Identifiability of Polychoric Models with Latent Elliptical Distributions

Che Cheng, Hau-Hung Yang and Yung-Fong Hsu

Department of Psychology, National Taiwan University, Taiwan , R.O.C.

Corresponding author: Yung-Fong Hsu; Email: yfhsu@ntu.edu.tw

Abstract

The family of polychoric models (PM) categories ordinal data with latent multivariate normal variables. This modeling framework is commonly used to study the association between ordinal variables, often leading to a polychoric correlation model (PCM). Moreover, PM subsumes several well-known psychometric models, such as the structural equation modeling (SEM) with ordinal data. That said, the identifiability of PM has not been addressed in the literature. Meanwhile, in recent years researchers have suggested that the latent variables underlying PM could be generalized to the family of elliptical distributions, such as the multivariate logistic and *t* distributions. This article concerns the identifiability of PM and PCM with latent elliptical distributions, for which we show that PM is not identifiable and PCM is identifiable. In particular, we prove the identifiability of the polychoric *t* correlation model based on the copula representation. We then move on to find the set of identifiability constraints of PM through an "equivalence-classes approach of identifiability," and demonstrate its use in two applications: one concerns the identifiability of PM on Likert scales and on comparative judgment, and the other concerns the identifiability of ordinal SEM and item factor analysis. Possible implications induced by these identifiability constraints are discussed.

Keywords: elliptical distributions; identifiability constraints; just-identified models; polychoric correlations; polychoric models

Treating ordinal variables as interval or continuous variables might produce biased results (Olsson, 1979). The (family of) *polychoric models* (PM) deals with ordinal data by categorizing them with latent multivariate normal variables. PM mainly appears in two situations. First, PM is used to study the association between ordinal variables, often leading to a *polychoric correlation model* (PCM). PCM assumes that the latent variables underlying two ordinal variables follow the bivariate normal distribution with zero mean and a correlation matrix known as the *polychoric correlation* (Olsson, 1979). In particular, if both variables are binary, a polychoric correlation reduces to a tetrachoric correlation (Bonett & Price, 2005; Pearson, 1901).

Second, PM subsumes several commonly used psychometric models. In particular, PM is a general model of structural equation modeling with ordinal data (ordinal SEM; Muthén, 1984). Ordinal SEM models PM's mean vector and covariance matrix as a function of some parameters. Several models are special cases of ordinal SEM, such as the graded response model (Samejima, 1968; Samejima, 1997) and the family of item factor analysis models (Wirth & Edwards, 2007).

Although PM is the basis of several models, its identifiability has not been addressed in the literature; only the identification of PCM (which, as mentioned above, is a special case of PM) has been proved

(Almeida & Mouchart, 2003a). The importance of establishing the identifiability of a statistical model cannot be overstated. If a model is not identifiable, one may have two sets of parameters with the same probability distribution, even when the sample size approaches infinity. In this case, all estimators would not be consistent estimators.

To make the issue a bit more complicated, the normality assumption underlying PM has been challenged recently; researchers have suggested that the latent variables underlying PM could be generalized to the family of elliptical distributions, such as the multivariate logistic distribution and the multivariate t distribution (Jin & Yang-Wallentin, 2017; Kolbe et al., 2021; Roscino & Pollice, 2006). In light of this, in this article we explore these more general elliptical distributions. Two unsolved questions can be posed: (a) Are PM and/or PCM with latent elliptical distributions identifiable? (b) If any of them is not identifiable, can we find the minimal identifiability constraints? Practically, even if PCM is identifiable, it is not practical in some situations. For instance, when modeling the developmental changes of children, it is unreasonable to assume that all of the mean vectors are zero (McArdle et al., 2015; Muthén, 1984). Therefore, finding some other reasonable identifiability constraints is a task with practical significance.

This article aims at answering the above two questions and is organized as follows. Section 1 gives formal definitions of PM and PCM with elliptical distributions. We address the issue of (a) in Section 2. By generalizing Almeida and Mouchart's (2003a) argument, we show that PCM with elliptical distributions is identifiable. In particular, we prove the identification of the polychoric t correlation model based on the copula representation. On the other hand, PM with elliptical distributions is not identifiable. We address the issue of (b) in Section 3. We show that one can find the identifiability constraints of PM through the equivalence-classes approach of Tsai (2000, 2003). This approach can also help determine the measurement scales of latent variables. The minimal identifiability constraints of PM on Likert scales and also on comparative judgment are demonstrated. Moreover, we prove a theorem stating the necessary and sufficient conditions for the identifiability of ordinal SEM and item factor analysis. Section 4 is devoted to the discussion of possible implications and applications induced by these identifiability constraints.

1. Definitions of PM and PCM with elliptical distributions

Suppose there are K ordinal-scale item scores, $W_1, \ldots, W_K$ with support $\{1, \ldots, r_k\}$, respectively. Let $W = (W_1, \ldots, W_K)$ be the response vector. The family of polychoric models (PM) assumes that for an ordinal variable W_k, there is a corresponding vector of cut-offs (or thresholds) $a^{(k)} = \left(a_j^{(k)} : -\infty = a_0^{(k)} < a_1^{(k)} < \cdots < a_{r_k}^{(k)} = \infty \right)$ and a latent random variable X_k^* such that

$$W_k \leq j \iff X_k^* \leq a_j^{(k)}, \quad j = 1, \ldots, r_k.$$

That is,

$$P\left(\mathbf{W} \leq (j_1, \ldots, j_K)^\top \right) = P\left(\mathbf{X}^* \leq \left(a_{j_1}^{(1)}, \ldots, a_{j_K}^{(K)} \right)^\top \right). \tag{1}$$

where $\mathbf{X}^* = \left(X_1^*, \ldots, X_K^* \right)^\top$ follows a distribution function F. Thus, PM can be parametrized by

$$\theta_{PM} = (F, \mathcal{A}) \in \Theta_{PM},$$

where $\mathcal{A} = \{a^{(k)}\}_{k=1,\ldots,K}$, and Θ_{PM} is the parameter space for this parametrization.

In the following, we review the concept of spherical and elliptical distributions with some examples.

Definition 1.1. (Fang et al., 1990)

(i) Let $\mathbf{X}$ be a p-dimensional random vector. $\mathbf{X}$ is said to have a spherical distribution if for any orthogonal transformation Γ, $\mathbf{\Gamma X} \stackrel{d}{=} \mathbf{X}$, where $\stackrel{d}{=}$ means "equal in distribution."

(ii) Let $\mathbf{Y}$ be an n-dimensional random vector. $\mathbf{Y}$ is said to have an elliptical distribution with parameters $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ if

$$\mathbf{Y} \overset{d}{=} \boldsymbol{\mu} + \mathbf{A}^\top \mathbf{X},$$

where $\mathbf{A} \in \mathbb{R}^{p \times n}, \mathbf{A}^\top \mathbf{A} = \boldsymbol{\Sigma}$ with $\operatorname{rank}(\boldsymbol{\Sigma}) = p$.

The following proposition characterizes the spherical and elliptical distributions.

Proposition 1.2. *(Fang et al., 1990)*

(i) $\mathbf{X}$ *is spherically distributed if and only if there is a scalar function* $\phi(\cdot)$ *such that the characteristic function of* $\mathbf{X}$ *satisfies* $\psi_{\mathbf{X}}(\mathbf{t}) = \phi(\mathbf{t}^\top \mathbf{t})$. *The function* $\phi(\cdot)$ *is called the characteristic generator of the spherical distribution.*

(ii) *Under Definition 1.1(ii), The characteristic function of* $\mathbf{Y}$ *satisfies* $\psi_{\mathbf{Y}}(\mathbf{t}) = e^{i\mathbf{t}^\top \boldsymbol{\mu}} \phi(\mathbf{t}^\top \boldsymbol{\Sigma} \mathbf{t})$. *Therefore, an elliptical distribution can be parametrized by* $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi$, *denoted* $EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

The parameters of (a family of) elliptical distributions are named below.

Definition 1.3. $\mathbf{Y} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, where $\boldsymbol{\mu}$ is the location vector, $\boldsymbol{\Sigma}$ is the dispersion matrix or scatter matrix. If $\boldsymbol{\Sigma}$ is positive definite, then $\mathbf{P} = \operatorname{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}} \boldsymbol{\Sigma} \operatorname{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}}$ is called the *pseudo-correlation matrix*. If the first and second moments of $\mathbf{Y}$ exist, then $\boldsymbol{\mu}$ is called the *mean vector*, $\boldsymbol{\Sigma}$ is called the *covariance matrix*, and $\mathbf{P}$ is called the *correlation matrix*.

Some commonly used elliptical distributions include the normal, logistic, uniform, and t distributions.

Example 1.4. $\mathbf{Y} \sim EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$. Then $\mathbf{Y}$ follows

(i) a multivariate Normal distribution if $\phi(u) = \exp\left(-\frac{u}{2}\right)$. Also, we denote $\mathbf{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

(ii) a multivariate logistic distribution if $\phi(u) = 4\exp(-u)/(1 + \exp(-u))^2$.

(iii) a multivariate uniform distribution if $\phi(u) = 2I_{\{u<1\}}$.

(iv) a multivariate t distribution if $\phi_v(u) = (1 + u/v)^{-(v+2)/2}$. Also, we denote $\mathbf{Y} \sim t_{n,v}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The characteristic function of a multivariate t distribution depends on the degrees of freedom parameter v.

Definition 1.5.

(i) PM with a latent elliptical distribution characterized by ϕ is $P_{\theta_{PM}}$, where $\theta_{PM} = (EC_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi), \mathcal{A})$.

