X^*)$ is identified from

$$p(X^1, X^2, W^3) \;=\; \int p(W^3|X^*)p(X^1, X^2, X^*)dX^*. \tag{34}$$

Because we have identified $L_{W_l^3|X_l^*}$ for all l and they are invertible, we can identify

$p(X^2, X^3, X^*)$. Given the mapping Q^3 is one-to-one and continuous, we also identify the distribution of $p(X^3 \mid X^*)$ from $p(W^3 \mid X^*)$. That means we have identified

$$p(X^1, X^2, X^3, X^*) = p(X^3|X^*)p(X^2, X^3, X^*) \tag{35}$$

We summarize the result as follows:

Theorem 2. *Suppose that Assumptions 4, 5, and 6 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*

$$p(X^1, X^2, X^3) = \int p(X^1|X^*)p(X^2|X^*)p(X^3|X^*)p(X^*)dX^* \tag{36}$$

if

$$\kappa^s(L_{X^1|X^*}) + \kappa^s(L_{X^2|X^*}) + \kappa^s(L_{X^3|X^*}) \geq 2L + 2. \tag{37}$$

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

Proof: See the Appendix.

6 Summary

This paper establishes new identification results for multivariate measurement-error models in which all observed measurements contain correlated noise and none provides an injective mapping of the latent vector. By exploiting the structure of third-order cross-moments, I construct a three-way tensor whose decomposition is uniquely determined under Kruskal's rank condition. This tensor decomposition identifies the factor loading matrices up to permutation and scaling. Building on a linear structure, the paper then identifies the full distribution of each latent component by constructing appropriate measurements and applying either scalar or multivariate versions of Kotlarski's identity. As a result, the joint distribution of the latent vector and all measurement errors is determined nonparametrically without requiring injective measurements. With the injectivity, we also achieve the identification under testable and less restrictive conditions.

The results in this paper widen the scope of what can be identified in measurement-error models, demonstrating that multivariate latent structures can be recovered in settings that would otherwise appear underidentified using existing techniques. The identification in the linear case is constructive in the sense that it leads to a straightforward estimation procedure. This makes the approach useful for empirical work involving noisy regressors, survey misreporting, and multidimensional factor models in economics and the social sciences.

Furthermore, this paper provides a generalized Kruskal rank - signal rank - for integral operators, and show that identification holds for general nonlinear models under a signal rank condition similar to the well-known Kruskal rank condition. The newly-defined signal

rank is consistent with the existing Kruskal rank for matrices and the injectivity of integral operators, and can be used to describe how informative a measurement is in the multivariate case. Although our identification of nonlinear models is not as constructive as that of linear models, the results here can be seen as a small step toward developing a continuous analogue of Kruskal's classical theorem. As shown in Allman et al. (2009), Kruskal's theorem delivers strong identification results for models with discrete latent variables. Extending these ideas to continuous latent variables would require a version of Kruskal's theorem that works in a continuous setting, which, to the best of my knowledge, is still an open question. The Hu–Schennach theorem provides a partial analogue, but only under an injectivity condition. This paper provides constructive identification for linear measurement error models and defines signal rank of integral operators for the multivariate case, which moves the literature closer towards a solution of the important open question.

7 Appendix

7.1 Proof of Lemma 1

We start with the linear model

$$\begin{aligned} X^1 &= M^1 X^* + \varepsilon^1 \\ X^2 &= M^2 X^* + \varepsilon^2 \\ X^3 &= M^3 X^* + \varepsilon^3 \end{aligned} \tag{38}$$

Assumption 1.1 implies that for $i \neq j$ and $i \neq k$

$$E(\varepsilon^i \times X^j \times X^k) = 0.$$

We then have

$$\begin{aligned} E(X_i^1 \times X_u^2 \times X_v^3) &= \sum_{j=1}\sum_{k=1}\sum_{l=1} E(m_{i,j}^1 X_j^* \times m_{u,k}^2 X_k^* \times m_{v,l}^3 X_l^*) \\ &= \sum_{j=1}\sum_{k=1}\sum_{l=1} m_{i,j}^1 \times m_{u,k}^2 \times m_{v,l}^3 \times E[X_j^* X_k^* X_l^*] \end{aligned} \tag{39}$$