(ii) PCM with a latent elliptical distribution characterized by ϕ is $P_{\theta_{PCM}}$, where $\theta_{PCM} = (EC_n(\mathbf{0}, \mathbf{P}, \phi), \mathcal{A})$.

Unless specifically stated, in this article we study PCM and PM within the realm of elliptical distributions.

2. Identifiability of PCM with elliptical distributions

We first define some terms concerning model identifiability:

Definition 2.1. $\mathcal{P}_\theta = \{P_\theta : \theta \in \Theta\}$ is a parametric model. Let $\theta^* = g(\theta) \in \Theta^* \subset \Theta$. Then

(i) For any $\theta_1, \theta_2 \in \Theta, \theta_1, \theta_2$ are empirically indistinguishable if $P_{\theta_1} = P_{\theta_2}$, i.e., for any response w, $P_{\theta_1}(w) = P_{\theta_2}(w)$.

(ii) $\mathcal{P}_\theta$ is identifiable (or θ is identifiable) if $P_{\theta_1} = P_{\theta_2} \rightarrow \theta_1 = \theta_2$ for all $\theta_1, \theta_2 \in \Theta$.

(iii) $\mathcal{P}_\theta$ is just-identified (or θ is just-identified) if $\forall P \in \mathcal{P}_\theta$, there is a unique θ such that $P = P_\theta$.

(iv) $\mathcal{P}_\theta$ is partially identifiable over θ^* (or θ^* is partially identifiable) if $P_{\theta_1} = P_{\theta_2} \to \theta_1{}^* = \theta_2{}^*$ for all $\theta_1, \theta_2 \in \Theta^*$.

(v) For any $\theta_0 \in \Theta$, the identified set of P_{θ_0} is $\{\theta : P_\theta = P_{\theta_0}\}$.

In ordinary language, two sets of parameters are empirically indistinguishable if they have the same probability structure for all possible outcomes. A model is identifiable if distinct sets of parameters correspond to distinct distribution functions; thereby, constructing consistent estimators is possible. Similarly, a model is partially identifiable over θ^* if distinct sets of θ^* correspond to distinct distribution functions.

The following proposition states that if the model is parametrized by θ, then the model is identifiable if and only if the model is just-identified. Therefore, if the model is parameterized by θ, there is no need to distinguish between whether the model is just-identified or over-identified. The distinguishment between just-identified or over-identified is meaningful only when considering the restricted model of a parameterized model (see Section 3.1).

Proposition 2.2. $\mathcal{P}_\theta = \{P_\theta : \theta \in \Theta\}$. $\mathcal{P}_\theta$ *is identifiable iff $\mathcal{P}_\theta$ is just-identified.*

Proof. ($\supseteq$) The uniqueness of θ implies the identification of $\mathcal{P}_\theta$.

($\subseteq$) The P_θ is parametrized by θ, so for any $P \in \mathcal{P}_\theta$, there is a θ such that $P = P_\theta$. The uniqueness of θ is guaranteed by the identification of P_θ. □

Note that parameter identification is a necessary condition for the existence of a consistent estimator. This proposition has been implicitly stated in some studies (such as Gu & Xu, 2019; Ouyang & Xu, 2022). San Martín and Quintana (2002) provided a formal formulation and a proof of the proposition (see below, where we also provide an alternative proof).

Proposition 2.3. *(San Martín & Quintana, 2002) $\mathcal{P}_\theta = \{P_\theta : \theta \in \Theta\}$ is a statistical model and $X_1, \ldots, X_n$ are independently, identically distributed random variables from P_θ. Further, g is an invertible function of. If $\mathcal{P}_\theta$ is not identifiable, then there is no consistent estimator of $g(\theta)$. That is, the identifiability of the parameter is a necessary condition for the existence of a consistent estimator.*

Proof. If the estimator $\delta_n = \delta(X_1, \ldots, X_n)$ is a consistent estimator of $g(\theta)$, then $\delta_n \to g(\theta)$ under P_θ for any $\theta \in \Theta$. Given $P_{\theta_1} = P_{\theta_2}$ while $\theta_1 \neq \theta_2$, we have $\delta_n \to g(\theta_1)$ under P_{θ_1} and $\delta_n \to g(\theta_2)$ under P_{θ_2}. Because $P_{\theta_1} = P_{\theta_2}$, δ_n has the same distribution under P_{θ_1} and P_{θ_2} based on the inversion formula. Since a convergence sequence has a unique limit, $g(\theta_1) = g(\theta_2)$. Because g is invertible, we have $\theta_1 = \theta_2$.—a contradiction. □

Identifiability is a property of a statistical model, sometimes defined as a property of parameters (Casella & Berger, 2002, p. 523). However, the statement "θ is identifiable" is ambiguous because it does not specify the statistical model in consideration. For example, Almeida and Mouchart (2003a) showed that PCM under the normality assumption is identifiable in PCM but not in PM. Indeed, the latter fact was derived from a proposition in Almeida and Mouchart (2003a).

Proposition 2.4. *(Almeida and Mouchart, 2003a) For any monotonic increasing functions $g_1, \ldots, g_k$, define $g = (g_1, \ldots, g_k)$ as a component-wise transformation. Then $\theta_{PM} = (F, A)$ is empirically indistinguishable from $\theta_{PM,g} = (F_g, A_g)$, where $F_g = F \circ g^{-1}$ and $A_g = \left\{ g_j \left(a_j^{(k)} \right) \right\}_{j=1,\ldots,r_k-1; k=1,\ldots,K}$.*

2.1. Olsson's (1979) argument of identification of PCM

Olsson (1979) worked on the identification problem under PCM (with the normality assumption) by counting the difference between the number of parameters and the number of independent proportions. For convenience, we define the degrees of freedom, *df*, as

$$df = \#\text{independent proportions} - \#\text{parameters}.$$

This definition is consistent with the common usage of degrees of freedom (Rodgers, 2019). Olsson (1979) argued that PCM is identified because $df \geq 0$. In particular, the model is *just-identified* if $df = 0$ and *over-identified* if $df > 0$. However, Proposition 2.4 implies that Olsson's (1979) argument of identifiability is flawed, in that even for PCM under the normality assumption, $\theta_{PCM} = (\mathcal{N}_K(\mathbf{0},\mathbf{P}),\mathcal{A})$ satisfying $df \geq 0$, there could be a model $\theta_{PCM,g} = \left(\mathcal{N}_K(\mathbf{0},\mathbf{P})_g,\mathcal{A}_g\right)$ indistinguishable from it. Thus, merely observing $df \geq 0$ for a model does not guarantee that the model is identifiable. A similar statement (showing that $df \geq 0$ is a necessary but not sufficient condition for model identification) also can be found in the SEM literature (the t rule in Bollen, 1989).

In the literature, Almeida and Mouchart (2003a) provided the first rigorous proof of identifiability of PCM under the normality assumption. In the following, we generalize their proof to the elliptical distribution case. Specifically, using Proposition 2.2, we show that PCM is always just-identified under its model assumptions (see Theorem 2.5).[1]

2.2. *Generalization of Almeida and Mouchart's (2003a) proof*

Motivated by Theorem 3.1 in Almeida and Mouchart (2003a), we now establish the conditions of identification under PCM with any latent continuous elliptical distributions.

Theorem 2.5. *Under PCM with a latent elliptical distribution characterized by $\phi,P_{\theta_{PCM}}$, where $\theta_{PCM} = (EC_K(\mathbf{0},\mathbf{P},\phi),\mathcal{A})$ and ϕ is known. If*

   (i) *ϕ defines a continuous distribution with strictly increasing univariate and bivariate CDFs on its support;*
   (ii) *the polychoric (pseudo-)correlation matrix $\mathbf{P}$ is positive definite,*
   (iii) *then PCM ($P_{\theta_{PCM}}$) is just-identified.*

Proof. It suffices to prove that if two pairs of parameters correspond to the same response probability vector, say, $(\mathbf{P},\mathcal{A})$ and $(\tilde{\mathbf{P}},\tilde{\mathcal{A}})$, then $(\mathbf{P},\mathcal{A}) = (\tilde{\mathbf{P}},\tilde{\mathcal{A}})$. Our proof is composed of two parts:

   (i) The cut-offs (or thresholds) are identifiable: Without loss of generality, consider the jth cut-off of item k, say, $a_j^{(k)}$ and $\tilde{a}_j^{(k)}$. Let the response vector $\mathbf{w} = (r_1,\dots,r_{k-1},j,r_{k+1},\dots,r_K)$. Then $P(\mathbf{W} \leq \mathbf{w}) = F_{EC_1(0,1,\phi)}\left(a_j^{(k)}\right) = F_{EC_1(0,1,\phi)}\left(\tilde{a}_j^{(k)}\right)$. Since the CDF is injective, $a_j^{(k)} = \tilde{a}_j^{(k)}$. Because $a_j^{(k)}$ can be any cut-off in $\mathcal{A}$, we have $\mathcal{A} = \tilde{\mathcal{A}}$.
   (ii) The pseudo-correlation matrix is identifiable: Consider a response vector $\mathbf{w} = (r_1,\dots,r_{l-1},i, r_{l-1},\dots,r_{k-1},j,r_{k+1},\dots,r_K)$, for which the lth response is i, and the kth response is j. We define
   $$g(\rho_{lk}) = P(\mathbf{W} \leq \mathbf{w}) = P\left(\mathbf{X}^* \leq \left(a_j^{(l)},a_i^{(k)}\right)^{\top}\right) = F_{EC_2(\mathbf{0},\mathbf{P}_{lk},\phi)}\left(a_j^{(l)},a_i^{(k)}\right), \text{ where } \mathbf{P}_{lk} = \begin{pmatrix} 1 & \rho_{lk} \\ \rho_{lk} & 1 \end{pmatrix}.$$
   Because $\mathbf{P}$ is positive definite, $\mathbf{P}_{lk}$ is also positive definite. So the Cholesky decomposition is unique: $\mathbf{P}_{lk} = \mathbf{\Lambda}\mathbf{\Lambda}^{\top} = \begin{pmatrix} 1 & 0 \\ \rho_{lk} & \sqrt{1-\rho_{lk}^2} \end{pmatrix}\begin{pmatrix} 1 & \rho_{lk} \\ 0 & \sqrt{1-\rho_{lk}^2} \end{pmatrix}$, and $\mathbf{\Lambda}$ is invertible. We have $g(\rho_{lk}) = P\left(\mathbf{\Lambda}^{-1}\mathbf{X}^* \leq \mathbf{\Lambda}^{-1}\left(a_j^{(l)},a_i^{(k)}\right)^{\top}\right) = F_{EC_2(\mathbf{0},\mathbf{I}_2,\phi)}\left(a_j^{(l)},\frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}}a_j^{(l)} + \frac{1}{\sqrt{1-\rho_{lk}^2}}a_i^{(k)}\right)$.

$\square$

---

[1] At first glance, the statement of just-identifiability of PCM seems somewhat counter-intuitive, especially it is contradictory to the statement that $df \geq 0$. However, we only state that PCM is just-identified under its model assumptions, but not under a more general framework. For instance, consider a set of data consisting of K ordinal-scale items, and each item has r_k possible categories with probability p_k. Then we can treat PCM as a model nested in this framework. In such a case, PCM may not be just-identified (see Almeida and Mouchart, 2003a).

Since the bivariate CDF is monotone, $g\left(\rho_{lk}\right)$ is strictly decreasing iff $h\left(\rho_{lk}\right) = \frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}}a_j^{(l)} + \frac{1}{\sqrt{1-\rho_{lk}^2}}a_i^{(k)}$ is strictly decreasing. Based on the calculus, we know that $h\left(\rho_{lk}\right)$ is strictly decreasing. Thus $g\left(\rho_{lk}\right)$ is strictly decreasing, and so ρ_{lk} is identified. Because i and j can be any pair of items, we have $\mathbf{P} = \tilde{\mathbf{P}}$.

Based on Proposition 2.2, the model is just-identified because it is identifiable and also a parametrized model.

Corollary 2.6. *The multivariate normal distribution, multivariate t distribution with known degrees-of-freedom, multivariate logistic distribution, and multivariate uniform distribution all satisfy the conditions in Theorem 2.5, so their corresponding $\theta_{PCM} = \left(EC_K\left(\mathbf{0},\mathbf{P},\phi\right),\mathcal{A}\right)$ is identifiable. In particular, the multivariate normal case corresponds to Theorem 3.1 in Almeida and Mouchart (2003a).*

Remark 2.7. *The multivariate t distribution $\left(t_v\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)\right)$ does not satisfy the conditions in Theorem 2.5, because the function $\phi(u) = \left(1+u/v\right)^{-(v+2)/2}$ depends on the degrees-of-freedom parameter v, which could be unknown.*

As mentioned in Remark 2.7, Theorem **2.5** cannot cover the case of PCM with a latent multivariate t distribution (abbreviated as *polychoric t correlation model*). In the following, we provide a proof for this case based on the copula approach proposed by Almeida and Mouchart (2003b). For brevity, in the following paragraphs, we denote the CDF of a distribution simply by the symbol of that distribution. For example, we denote the CDF of the t distribution with degrees of freedom v as t_v. And the CDF of a multivariate t distribution $t_v\left(\boldsymbol{\mu},\boldsymbol{\Sigma}\right)$ is abbreviated as $t_{v,\boldsymbol{\mu},\boldsymbol{\Sigma}}$.

2.3. The copula approach to the identifiability of PCM

First, we introduce the concept of copula and define the copulas of elliptical distributions, under which the bivariate t copula will be a special case.

Definition 2.8. (Hofert, 2018) A copula is a multivariate CDF with standard uniform univariate margins, that is, the Unif(0,1) margins.

Definition 2.9. The elliptical copula characterized by ϕ with *pseudo-correlation* $\mathbf{P} = \text{diag}\left(\boldsymbol{\Sigma}\right)^{-\frac{1}{2}}\boldsymbol{\Sigma}\,\text{diag}\left(\boldsymbol{\Sigma}\right)^{-\frac{1}{2}}$ is defined as

$$C_{\mathbf{P}}^{\phi}\left(u_1,u_2,\ldots,u_K\right) = F_{EC_K\left(\mathbf{0},\mathbf{P},\phi\right)}\left(F^{-1}_{EC_1(0,1,\phi)}\left(u_1\right),\ldots,F^{-1}_{EC_1(0,1,\phi)}\left(u_K\right)\right).$$

Definition 2.10. Let $|\rho| \le 1$ and $v > 0$. Then $C_{\rho,v}^t$ is a bivariate t copula if

$$C_{\rho,v}^t\left(u_1,u_2\right) = t_{v,\mathbf{0},\mathbf{P}}\left(t_v^{-1}\left(u_1\right),t_v^{-1}\left(u_2\right)\right),$$

where $\mathbf{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, u_1, and u_2 are within $[0,1]$.

Remark 2.11. *The elliptical copula does not depend on the scale of the elliptical distribution. That is,*

$$F_{EC_K\left(\mathbf{0},\mathbf{P},\phi\right)}\left(F^{-1}_{EC_1(0,1,\phi)}\left(u_1\right),\ldots,F^{-1}_{EC_1(0,1,\phi)}\left(u_K\right)\right)$$
$$= F_{EC_K\left(\mu,\boldsymbol{\Sigma},\phi\right)}\left(F^{-1}_{EC_1\left(\mu_1,\sigma_1^2,\phi\right)}\left(u_1\right),\ldots,F^{-1}_{EC_1\left(\mu_K,\sigma_K^2,\phi\right)}\left(u_K\right)\right),$$

where $\mathbf{P} = \text{diag}\left(\boldsymbol{\Sigma}\right)^{-\frac{1}{2}}\boldsymbol{\Sigma}\,\text{diag}\left(\boldsymbol{\Sigma}\right)^{-\frac{1}{2}}$, $\left(\boldsymbol{\Sigma}\right)_{kk} = \sigma_k^2$, and $\boldsymbol{\mu} = \left(\mu_1,\ldots,\mu_K\right)^\top$. Therefore, to avoid overparameterization, the parameters of elliptical copula do not contain $\mu_k,\sigma_k,k = 1,\ldots,K$.

Almeida and Mouchart (2003b) found that PCM can be reparametrized by the copulas. Under this parametrization, the cut-offs are no longer the parameters to be considered. We formulate this property as a lemma.

Lemma 2.12. *(Almeida & Mouchart, 2003b) Let $\theta_{PCM} = \left(EC_p\left(\mathbf{0},\mathbf{P},\phi\right),\mathcal{A}\right)$. If ϕ is a continuous distribution with strictly increasing univariate and bivariate CDFs on its support. Also, the k-th marginal is*

$F_{EC_1(0,1,\phi)}$. *Let* $u_j^{(k)} = F_{EC_1(0,1,\phi)}\left(a_j^{(k)}\right), j = 1, \ldots, r_k$, *and* $u^{(k)} = \left(u_j^{(k)} : 0 = u_0^{(k)} < u_1^{(k)} < \cdots < u_{r_k}^{(k)} = 1\right)$. *Then PCM can be reparametrized by a one-to-one correspondence function* $\psi : \Theta_{PCM} \to \Theta_{PCM}^{CO}$, *defined by* $(\phi, \mathbf{P}) \mapsto C_{\mathbf{P}}^\phi$, *where*

$$\theta_{PCM}^{CO} = \left(C_{\mathbf{P}}^\phi, U\right) \in \Theta_{PCM}^{CO},$$

$U = \left\{u^{(k)} : k = 1, \ldots, K\right\}$, *and* Θ_{PCM}^{CO} *is the parameter space for this parametrization.*

Proof. The proof can be derived from Equation (1):

$$P\left(\mathbf{W} \le (j_1, \ldots, j_K)^\top\right) = P\left(\mathbf{X}^* \le \left(a_{j_1}^{(1)}, \ldots, a_{j_K}^{(K)}\right)^\top\right).$$
$$= C_{\mathbf{P}}^\phi\left(F_{EC_1(0,1,\phi)}\left(a_{j_1}^{(1)}\right), \ldots, F_{EC_1(0,1,\phi)}\left(a_{j_K}^{(K)}\right)\right)^\top\right)$$
$$= C_{\mathbf{P}}^\phi\left(u_{j_1}^{(1)}, \ldots, u_{j_K}^{(K)}\right).$$

That is, two sets of parameters θ_{PCM} and $\psi(\theta_{PCM})$ define the same CDF. Moreover, the assumption of ϕ guarantees that $F_{EC_1(0,1,\phi)}$ is invertible. It implies that the cut-offs can be uniquely determined through $u_j^{(k)}$ by $F_{EC_1(0,1,\phi)}^{-1}\left(u_j^{(k)}\right) = a_j^{(k)}, j = 1, \ldots, r_k$. So the function ψ defines a one-to-one correspondence between two parameter spaces. □

Using Lemma 2.12, in the following, we construct another proof of Theorem 2.5 based on the parametrization of copulas.