Assumption 1.2 implies that for $l \neq j$ and $l \neq k$

$$\begin{aligned} E[X_j^* X_k^* X_l^*] &= E\left(E[X_j^* X_k^* | X_l^*] \times X_l^*\right) \\ &= E\left(E[X_j^* X_k^*] \times X_l^*\right) \\ &= E[X_j^* X_k^*] E(X_l^*) \\ &= 0 \end{aligned} \tag{40}$$

Therefore,

$$E(X_i^1 \times X_u^2 \times X_v^3) = \sum_{l=1}^{l=L} m_{i,l}^1 \times m_{u,l}^2 \times m_{v,l}^3 \times E[(X_l^*)^3] \quad (41)$$

These third order moments form a 3-way tensor. The Kruskal's Theorem (Kruskal (1977b)) implies that the elements on the right hand side, i.e., M^1, M^2, M^3 and $E[(X_i^*)^3]$, are unique up to permutation and scaling under the Kruskal rank condition in Equation (8) . That means that we have identified the factor loadings. QED.

7.2 Proof of Theorem 1

Given the independence assumption

$$p(\varepsilon^1, \varepsilon^2, \varepsilon^3) = p(\varepsilon^1) \times p(\varepsilon^2) \times p(\varepsilon^3) \quad (42)$$

and

$$p(X^*) = p(X_1^*) \times ... \times p(X_L^*) \quad (43)$$

we first show how to identify the distribution of each X_l^*. Then, the distributions of ε^i are identified by deconvolution from X^i.

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 and Q^2 s.t.

$$\begin{aligned} W^1 \equiv Q^1 X^1 &= Q^1 M^1 X^* + Q^1 \varepsilon^1 \quad (44) \\ &= \begin{pmatrix} I_{\kappa^1 \times \kappa^1} & R^1_{\kappa^1 \times (L-\kappa^1)} \\ 0 & R^1_{(K^1-\kappa^1)\times(L-\kappa^1)} \end{pmatrix} X^* + Q^1 \varepsilon^1 \\ &\equiv \begin{pmatrix} 1 & 0 & R^1_{1\times(L-\kappa^1)} \\ 0 & I_{(\kappa^1-1)\times(\kappa^1-1)} & R^1_{(\kappa^1-1)\times(L-\kappa^1)} \\ 0 & 0 & R^1_{(K^1-\kappa^1)\times(L-\kappa^1)} \end{pmatrix} X^* + Q^1 \varepsilon^1 \end{aligned}$$

where R^1 refers to a generic remainder matrix in M^1 with its dimensions in the subscript. Matrices $R^1_{(K^1-\kappa^1)\times(L-\kappa^1)}$ may not appear if $\kappa^1 = K^1$. The first element in W^1 can be considered as a measurement of X_1^* with other elements as its measurement error. Let $[M]_1$ stand for the first row of matrix M. WLOG, we consider the first element in X^*

$$\begin{aligned} W_1^1 &= X_1^* + R^1_{1\times(L-\kappa^1)}(X^*_{\kappa^1+1}, ..., X_L^*)^T + [Q^1]_1 \varepsilon^1 \\ &\equiv X_1^* + e_1 \quad (45) \end{aligned}$$

Next, we generate a second measurement of X_1^* from X^2, in which there is no other common factor than X_1^*. We will show that the Kruskal rank condition guarantees the existence of such a measurement. For that reason, we generate a $\kappa^2 \times \kappa^2$ identity matrix from the right side, i.e., starting from X_L^* in the matrix M^2.

Similarly, there exist Q^2 corresponding to M^2 in measurement X^2 s.t.

$$\begin{aligned} W^2 \equiv Q^2 X^2 &= Q^2 M^2 X^* + Q^2 \varepsilon^2 \\ &= \begin{pmatrix} R^2_{(K^2-\kappa^2)\times(L-\kappa^2)} & 0 \\ R^2_{(\kappa^2)\times(L-\kappa^2)} & I_{\kappa^2\times\kappa^2} \end{pmatrix} X^* + Q^2\varepsilon^2 \\ &\equiv \begin{pmatrix} R^2_{(K^2-\kappa^2)\times(L-\kappa^2)} & 0 & 0 \\ R^2_{1\times(L-\kappa^2)} & 1 & 0 \\ R^2_{(\kappa^2-1)\times(L-\kappa^2)} & 0 & I_{(\kappa^2-1)\times(\kappa^2-1)} \end{pmatrix} X^* + Q^2\varepsilon^2 \\ &\equiv \begin{pmatrix} \begin{pmatrix} R^2_{(K^2-\kappa^2)\times(L-\kappa^2)} & 0 \\ R^2_{1\times(L-\kappa^2)} & 1 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} R^2_{(\kappa^2-1)\times(L-\kappa^2)} & 0 \end{pmatrix} & I_{(\kappa^2-1)\times(\kappa^2-1)} \end{pmatrix} X^* + Q^2\varepsilon^2 \\ &\equiv \begin{pmatrix} R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)} & 0 \\ R^2_{(\kappa^2-1)\times(L-\kappa^2+1)} & I_{(\kappa^2-1)\times(\kappa^2-1)} \end{pmatrix} X^* + Q^2\varepsilon^2 \end{aligned} \tag{46}$$

where R^2 refers to a generic remainder matrix in M^2 with its dimensions in the subscript. In these derivations, we only reorganize the matrix so that we are able to locate a suitable measurement. We then pick an element in W^2 which does not contain $(X^*_{\kappa^1+1}, ..., X^*_L)$, which are in W_1^1, but does contain X_1^*. We consider the i-th row of $R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)}$ for all $i \leq L - \kappa^2 + 1$. 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

$$\begin{aligned} W_i^2 &= [R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)}]_{i,1} X_1^* + \sum_{j=2}^{L-\kappa^2+1} [R^2_{(K^2-\kappa^2+1)\times(L-\kappa^2+1)}]_{i,j} X_j^* + [Q^2]_i \varepsilon^2 \\ &\equiv r X_1^* + e_2 \end{aligned} \tag{47}$$

In summary, e_2 in Equation (47) contains $(X_2^*, ..., X^*_{L-\kappa^2+1}, \varepsilon^2)$ and e_1 in Equation (45) contains $(X^*_{\kappa^1+1}, ..., X^*_L, \varepsilon^1)$. Notice that the Kruskal rank condition in Equation (8) guarantees that

$$\kappa^1 + \kappa^2 \geq L + 2 \tag{48}$$

because $L \geq \kappa^3$ and $\kappa^1 + \kappa^2 + \kappa^3 \geq 2L + 2$. That means

$$\kappa^1 + 1 > L - \kappa^2 + 1. \tag{49}$$

Therefore, (X_1^*, e^1, e^2) have no common elements and therefore are jointly independent.

We then reveal the distribution of X_1^* from two scalar measurements (W_1^1, W_1^2) given that $r \neq 0$ has been identified. For each latent dimension, $p(X_1^*)$ is identified from its ch.f. by Kotlarski's Identity

$$\phi_{X_1^*}(t) = \exp\left(\int_0^t \frac{\left[\frac{\partial}{\partial t_2}\phi_{W_1^1,W_1^2}\left(s, \frac{t_2}{\gamma}\right)\right]_{t_2=0}}{\phi_{W_1^1}(s)} ds\right). \tag{50}$$

where ϕ_X is ch.f. of X. Given the definition of the Kruskal rank, this identification procedure applies to all X_l^* with any permutation of the columns in M^i or X_l^* in X^*. That means distributions of X^*, ε^1, ε^2 are nonparametrically identified with a closed-form solution as follows:

$$\phi_{X^*} = \phi_{X_1^*} \times ... \times \phi_{X_L^*} \tag{51}$$

$$\phi_{\varepsilon^i} = \frac{\phi_{X^i}}{\phi_{M^i X^*}} \tag{52}$$

where $\phi_{X^i} = \phi_{\varepsilon^i}\phi_{M^i X^*} \neq 0$. In summary, we have identified the distributions of X_1^*, ... ,X_l^*, ε^1, ε^2, ε^3 from the joint distribution $p(X^1, X^2, X^3)$. In other words, the joint distribution $p(X^1, X^2, X^3)$ uniquely determines the joint distribution $p(X^1, X^2, X^3, X^*)$, where

$$p(X^1, X^2, X^3, X^*) = p_{\varepsilon^1}(X^1 - M^1 X^*)p_{\varepsilon^2}(X^2 - M^2 X^*)p_{\varepsilon^3}(X^3 - M^3 X^*)p(X_1^*)...p(X_L^*). \tag{53}$$

QED.