Proof of Theorem 2.5 based on copula parametrization. Based on condition (i) and Lemma 2.12, the identifiability of PCM under the copula parameterization (θ_{PCM}) implies the identifiability of PCM (θ_{PCM}^{CO}). We claim that if two pairs of copula parametrized parameters, $\left(C_{\mathbf{P}}^\phi, U\right)$ and $\left(C_{\tilde{\mathbf{P}}}^\phi, \tilde{U}\right)$, correspond to the same response probability vector, then $\left(C_{\mathbf{P}}^\phi, U\right) = \left(C_{\tilde{\mathbf{P}}}^\phi, \tilde{U}\right)$. This can be done by the following:

(i) The marginals are identifiable:

For any k and j, by definition, we have $u_j^{(k)} = P(W_j \le j)$. Based on the law of large numbers, $|\{W_j \le j\}|/n$ is a consistent estimator of $u_j^{(k)}$. So based on Proposition 2.3, $u_j^{(k)}$ is identifiable.

(ii) The pseudo-correlation matrix is identifiable:

Let $\mathbf{P}_{lk} = \begin{pmatrix} 1 & \rho_{lk} \\ \rho_{lk} & 1 \end{pmatrix}$. Define $g(\rho_{lk}) = C_{\mathbf{P}}^\phi(u_l, u_k)$. Similar to the proof of Theorem 2.5, we can show that $g(\rho_{lk})$ is strictly decreasing. □

We note by passing that the identifiability of the pseudo-correlation matrix also leads to a property of elliptical copulas:

Corollary 2.13. $\partial_{\rho_{lk}} C_{\mathbf{P}}^\phi(u_l, u_k) < 0$ *for all* $l = 1, \ldots, K; k = 1, \ldots, K; l \ne k$.

2.4. Identification of PCM with a latent multivariate t distribution

The following theorem establishes the identification of the polychoric *t* correlation model that Theorem 2.5 cannot cover. As aforementioned, the proof is based on the reparameterization of parameters by the copula.

Theorem 2.14. *Let* $\theta_{PCM,t} = (t_v(\mathbf{0},\mathbf{P}),\mathcal{A})$. *If the polychoric (pseudo-)correlation matrix* $\mathbf{P}$ *is positive definite, then the polychoric t correlation model* $(P_{\theta_{PCM,t}})$ *is just-identified.*

Proof. If two pairs of parameters correspond to the same response probability vector, say, $\left(C_{\mathbf{P}}^{\phi(v)}, U\right)$ and $\left(C_{\tilde{\mathbf{P}}}^{\phi(\tilde{v})}, \tilde{U}\right)$, we claim that $\left(C_{\mathbf{P}}^{\phi}, U\right) = \left(C_{\tilde{\mathbf{P}}}^{\phi}, \tilde{U}\right)$. Based on Lemma 2.12 and Corollary A.3, establishing $\left(C_{\mathbf{P}}^{\phi}, U\right) = \left(C_{\tilde{\mathbf{P}}}^{\phi}, \tilde{U}\right)$ suffices to complete the proof. Also, $C_{\mathbf{P}}^{\phi}$ can be uniquely determined by all bivariate copulas implied by $C_{\mathbf{P}}^{\phi}$, i.e., $\left\{C_{\rho_{lk},v}^{t}(u_l,u_k)\right\}_{l=1,\dots,K;k=1,\dots,K;l\neq k}$, so establishing $\left(C_{\rho_{12},v}^{t}, \dots, C_{\rho_{(K-1)K},v}^{t}, U\right) = \left(C_{\tilde{\rho}_{12},v}^{t}, \dots, C_{\tilde{\rho}_{(K-1)K},v}^{t}, \tilde{U}\right)$ suffices to complete the proof. $\square$

Our proof is composed of two parts:

(i)   The marginals are identifiable:

The proof is the same as the corresponding part of the "*Proof of Theorem 2.5 based on copula parametrization*" in the previous section.

(ii)   The bivariate copulas are identifiable:

Consider a response vector $\mathbf{w} = (r_1,\dots,r_{l-1},i,r_{l-1},\dots,r_{k-1},j,r_{k+1},\dots,r_p)$, for which the lth response is i, and the kth response is j. We claim that both $\partial_{\rho_{lk}} C_{\rho_{lk},v}^{t}(u_l,u_k) < 0$ and $\partial_v C_{\rho_{lk},v}^{t}(u_l,u_k) < 0$. For $P(\mathbf{W} \leq \mathbf{w}) = C_{\rho_{lk},v}^{t}\left(u_i^{(k)}, u_j^{(l)}\right)$, there is a unique copula $C_{\rho_{lk},v}^{t}$ satisfying the equation. Thus $C_{\rho_{lk},v}^{t}$ is identified, which completes this part of proof. The fact that $\partial_{\rho_{lk}} C_{\rho_{lk},v}^{t}(u_l,u_k) < 0$ has been established by Corollary 2.13. It remains to show that $\partial_v C_{\rho_{lk},v}^{t}(u_l,u_k) < 0$. Let $\tilde{v} > v$. Without loss of generality, we may assume that both $u_l = P(W_l \leq w_l) \leq 0.5$, and $u_k = P(W_k \leq w_k) \leq 0.5$. If u_l or $u_k > 0.5$, then the response can be reversely coded to achieve both u_l and $u_k \leq 0.5$. Because $\mathbf{P}$ is positive definite, $\mathbf{P}_{lk}$ is also positive definite. Thus, the Cholesky decomposition $\mathbf{P}_{lk} = \mathbf{\Lambda}\mathbf{\Lambda}^{\top}$ is unique and $\mathbf{\Lambda}$ is invertible. Based on Lemma A.4 and the property of the multivariate t distribution, we have $P(\mathbf{W} \leq \mathbf{w}) = P\left(\frac{\mathbf{\Lambda Z}}{\sqrt{V/v}} < \begin{pmatrix} a_j^{(l)} \\ a_i^{(k)} \end{pmatrix}\right)$, where $\mathbf{Z} \sim \mathcal{N}_2(\mathbf{0},\mathbf{I}_2)$ and $V \sim \chi_v^2$. Therefore,

$$
\begin{aligned}
P(\mathbf{W} \leq \mathbf{w}) &= P\left(\mathbf{Z} < \mathbf{\Lambda}^{-1}\begin{pmatrix} a_j^{(l)} \\ a_i^{(k)} \end{pmatrix}\sqrt{V/v}\right) \\
&= P\left(\mathbf{Z} < \begin{pmatrix} a_j^{(l)} \\ \frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}}a_j^{(l)} + \frac{1}{\sqrt{1-\rho_{lk}^2}}a_i^{(k)} \end{pmatrix}\sqrt{V/v}\right) \\
&= P\left(\mathbf{Z} < \begin{pmatrix} t_v^{-1}(u_l) \\ \frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}}t_v^{-1}(u_l) + \frac{1}{\sqrt{1-\rho_{lk}^2}}t_v^{-1}(u_k) \end{pmatrix}\sqrt{V/v}\right) \\
&= \int_{u=0}^{\infty} P\left(\mathbf{Z} < \begin{pmatrix} t_v^{-1}(u_l) \\ \frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}}t_v^{-1}(u_l) + \frac{1}{\sqrt{1-\rho_{lk}^2}}t_v^{-1}(u_k) \end{pmatrix}\sqrt{V/v} \Big| V = v\right) dF_V(v) \\
&= \int_{u=0}^{\infty} \Phi\left(t_v^{-1}(u_l)\sqrt{\frac{v}{v}}\right)\Phi\left(\left(\frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}}t_v^{-1}(u_l) + \frac{1}{\sqrt{1-\rho_{lk}^2}}t_v^{-1}(u_k)\right)\sqrt{\frac{v}{v}}\right) dF_V(v)\ (\mathbf{Z} \perp V) \\
&< \int_{u=0}^{\infty} \Phi\left(t_{\tilde{v}}^{-1}(u_l)\sqrt{\frac{v}{\tilde{v}}}\right)\Phi\left(\left(\frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}}t_{\tilde{v}}^{-1}(u_l) + \frac{1}{\sqrt{1-\rho_{lk}^2}}t_{\tilde{v}}^{-1}(u_k)\right)\sqrt{\frac{v}{\tilde{v}}}\right) dF_V(v)
\end{aligned}
$$

(Proposition A.5.)

$$= P\left(\mathbf{Z} < \left(\frac{t_{\tilde{v}}^{-1}(u_l)}{\frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}} t_{\tilde{v}}^{-1}(u_l) + \frac{1}{\sqrt{1-\rho_{lk}^2}} t_{\tilde{v}}^{-1}(u_k)} \right) \sqrt{V/\tilde{v}} \right)$$

$$\leq P\left(\mathbf{Z} < \left(\frac{t_{\tilde{v}}^{-1}(u_l)}{\frac{-\rho_{lk}}{\sqrt{1-\rho_{lk}^2}} t_{\tilde{v}}^{-1}(u_l) + \frac{1}{\sqrt{1-\rho_{lk}^2}} t_{\tilde{v}}^{-1}(u_k)} \right) \sqrt{V+V^*/\tilde{v}} \right) \left(V^* \sim \chi_{\tilde{v}-v}^2\right)$$

$$= P\left(\mathbf{Z} < \boldsymbol{\Lambda}^{-1} \begin{pmatrix} \tilde{a}_j^{(l)} \\ \tilde{a}_i^{(k)} \end{pmatrix} \sqrt{V+V^*/\tilde{v}} \right).$$

That is, $C_{\rho_{lk},v}^t(u_1, u_2) < C_{\rho_{lk},\tilde{v}}^t(u_1, u_2)$, and so $\partial_v C_{\rho_{lk},v}^t(u_1, u_2) < 0$. Because l and k are not specific, we have that $\partial_v C_{\rho_{lk},v}^t(u_l, u_k) < 0$ for any l and k.