7.3 Proof of Proposition 1

First, Lemma 1 suggests that Assumptions 1 and the Kruskal rank condition in Equation (8) guarantee that factor loading matrices M^1, M^2, M^3 in Equation 2 are identified up to permutation and scaling. We then show that all the distributions are still identified when the elements in X^* are correlated. Assumption 3.2 implies that we can use X^1 a primary measurement of X^* and combine X^2 and X^3 as a secondary measurement. We then use the multivariate Kotlarski's identity to identify the joint distribution of X^*.

The full column rank conditions in Equation (21) guarantees that there exist Q_1 and Q_{23} such that

$$W^1 \equiv Q^1 X^1 = X^* + Q^1 \varepsilon^1 \tag{54}$$

$$W^{23} \equiv Q^{23}\begin{pmatrix} X^2 \\ X^3 \end{pmatrix} = X^* + Q^{23}\begin{pmatrix} \varepsilon^2 \\ \varepsilon^3 \end{pmatrix} \tag{55}$$

We can then identify the distribution $p(X^*)$ from its ch.f. as follows: for $t, s \in \mathbb{R}^L$ and $\lambda \in [0,1]$

$$\phi_{X^*}(t) = \exp\left(\int_0^1 \frac{\left[\nabla_s \phi_{W^1,W^{23}}(\lambda t, s)\right]_{s=0}^T t}{\phi_{W^1}(\lambda t)} d\lambda\right). \tag{56}$$

The distributions of all other elements are also identified. We have

$$\phi_{\varepsilon^i} = \frac{\phi_{X^i}}{\phi_{M^i X^*}} \tag{57}$$

where $\phi_{X^1} = \phi_{\varepsilon^i}\phi_{M^i X^*} \neq 0$ implies that $\phi_{M^i X^*} \neq 0$ In summary, we have identified the distributions of X^*, ε^1, ε^2, ε^3 from the joint distribution $p(X^1, X^2, X^3)$. In other words, the joint distribution $p(X^1, X^2, X^3)$ uniquely determines the joint distribution $p(X^1, X^2, X^3, X^*)$, where

$$p(X^1, X^2, X^3, X^*) = p_{\varepsilon^1}(X^1 - M^1 X^*) p_{\varepsilon^2}(X^2 - M^2 X^*) p_{\varepsilon^3}(X^3 - M^3 X^*) p(X^*). \tag{58}$$

QED.

7.4 Proof of Theorem 2

Here we show the identification for the genearl nonlinear case. Assumption 4 leads to

$$\begin{aligned} p(X^1, X^2, X^3) &= \int p(X^1|X^*) p(X^2|X^*) p(X^3|X^*) p(X^*) dX^* \\ &= \int p(X^1|X^*) p(X^2|X^*) p(X^3|X^*) p(X_1^*) \times ... \times p(X_L^*) dX^* \end{aligned} \tag{59}$$

The goal is to identify every probability function within the integral on the right hand side. 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:

$$X^* = (X_1^*; X_{L-\kappa_2^s+2}^*, X_{L-\kappa_2^s+3}^*, ..., X_L^*; X_2^*, X_3^*, ..., X_{L-\kappa_2^s+1}^*)^T \tag{60}$$

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

$$\kappa_i^s := \kappa^s(L_{X^i|X^*}). \tag{61}$$

We start with the joint distribution

$$p(W^1, W^2) = \int p(W^1|X^*) p(W^2|X^*) p(X_1^*) \times ... \times p(X_L^*) dX^* \tag{62}$$

For the first elements in W^1 and W^2 satisfy

$$\begin{aligned} & p(W_1^1, W_1^2) \\ = & \int p(W_1^1|X^*)p(W_1^2|X^*)p(X_1^*)...p(X_L^*)dX^* \\ = & \int p(W_1^1|X_1^*, ..., X_L^*)p(W^2|X_1^*; X_{L-\kappa_2^s+2}^*, X_{L-\kappa_2^s+3}^*, ..., X_L^*; X_2^*, X_3^*, ..., X_{L-\kappa_2^s+1}^*) \\ & \times p(X_1^*) \times ... \times p(X_L^*)dX^* \end{aligned} \tag{63}$$