We have proved that $\left(C_{\rho_{12},v}^t, \ldots, C_{\rho_{(K-1)K},v}^t, U\right) = \left(C_{\tilde{\rho}_{12},v}^t, \ldots, C_{\tilde{\rho}_{(K-1)K},v}^t, \tilde{U}\right)$ if two pairs of parameters correspond to the same response probability vector, and this leads to the identification of the polychoric t correlation model.

3. Identifiability and identifiability constraints of PM

In Section 2, we proved the identification of PCM. In comparison with PCM, the more general model PM is not identifiable, for which we now describe. We first present a proposition.

Proposition 3.1. *For a polychoric model* $\theta_{PM} = \left(EC_K(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi), \mathcal{A}\right)$*, suppose the transformation* $g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}$*, with* $B = diag\{b_1, \ldots, b_K\}, b_k > 0, \mathbf{c} \in \mathbb{R}^K$ *holds. Then* θ_{PM} *is empirically indistinguishable from* $\theta_{PM,g}$*, where* $\theta_{PM,g} = \left(EC_K(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}, \phi), \tilde{\mathcal{A}}\right)$ *and*

$$\begin{cases} \tilde{\boldsymbol{\mu}} = \mathbf{B}\boldsymbol{\mu} + \mathbf{c} \\ \tilde{\boldsymbol{\Sigma}} = \mathbf{B}^\top \boldsymbol{\Sigma} \mathbf{B} \\ \tilde{\mathcal{A}} = \mathcal{A}_{B,c} := \left\{ b_k a_j^{(k)} + c_k \right\}_{j=1,\ldots,r_k-1; k=1,\ldots,K} \end{cases}.$$

Proof. Let $\tilde{\mathbf{X}}^* = \mathbf{BX}^* + \mathbf{c}$. Then

$$P\left(\mathbf{W} \leq (j_1, \ldots, j_K)\right) = P\left(\mathbf{X}^* \leq \left(a_{j_1}^{(1)}, \ldots, a_{j_K}^{(K)}\right)^\top\right)$$

$$= P\left(\mathbf{BX}^* + \mathbf{c} \leq \mathbf{B}\left(a_{j_1}^{(1)}, \ldots, a_{j_K}^{(K)}\right)^\top + \mathbf{c}\right)$$

$$= P\left(\tilde{\mathbf{X}}^* \leq \left(b_1 a_{j_1}^{(1)} + c_1, \ldots, b_K a_{j_K}^{(K)} + c_K\right)^\top\right).$$

$\square$

Corollary 3.2. *PM is not identifiable because* θ_{PM} *and* $\theta_{PM,g}$ *are empirically indistinguishable, and both of them belong to* $P_{\theta_{PM}}$*.*

Thus, some workable identifiability constraints are needed to make PM identifiable. Indeed, the polychoric correlation model is constructed by putting some constraints on the location vector $\boldsymbol{\mu}$ and the dispersion matrix $\boldsymbol{\Sigma}$ of PM. However, these identifiability constraints are not practical in some situations. For instance, when modeling the developmental changes of children, it is not well grounded to assume that the location vector $\boldsymbol{\mu}$ is zero (McArdle et al., 2015; Muthén, 1984). Therefore, finding some other reasonable identifiability constraints is a task with practical significance.

3.1. The equivalence-classes approach of identifiability (ECAI)

An effective way to find the identifiability constraints was proposed by Tsai (2000, 2003) in the context of Thurstonian modeling of comparative judgment. Because this approach depends on the mathematical

concept *of equivalence classes* (Dummit & Foote, 2004), we call it the *Equivalence-Classes Approach of Identifiability* (ECAI).

Traditionally, the identifiability of models needs to be proved case by case. ECAI provides a different route for derivation. Importantly, Tsai (2003) applied ECAI and proved that Case III and Case V of the Thurstonian models of comparative judgment are identifiable.

ECAI can help researchers to find a model's identifiability constraints. Furthermore, it can be utilized to determine whether a (restricted) model is just-identified or over-identified. It utilizes the following two principles to construct the identifiability of the restricted model:

(i) For a non-identified general model, find the set of all parameters empirically indistinguishable from the true parameter, i.e., the identified set.

(ii) After adding some constraints to the model, if there is at most one element in the predefined set, then the restricted model is identified, and the constraints are identifiability constraints. Moreover, if there is one and only one element in the predefined set, then the restricted model is just-identified; otherwise, the model is over-identified.

We can formulate ECAI as follows.

Proposition 3.3. *Under Definition 2.1,*

(i) $P_{\theta_1} = P_{\theta_2}$ *defines an equivalence relation on* θ, *denoted by* $\theta_1 \sim \theta_2$.

(ii) *The identified sets are the equivalence classes defined by the relation. Therefore, the identified set of* P_{θ_0} *can be denoted as* $[\theta]$ *for all* $\theta \in \Theta$ *satisfying* $P_\theta = P_{\theta_0}.\theta$ *is called a representative of the equivalence class. That is,* $\{\theta | P_\theta = P_{\theta_0}\} = [\theta_0] = [\theta]$ *for* $\theta \sim \theta_0$.

(iii) *The set of equivalence classes forms a partition of* Θ, *that is, for any* $\theta_0 \in \Theta, \theta_0$ *belongs to one and only one of the equivalence classes.*

(iv) $\mathcal{P}$ *is (just-)identifiable if and only if for each identified set* $[\theta_0]$, *there is a unique element,* θ_0, *in it, i.e.,* $[\theta_0] = \{\theta_0\}$. *Namely,* $[\theta_0]$ *is a singleton.*

(v) $\mathcal{P}$ *is partially identifiable over* θ^* *if and only if for each identified set* $[\theta_0]$, *there is a unique element in* $g[\theta_0] = \{\theta^* | \theta^* = g(\theta), \theta \sim \theta_0\}$.

Proof. (i) is straightforward by the definition of the equality of functions. (iii) follows by Proposition 2 in 0.1 of Dummit and Foote (2004). (iv) and (v) can be proved by (iii), Proposition 2.2 and the definition of identifiability. □

Proposition 3.3(iv) shows the sufficient and necessary condition for model identification. A statistical model might not be identifiable unless one adds some identifiability constraints to it. As mentioned above, Thurstonian models are identifiable with the Case III or Case V assumptions (Thurstone, 1927; Tsai, 2000). Here is another example. The confirmatory factor analysis model assumes that either the variance of the latent factor is any positive constant, or the loading of an indicator is any nonzero constant; otherwise, the model is not identifiable (Kline, 2023).

Definition 3.4. (Identifiability constraints) Under Definition 2.1, let $\psi_1, \ldots, \psi_J$ be functions of θ. Then $\mathcal{P}$ is identifiable with constraints $\psi_1, \ldots, \psi_J$ if the model with the parameter space under constraints, $\Theta_c = \Theta \cap \{\theta | P_\theta = P_{\theta_0}, \psi_1(\theta) = 0, \ldots, \psi_J(\theta) = 0\}$, is identifiable. We call $\psi_1, \ldots, \psi_J$ the identifiability constraints if $\mathcal{P}$ is identifiable with these constraints but not identifiable without these constraints.

Proposition 3.5. *Let* $[\theta_0]_{\psi_1, \ldots, \psi_J} = \{\theta \in \Theta | P_\theta = P_{\theta_0}, \psi_1(\theta) = 0, \ldots, \psi_J(\theta) = 0\}$ *and* Θ_c *is not empty. The following three statements are equivalent:*

(i) $\mathcal{P}$ *is identifiable with constraints* $\psi_1, \ldots, \psi_J$

(ii) *For any* $\theta_0^* \in \Theta_c, [\theta_0^*]_{\psi_1, \ldots, \psi_J} = \{\theta_0^*\}$.

(iii) *For any* $\theta_0 \in \Theta, [\theta_0]_{\psi_1, \ldots, \psi_J}$ *has at most one element.*

Proof. (ii) can be easily derived from Proposition 3.3(iv). (iii) implies (ii) naturally. (ii) ⇒(iii) can be proved through contraposition. If there is a $\theta_0 \in \Theta$ such that $[\theta_0]_{\psi_1,\dots,\psi_J}$ has more than one elements. Because $[\theta_0]_{\psi_1,\dots,\psi_J} \subseteq \Theta_c$, then (ii) is violated. Thus (ii) ⇒(iii). □

The distinction between just-identified and over-identified (restricted) models can be recharacterized below using the concept of identifiability constraints.

Definition 3.6. Under Definitions 2.1 and 3.4,

  (i) $\mathcal{P}$ is just-identified with constraints $\psi_1,\dots,\psi_J$ if for any $\theta_0 \in \Theta, [\theta_0]_{\psi_1,\dots,\psi_J}$ has a unique element. Moreover, $\psi_1,\dots,\psi_J$ are called *minimal constraints* if removing any of it leads to non-identification.
  (ii) $\mathcal{P}$ is over-identified with constraints $\psi_1,\dots,\psi_J$ if there exists $\theta_0 \in \Theta$, such that $[\theta_0]_{\psi_1,\dots,\psi_J}$ is empty.

We take the one-way ANOVA model (Rice, 2007) as an example to demonstrate the relation of non-identified, just-identified, and over-identified models. Let K be the number of groups. In this model, the parameters $(\mu, \alpha_1, \dots, \alpha_K)^\top \in \mathbb{R}^{K+1}$ are not identifiable because both $(\mu, \alpha_1, \dots, \alpha_K)^\top$ and $(\mu+1, \alpha_1-1, \dots, \alpha_K-1)^\top$ belong to $[(\mu, \alpha_1, \dots, \alpha_K)^\top]$. $\sum_k \alpha_k = 0$ is an identifiability constraint because $[(\mu, \alpha_1, \dots, \alpha_K)^\top]_{\sum_k \alpha_k = 0} = \{(\mu, \alpha_1, \dots, \alpha_K)^\top\}$. Also, "$\sum_k \alpha_k = 0$ and $\alpha_1 = 0$" are identification constraints because a subset of the constraints ($\sum_k \alpha_k = 0$) can identify the model. Moreover, $\sum_k \alpha_k = 0$ is a minimal constraint because for any $(\mu, \alpha_1, \dots, \alpha_K)^\top \in \mathbb{R}^{K+1}$, there exists a unique element $(\mu, \alpha_1 - \sum_k \alpha_k, \dots, \alpha_K - \sum_k \alpha_k)^\top$ belonging to $[(\mu, \alpha_1, \dots, \alpha_K)^\top]$ that removing $\sum_k \alpha_k = 0$ leads to non-identification. However, "$\sum_k \alpha_k = 0$ and $\alpha_1 = 0$" produce an over-identified model because there is a vector $(\mu, 0, 1, \dots, 1)^\top$ such that $[(\mu, 0, 1, \dots, 1)^\top]_{\sum_k \alpha_k = 0, \alpha_1 = 0} = \phi$. That is, there is no parameter which is empirically equivalent to $(\mu, 0, \dots, 1)^\top$ and simultaneously satisfies both $\sum_k \alpha_k = 0$ and $\alpha_1 = 0$.

Another characterization of just-identification is the existence of a one-to-one correspondence between the identified parameters and the parameters with identification constraints (see Chang et al., 2017). It turns out this characterization is equivalent to the definition of just-identification:

Proposition 3.7. *Under Definitions 2.1, 3.4, and 3.6., let the model with $\psi_1,\dots,\psi_J$ constraints be just-identified with the parameter space under constraints Θ_c. Then the model with $\phi_1,\dots,\phi_L$ constraints is just-identified with the parameter space under constraints $\Theta_{\tilde{c}}$ if and only if there is a one-to-one correspondence (of empirically indistinguishable models) between Θ_c and $\Theta_{\tilde{c}}$.*

Proof. (⊇) Because the model with $\psi_1,\dots,\psi_J$ constraints is just-identified, for any $\theta_0 \in \Theta, [\theta_0]_{\psi_1,\dots,\psi_J}$ has a unique element. Then $[\theta_0]_{\phi_1,\dots,\phi_L}$ is also not empty based on the one-to-one correspondence. It indicates that the model with $\phi_1,\dots,\phi_L$ constraints is just-identified.

(⊆) Based on just-identification of the two models, both $[\theta_0]_{\psi_1,\dots,\psi_J}$ and $[\theta_0]_{\phi_1,\dots,\phi_L}$ are singletons mapped by θ_0. Thus, we can construct the correspondence between Θ_c and $\Theta_{\tilde{c}}$ based on these mappings. This completes the proof. □

The following theorem links ECAI to identifiability constraints.

Definition 3.8. $\mathcal{P}_{A,\theta} = \{P_\theta : \theta \in \Theta_A\}$ and $\mathcal{P}_{B,\theta} = \{P_\theta : \theta \in \Theta_B\}$ are two models where $\Theta_B \subseteq \Theta_A$. Then $\mathcal{P}_{A,\theta}$ is a general model of $\mathcal{P}_{B,\theta}$ and $\mathcal{P}_{B,\theta}$ is a restricted model of $\mathcal{P}_{A,\theta}$, denoted by $\mathcal{P}_{B,\theta} \subseteq \mathcal{P}_{A,\theta}$.

Theorem 3.9. *Let $\psi_1,\dots,\psi_J$ be functions of θ. $\Theta_B = \{\theta \in \Theta_A | \psi_1(\theta) = 0, \dots, \psi_J(\theta) = 0\}$. Thus $\mathcal{P}_{B,\theta} \subseteq \mathcal{P}_{A,\theta}$. Then*

  (i) *If $\mathcal{P}_{A,\theta}$ is identifiable, then $\mathcal{P}_{B,\theta}$ is identifiable.*
  (ii) *If $\mathcal{P}_{B,\theta}$ is identifiable, then $\psi_1,\dots,\psi_J$ are identifiability constraints of $\mathcal{P}_{A,\theta}$. In particular, If $\mathcal{P}_{B,\theta}$ is just-identified and removing any of $\psi_1,\dots,\psi_J$ leads to non-identification., then $\psi_1,\dots,\psi_J$ are minimal constraints.*

(iii) *(ECAI) $\mathcal{P}_{A,\theta}$ is not identifiable. $\theta_0^* \in \Theta_B$. Under $\mathcal{P}_{A,\theta}$, if $\left[\theta_0^*\right]_{\psi_1,\dots,\psi_J}$ is a singleton or empty, then $\mathcal{P}_{B,\theta}$ is identifiable. Moreover, $\psi_1,\dots,\psi_J$ are identifiability constraints of $\mathcal{P}_{A,\theta}$. Further, if $\left[\theta_0\right]_{\psi_1,\dots,\psi_J}$ is a singleton for any $\theta_0 \in \Theta_A$, then $\mathcal{P}_{B,\theta}$ is just-identified. Otherwise $\mathcal{P}_{B,\theta}$ is over-identified.*

Proof. The results are immediate consequences of Proposition **3.3**(iv) and Definition 3.6. □

Another related issue is that the equivalent class of a non-identifiable model consists of all permissible transformations of a model, which can determine the scale of measurement (Luce et al., 1990, Chapter 20; Stevens, 1946). The scale of a measurement can be identified as follows.

Definition 3.10. *Let $\mathcal{P}_\theta = \{P_\theta : \theta \in \Theta\}$ be a model and X be a random variable $X \sim P_\theta$. Define $G = \left\{g : g(X) \sim P_{\theta^*}, \exists \theta^* \in \Theta\right\}.$*

(i) *X is a nominal scale if G consists of all injective functions.*
(ii) *X is an ordinal scale if G consists of all monotone increasing functions.*
(iii) *X is an interval scale if $G = \{g(x) : g(x) = bx + c, b > 0\}$.*
(iv) *X is a ratio scale if $G = \{g(x) : g(x) = bx, b > 0\}$.*
(v) *X is an absolute scale if $G = \{id\}$.*

3.2. Identified sets of PM

Theorem 3.11 finds the equivalent class for PM. The theorem states that all empirically equivalent PMs can be linearly transformed into each other. So, the equivalent class of PM is the set of all possible linearly transformed PMs of it, and the latent vectors corresponding to PM are interval scales.

Theorem 3.11. *Given the conditions in Theorem 2.5, consider PM $\theta_{PM} = (EC_K(\boldsymbol{\mu},\boldsymbol{\Sigma},\phi),\mathcal{A})$. Also,*

$$[\theta_{PM}] = \left\{\theta_{PM,g} : g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}, \mathbf{B} = diag(b_1,\dots,b_K), b_k > 0, \mathbf{c} \in \mathbb{R}^K\right\}. \quad (2)$$

Then all entries of $\mathbf{X}^* = \left(X_1^*,\dots,X_K^*\right)^\top$ are interval scales, and the equivalent class of PM consists of models $\theta_{PM,EC,g} = \left(EC_K(\tilde{\boldsymbol{\mu}},\tilde{\boldsymbol{\Sigma}},\phi),\tilde{\mathcal{A}}\right)$ satisfying

$$\begin{cases} \tilde{\boldsymbol{\mu}} = \mathbf{B}\boldsymbol{\mu} + \mathbf{c} \\ \tilde{\boldsymbol{\Sigma}} = \mathbf{B}^\top \boldsymbol{\Sigma} \mathbf{B} \\ \tilde{\mathcal{A}} = \mathcal{A}_{B,c} := \left\{b_k a_j^{(k)} + c_k\right\}_{j=1,\dots,r_k-1;k=1,\dots,K} \end{cases}.$$

Proof. ($\Leftarrow$) is straightforward by Proposition 3.1.