Given

$$\mathrm{J}(p_{W^1|X^*}) = \begin{pmatrix} I_{\kappa_1^s \times \kappa_1^s} & R_{\kappa_1^s \times (L-\kappa_1^s)} \\ R_{(K_1-\kappa_1^s)\times\kappa_1^s} & R_{(K_1-\kappa_1^s)\times(L-\kappa_1^s)} \end{pmatrix} \tag{64}$$

and

$$\mathrm{J}(p_{W^2|X^*}) = \begin{pmatrix} I_{\kappa_2^s \times \kappa_2^s} & R_{\kappa_2^s \times (L-\kappa_2^s)} \\ R_{(K_2-\kappa_2^s)\times\kappa_2^s} & R_{(K_2-\kappa_2^s)\times(L-\kappa_2^s)} \end{pmatrix} \tag{65}$$

they imply that W_1^1 does not contain $(X_2^*, ..., X_{\kappa_1^s}^*)$ and W_1^2 does not contain $(X_{L-\kappa_2^s+2}^*, ..., X_L^*)$ so that

$$\begin{aligned} & p(W_1^1, W_1^2) \\ = & \int p(W_1^1|X^*)p(W_1^2|X^*)p(X_1^*)...p(X_L^*)dX^* \\ = & \int p(W_1^1|X_1^*, X_{\kappa_1^s+1}^*, ..., X_L^*)p(W^2|X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*)p(X_1^*)...p(X_L^*)dX^* \end{aligned} \tag{66}$$

Note that $\kappa^s{}_3 \leq L$. The signal rank condition in Equation 37 implies

$$\kappa_1^s + \kappa_2^s \geq L + 2 \tag{67}$$

and

$$\kappa_1^s + 1 > L - \kappa_2^s + 1 \tag{68}$$

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

iosyncratic elements in W_1^1, and W_1^2 are integrated out as follows:

$$\begin{aligned} & p(W_1^1, W_1^2) \\ = & \int p(W_1^1|X_1^*, X_{\kappa_1^s+1}^*, ..., X_L^*) \times \\ & \times p(W_1^2|X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*) p(X_1^*)...p(X_L^*)dX^* \\ = & \int p(W_1^1, X_1^*, X_{\kappa_1^s+1}^*, ..., X_L^*) \times \\ & \times p(W_1^2|X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*) p(X_2^*)...p(X_{L-\kappa_2^s+1}^*)dX^* \\ = & \int p(W_1^1, X_1^*) p(W_1^2|X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*) p(X_2^*)...p(X_{L-\kappa_2^s+1}^*) dX_1^* dX_2^* ... dX_{L-\kappa_2^s+1}^* \\ = & \int p(W_1^1|X_1^*) p(W_1^2, X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*) dX_1^* dX_2^* ... dX_{L-\kappa_2^s+1}^* \\ = & \int p(W_1^1|X_1^*) p(W_1^2|X_1^*) p(X_1^*) dX_1^* \end{aligned} \tag{69}$$

Next, Assumption 5 implies that $L_{X^3|X^*}$ has a full signal rank with $W^3 = Q_3(X^3)$ and

$$\mathrm{J}(p_{W^3|X^*}) = \begin{pmatrix} I_{L\times L} \\ R_{(K_3-L)\times L} \end{pmatrix} \tag{70}$$

and therefore,

$$p(W_{1:L}^3|X^*) = \Pi_{l=1}^L p(W_l^3|X_l^*). \tag{71}$$

We then consider

$$\begin{aligned} & p(W_1^1, W_1^3) \\ = & \int p(W_1^1|X_1^*, X_{\kappa_1^s+1}^*, ..., X_L^*) p(W_1^3|X_1^*) p(X_1^*)...p(X_L^*)dX^* \\ = & \int p(W_1^1|X_1^*) p(W_1^3|X_1^*) p(X_1^*) dX_1^* \end{aligned} \tag{72}$$

Similarly, we have

$$\begin{aligned} p(W_1^3, W_1^2) &= \int p(W_1^3|X_1^*) p(W_1^2|X_1^*, X_2^*, ..., X_{L-\kappa_2^s+1}^*) p(X_1^*)...p(X_L^*)dX^* \\ &= \int p(W_1^3|X_1^*) p(W_1^2|X_1^*) p(X_1^*) dX_1^* \end{aligned} \tag{73}$$