($\Rightarrow$) Consider two PMs, $\theta_{PM} = (EC_K(\boldsymbol{\mu},\boldsymbol{\Sigma},\phi),\mathcal{A})$ and $\theta^*_{PM} = \left(EC_K(\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*,\phi),\mathcal{A}^*\right)$, where $\theta_{PM} \sim \theta^*_{PM}$. We claim that

$$\theta^*_{PM} \in \left\{\theta_{PM,g} : g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}, \mathbf{B} = diag(b_1,\dots,b_K), b_k > 0, \mathbf{c} \in \mathbb{R}^K\right\},$$

that is, there is a function $\bar{g}$ such that $\theta_{PM,\bar{g}} = \theta^*_{PM}$, where $\bar{g}(\mathbf{x}) = \overline{\mathbf{B}}\mathbf{x} + \overline{\mathbf{c}}$, Define $g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}$ where $B = \text{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}}$ and $\mathbf{c} = -\text{diag}(\boldsymbol{\Sigma})^{-\frac{1}{2}}\boldsymbol{\mu}$, and $g^*(\mathbf{x}) = \mathbf{B}^*\mathbf{x} + \mathbf{c}^*$ where $\mathbf{B}^* = \text{diag}(\boldsymbol{\Sigma}^*)^{-\frac{1}{2}}$ and $\mathbf{c}^* = -\text{diag}(\boldsymbol{\Sigma}^*)^{-\frac{1}{2}}\boldsymbol{\mu}^*$. Based on Proposition 3.1, we have $\theta_{PM,g} \sim \theta_{PM} \sim \theta^*_{PM} \sim \theta^*_{PM,g^*}$. Both $\theta_{PM,g}$ and $\theta^*_{PM,\bar{g}}$ are PCM. Because of the identification of PCM (i.e., Theorem 2.5 and Theorem 2.14), there cannot be two different PCMs that are empirically indistinguishable. Therefore $\theta_{PM,g} = \theta^*_{PM,g^*}$, and so $\theta_{PM,(g^*)^{-1}\circ g} = \theta^*_{PM}$, i.e., $\bar{g} = \left(g^*\right)^{-1} \circ g$. □

By applying Proposition 3.7, Theorem **3.9**, and Theorem 3.11, we show in the following that the constraints of θ_{PCM} are minimal constraints. We then construct the criteria for the minimal constraints of the identifiability constraint in PM.

Corollary 3.12. *For PM the constraints that $\boldsymbol{\mu} = \mathbf{c}$ and $diag(\boldsymbol{\Sigma}) = \mathbf{b}$ (where $\mathbf{b} \in \mathbb{R}_+^K$ with $\mathbf{c} \in \mathbb{R}^K$,) are the minimal constraints. In particular, for PCM, the constraints that $\boldsymbol{\mu} = \mathbf{0}$ and $diag(\boldsymbol{\Sigma}) = \mathbf{1_K}$ are the minimal constraints.*

Proof. Equation (2) shows that $[\theta_{PM}]$ will not be a singleton when any of the two constraints $\boldsymbol{\mu} = \mathbf{c}$ and $diag(\boldsymbol{\Sigma}) = \mathbf{b}$ is removed. Therefore $\boldsymbol{\mu} = \mathbf{c}$ and $diag(\boldsymbol{\Sigma}) = \mathbf{b}$ consist of the minimal constraints. The particular scenario concerning the minimal constraints of PCM is also supported by the application of Theorem 3.9(ii). □

Corollary 3.13. *Under PM, let $\theta_{PM} = (EC_K(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi), \mathcal{A})$. The following three statements are equivalent:*

(i) *$\psi_1, \ldots, \psi_J$ are the minimal constraints.*
(ii) *There is a bijection from $\{\theta_{PCM}\}$ to $\{\theta_{PM} | \psi_1(\theta_{PM}) = \cdots = \psi_J(\theta_{PM}) = 0\}$, and removing any of $\psi_1, \ldots, \psi_J$ leads to non-identification.*
(iii) *$\psi_1, \ldots, \psi_J$ are identifiable constraints, there is a surjection (of empirically indistinguishable models) from $\{\theta_{PCM}\}$ to $\{\theta_{PM} | \psi_1(\theta_{PM}) = \cdots = \psi_J(\theta_{PM}) = 0\}$, and removing any of $\psi_1, \ldots, \psi_J$ leads to non-identification.*

Proof. $(i) \iff (ii)$ By Proposition 3.7 and Corollary 3.12, $\psi_1, \ldots, \psi_J$ are the minimal constraints if and only if there is an injection from $\{\theta_{PCM}\}$ to $\{\theta_{PM} | \psi_1(\theta_{PM}) = \cdots = \psi_J(\theta_{PM}) = 0\}$.
$(ii) \iff (iii)$ The identification can guarantee the injection. The reverse is also true. □

In the following, we demonstrate two applications of the ECAI approach along with our theorems concerning the identified sets of PM. First, we will use ECAI to find the identifiability constraints of PM on Likert scale (LS) and on comparative judgment (CJ) items. Second, using the ordinal SEM and item factor analysis as an example, we will illustrate the use of ECAI in establishing the identifiability of (restricted) models of PM.

3.3. Application 1: Identifiability constraints of PM on Likert scales and on comparative judgment

As mentioned earlier, PM subsumes several commonly used psychometric models in analyzing LS and CJ items. For LS, the graded response model (Samejima, 1968; 1997) and item factor analysis models (Wirth & Edwards, 2007) are restricted PMs. For CJ, Thurstonian models (Thurstone, 1927) and their advances (Bockenholt & Tsai, 2001; Brown & Maydeu-Olivares, 2013) are also restricted models of PM. The following theorem shows some identifiability constraints of PM on LS and CJ items. For convenience, CJ with r-point ordinal preference responses is called *r-point CJ* (Agresti, 1992; Brown & Maydeu-Olivares, 2018), whereas LS with r-point ordinal responses is called *r-point LS*.

Theorem 3.14.

(i) *Consider K items in a Q-point LS (Q > 2). If the first cut-offs of items are set to zero, and the final cut-offs are set to one ($a_1^{(k)} = 0, a_{Q-1}^{(k)} = 1$), then these constraints are minimal, and the PM is just-identified. Moreover, all entries of $\mathbf{X}^* = (X_1^*, \ldots, X_K^*)^\top$ are absolute scales. We call the above constraints the "global scale constraints (GSC)."*

(ii) *Consider K items in a (2H)-point LS/CJ (H ≥ 1). If the middle cut-offs of LS/CJ items are set to zero ($a_{H+1}^{(k)} = 0$), then $[\theta_{PM}] = \{\theta_{PM,g} : g(\mathbf{x}) = \mathbf{Bx}, \mathbf{B} = diag(b_1, \ldots, b_K), b_k > 0\}$. So all entries of $\mathbf{X}^* = (X_1^*, \ldots, X_K^*)^\top$ are ratio scales. Moreover, if $diag(\boldsymbol{\Sigma}) = \mathbf{1_K}$, namely, $\boldsymbol{\Sigma} = \mathbf{P}$, then these constraints are minimal and the PM is just-identified. We call $diag(\boldsymbol{\Sigma}) = \mathbf{1_K}$ the "unit variance constraints (UVC)" because it assumes that the variances of scales in X_1^* to X_k^* are one.*

(iii) *Consider K items in a (2H + 1)-point LS/CJ items (H ≥ 1). If the middle cut-offs of LS/CJ items are symmetric around 0 ($a_H^{(k)} + a_{H+1}^{(k)} = 0$), then $[\theta_{PM}] = \{\theta_{PM,g} : g(\mathbf{x}) = \mathbf{Bx}, \mathbf{B} = diag(b_1, \ldots, b_K), b_k > 0\}$. So all entries of $\mathbf{X}^* = (X_1^*, \ldots, X_K^*)^\top$ are ratio scales. Moreover, if $diag(\boldsymbol{\Sigma}) = \mathbf{1_K}$, namely, $\boldsymbol{\Sigma} = \mathbf{P}$, then these constraints are minimal and the PM is just-identified.*

(iv) *Consider K items in a Q-point LS/CJ (Q > 2). If the cut-offs of extreme preferences are set to be − 1 and 1 (i.e., $a_1^{(k)} = -1, a_{Q-1}^{(k)} = 1$), then these constraints are minimal and the PM is just-identified. Moreover, all entries of $\mathbf{X}^* = \left(X_1{}^*, \ldots, X_K{}^* \right)^\top$ are absolute scales. We call the above constraints the "extremity constraints (EC)."*

Proof. Because $[\theta_{PM}] = \left\{ \theta_{PM,g} : g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}, \mathbf{B} = diag(b_1, \ldots, b_K), b_k > 0, \mathbf{c} \in \mathbb{R}^K \right\}$, we only need to consider whether $b_k = 1$ and $c_k = 0$ for all k. If both hold, then any PM-based model with these constraints is identifiable. Moreover, by Corollary 3.13, if there is a surjection from PCM to PM with the aforementioned constraints, then the PM is just-identified.

(i) For any k, $a_1^{(k)} = 0, a_{Q-1}^{(k)} = 1$, so

$$\begin{cases} b_k a_1^{(k)} + c_k = c_k = 0 \\ b_k a_{Q-1}^{(k)} + c_k = b_k + c_k = 1 \end{cases} \Rightarrow \begin{cases} b_k = 1 \\ c_k = 0. \end{cases}$$

Therefore $g(x) = x$. For any PCM, we can construct a surjection by rescaling $a_1^{(k)}$ to 0 and $a_{Q-1}^{(k)}$ to 1 for all items. Therefore, the model is just-identified. These constraints are minimal because they are linearly independent.

(ii) For any k, $a_{H+1}^{(k)} = 0$, and $b_k a_{H+1}^{(k)} + c_k = 0$. So $c_k = 0$ for all k. We now prove the identification under the UVC. Let $\theta_{PM,EC,g}$ be empirically indistinguishable from $\theta_{PM,EC}$ for some $g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}$. Because $\sigma_k^2 = b_k^2 \tilde{\sigma}_k^2$ and $diag(\boldsymbol{\Sigma}) = diag(\tilde{\boldsymbol{\Sigma}}) = \mathbf{1_K}$, we have $\mathbf{B} = \mathbf{I}_k$. For any PCM, we can construct a surjection by shifting $a_{H+1}^{(k)}$ to 0 for all items. Therefore, the model is just-identified. These constraints are also minimal because they are linearly independent. (iii) can be proved in a similar way.