In operator form, we have

$$\begin{aligned} L_{W_1^1,W_1^2} &= L_{W_1^1|X_1^*} D_{X_1^*} (L_{W_1^2|X_1^*})^T \\ L_{W_1^1,W_1^3} &= L_{W_1^1|X_1^*} D_{X_1^*} (L_{W_1^3|X_1^*})^T \\ L_{W_1^3,W_1^2} &= L_{W_1^3|X_1^*} D_{X_1^*} (L_{W_1^2|X_1^*})^T \end{aligned} \tag{74}$$

Given that these operators are invertable, we have

$$\begin{aligned} & L_{W_1^3,W_1^2} (L_{W_1^1,W_1^2})^{-1} L_{W_1^1,W_1^3} \\ = \; & L_{W_1^3|X_1^*} D_{X_1^*} (L_{W_1^2|X_1^*})^T (L_{W_1^1|X_1^*} D_{X_1^*} (L_{W_1^2|X_1^*})^T)^{-1} L_{W_1^1|X_1^*} D_{X_1^*} (L_{W_1^3|X_1^*})^T \\ = \; & L_{W_1^3|X_1^*} D_{X_1^*} (L_{W_1^3|X_1^*})^T \end{aligned} \tag{75}$$

where the left hand side is identified from three observed joint distributions. Assumption 6 then implies that $L_{W_l^3|X_l^*}$ and $D_{X_l^*}$, i.e., $p(W_1^3, X_1^*)$, for all l are identified.

Finally, we show the whole distribution $p(X^1, X^2, X^3, X^*)$ is identified. We consider

$$\begin{aligned} p(X^1, X^2, W^3) &= \int p(W^3|X^*) p(X^1, X^2, X^*) dX^* \\ &= \int p(W_1^3|X_1^*) ... p(W_L^3|X_L^*) p(X^1, X^2, X^*) dX^* \end{aligned} \tag{76}$$

Because we have identified $L_{W_l^3|X_l^*}$ for all l and they are invertible, we can identify $p(X^2, X^3, X^*)$. 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

$$p(X^1, X^2, X^3, X^*) = p(X^3|X^*) p(X^2, X^3, X^*). \tag{77}$$

QED.

References

Allman, Elizabeth S, Catherine Matias, and John A Rhodes, "Identifiability of Parameters in Latent Structure Models with Many Observed Variables," *The Annals of Statistics*, 2009, pp. 3099–3132.

Bonhomme, Stéphane, Koen Jochmans, and Jean-Marc Robin, "Nonparametric identification in finite mixtures of nonparametric product measures," *Journal of the Royal Statistical Society: Series B*, 2016, *78* (4), 1–32.

Carroll, J. Douglas and Jih-Jie Chang, "Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of "Eckart-Young" Decomposition," *Psychometrika*, 1970, *35* (3), 283–319.

Harshman, Richard A., "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, 1970, *1* (16), 1–84.

Hu, Yingyao, "Identification of Nonparametric Measurement Error Models with Discrete Data," *Econometrica*, 2008, *76* (1), 195–216.

_ **and Ji-Liang Shiu**, "A Simple Test of Completeness in a Class of Nonparametric Specification," *Econometric Reviews*, 2022, *41* (4), 373–399.

_ **and Susanne Schennach**, "Instrumental Variable Treatment of Nonclassical Measurement Error Models," *Econometrica*, 2008, *76*, 195–216.

Kotlarski, Ignacy, "On Some Characteristic Functions and a New Proof of a Basic Lemma in the Theory of Characteristic Functions," *Annals of Mathematical Statistics*, 1967, *38*, 1719–1722.

Kruskal, Joseph B., "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and Its Applications*, 1977, *18* (2), 95–138.

_ , "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Linear Algebra and its Applications*, 1977, *18* (2), 95–138.

Sidiropoulos, Nicholas D., Rasmus Bro, and Georgios B. Giannakis, "Parallel Factor Analysis in Sensor Array Processing," *IEEE Transactions on Signal Processing*, 2000, *48* (8), 2377–2388.