(iv) For any k, $a_1^{(k)} = -1, a_{Q-1}^{(k)} = 1$ so

$$\begin{cases} b_k a_1^{(k)} + c_k = -b_k + c_k = -1 \\ b_k a_{Q-1}^{(k)} + c_k = b_k + c_k = 1 \end{cases} \Rightarrow \begin{cases} b_k = 1 \\ c_k = 0. \end{cases}$$

Therefore $g(\mathbf{x}) = \mathbf{x}$. For any PCM, we can construct a surjection by rescaling $a_1^{(k)}$ to −1 and $a_{Q-1}^{(k)}$ to 1 for all items. Therefore, the model is just-identified. □

Theorem 3.14(i) and (iv) show that PM is identifiable with GSC or EC for LS items. Indeed, GSC or EC would be appropriate for LS with different context labels. For instance, Casper et al. (2020) enumerated several types of labels, such as agreement, similarity, and frequency. Within this context, EC might be better than GSC for the agreement anchor because agreement and disagreement are two extreme attitudes. For example, for a 7-point LS from 1 (strongly disagree) to 7 (strongly agree), we could set the bound of responding "strongly disagree" to −1 and "strongly agree" to 1. In contrast, for the similarity and frequency anchor, GSC might be better than EC, because GSC is better at representing the context of similarity or frequency. For example, for a 7-point LS from 1 (not at all like me) to 7 (extremely like me), we could set the cut-offs of "not at all like me" to 0 and "extremely like me" to 1.

Under GSC, because the zero and unit are prespecified, the scale is an absolute scale, i.e., a scale with an absolute zero and absolute unit (Zwislocki & Goodman, 1980). The absolute scale is a type of measurement scale that extends Stevens' (1946) classification (Luce et al., 1990). An absolute scale might be helpful in practice because all calculations are permissible for the absolute scale.

Moreover, Theorem 3.14(ii) and (iii) show that PM is identifiable with UVC or EC for CJ items. Under the Thurstonian modeling framework, $\boldsymbol{\mu}$ is the differences of the means of latent processes. In order to overcome the identifiability problem, researchers have made assumptions such as Case III or Case V to identify the model parameters (Thurstone, 1927; Tsai, 2000). Here we demonstrate that other constraints, such as UVC and EC, can help researchers to identify $\boldsymbol{\mu}$. (Note that the latent scales under PCM and UVC share a common unit but have different origins. The relation between these two is similar to a standardized scale $z = (x - \mu)/\sigma$ and a scale normalized by the standard deviation: $y = x/\sigma$.)

We mentioned earlier that Case V implies the covariance matrix underlying CJ items $\boldsymbol{\Sigma} = c\mathbf{I}$ for some c, where $\boldsymbol{\Sigma}$ is the covariance matrix of latent differences (Thurstone, 1927; Tsai, 2000). Let the variance of the discriminal process be 1/2, then $\boldsymbol{\Sigma} = \mathbf{I}$, indicating that UVC ($diag(\boldsymbol{\Sigma}) = \mathbf{1}_K$) is a general model of Case V. The identification of UVC implies that the independence assumption in Case V is unnecessary for identifying the covariance matrix underlying CJ items. Assuming UVC rather than Case V can avoid the risk of making wrong uncorrelated assumptions. In practice, UVC can be applied to multidimensional scaling (MDS). Recall that MDS utilizes $\boldsymbol{\mu}$ to infer multidimensional relative positions among different alternatives, for which Case V is typically assumed (Torgerson, 1952). UVC can identify $\boldsymbol{\mu}$ without assuming that the latent variables are uncorrelated.

3.4. *Application 2: Identifiability of restricted models of PM*

PM subsumes several commonly used psychometric models. We can use this hierarchy to prove the identifiability of some commonly used psychometric models. We first prove a lemma.

Lemma 3.15. *Let $\theta(\omega) = EC_K(\boldsymbol{\mu}(\omega), \boldsymbol{\Sigma}(\omega), \phi)$ be an elliptical distribution parametrized by a real vector $\omega \in \Omega$, then the empirically indistinguishable relation defines the family of equivalence classes $[\theta(\omega)] = \{\theta(\omega^*) : \boldsymbol{\mu}(\omega) = \boldsymbol{\mu}(\omega^*), \boldsymbol{\Sigma}(\omega) = \boldsymbol{\Sigma}(\omega^*), \omega^* \in \Omega\}$. Consequently, the equivalence class of the PM with parameter $\theta_{PM}(\omega) = (EC_K(\boldsymbol{\mu}(\omega), \boldsymbol{\Sigma}(\omega), \phi), \mathcal{A})$ is*

$$[\theta_{PM}(\omega)] = \left\{\theta_{PM,g}(\omega^*) : g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}, \mathbf{B} = diag(b_1, \ldots, b_K), b_k > 0, \mathbf{c} \in \mathbb{R}^K, \omega^* \in \theta^{-1}[\theta(\omega)]\right\}. \quad (3)$$

Proof. Based on Theorem 3.11, $[\theta_{PM}(\omega)] = \bigcup_{\omega^* \in \theta^{-1}[\theta(\omega)]} \{\theta_{PM,g}(\omega^*) : g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}, \mathbf{B} = diag(b_1, \ldots, b_K),$

$b_k > 0, \mathbf{c} \in \mathbb{R}^K\}$. The union of equivalence relations satisfies reflexivity and symmetry natually. Also, the equivalence of probability structure satisfies transitivity. Thus, the union froms another equivalence class. $\qquad \square$

Based on the above lemma, we can easily derive that If $\theta(\omega) = EC_K(\boldsymbol{\mu}(\omega), \boldsymbol{\Sigma}(\omega), \phi)$ is just-identified/identifiable, then PM with parameter $\theta_{PM}(\omega) = (EC_K(\boldsymbol{\mu}(\omega), \boldsymbol{\Sigma}(\omega), \phi), \mathcal{A})$ is an interval scale. An application is stated in the following theorem, which presents necessary and sufficient conditions for the identifiability of the ordinal SEM. Notably, the item factor analysis (FA) model, also named as the confirmatory multidimensional item response theory (IRT), can be granted as a special case of ordinal SEM (Asparouhov & Muthén, 2020; Reckase, 2009; SAS Institute Inc., 2024; Takane & de Leeuw, 1987; Wirth & Edwards, 2007).

Corollary 3.16. *An SEM model, $\theta(\omega) = EC_K(\boldsymbol{\mu}(\omega), \boldsymbol{\Sigma}(\omega), \phi)$, is identifiable if and only if the corresponding ordinal SEM model is an interval scale.*

Proof. ($\Rightarrow$) Based on the identification of the SEM model, there is only one element, ω, in $\theta^{-1}[\theta(\omega)]$. Thus Equation (3) reduces to $[\theta_{PM}(\omega)] = \{\theta_{PM,g}(\omega^*) : g(\mathbf{x}) = \mathbf{Bx} + \mathbf{c}, \mathbf{B} = diag(b_1, \ldots, b_K), b_k > 0, \mathbf{c} \in \mathbb{R}^K\}$, which is the definition of the interval scale. Reversely, if the ordinal SEM model is not an interval scale, then $\theta^{-1}[\theta(\omega)]$ has more than one elelments. Thus, the SEM model is not identifiable. $\qquad \square$

Theorem 3.17. *(Sufficient and necessary condition for (just-)identifiability of ordinal SEM) The ordinal SEM model, $\theta_{PM}(\omega) = (EC_K(\boldsymbol{\mu}(\omega), \boldsymbol{\Sigma}(\omega), \phi), \mathcal{A})$, with constraints $\psi_1, \ldots, \psi_J$, is identifiable if and only if the SEM model is identifiable and $[\theta_{PM}(\omega)]_{\psi_1, \ldots, \psi_J}$ admits both $\mathbf{B} = \mathbf{I}_k$ and $\mathbf{c} = 0$ in Equation (3). Moreover,*

it is just-identifiable if and only if the SEM model is just-identifiable and for any ω such that $\theta(\omega) = EC_K(\mathbf{0}, \mathbf{P}, \phi)$ there is one and only one corresponding $[\theta_{PM}(\omega)]_{\psi_1,\ldots,\psi_J}$. In addition, $\psi_1,\ldots,\psi_J$ are minimal constraints if removing any of $\psi_1,\ldots,\psi_J$ leads to non-identification.

Proof. ($\Rightarrow$) Using an argument similar to that of Corollary 3.16, the ordinal SEM model with constraints admits a partition finer than that of an interval scale, and so the SEM model is identifiable. Moreover, we have

$$[\theta_{PM}(\omega)] = \left\{ \theta_{PM,g}(\omega) : g(\mathbf{x}) = \mathbf{B}\mathbf{x} + \mathbf{c}, \mathbf{B} = diag(b_1,\ldots,b_K), b_k > 0, \mathbf{c} \in \mathbb{R}^K \right\}.$$

Now that the ordinal SEM model is identifiable. It implies $g(\mathbf{x}) = \mathbf{x}$, which is equivalent to $\mathbf{B} = \mathbf{I}_k$ and $\mathbf{c} = \mathbf{0}$.

($\Leftarrow$) Because the SEM model is identifiable, and $\psi_1,\ldots,\psi_J$ admits both $\mathbf{B} = \mathbf{I}_k$ and $\mathbf{c} = \mathbf{0}$, the equivalence class $[\theta_{PM}(\omega)]_{\psi_1,\ldots,\psi_J}$ will not have more than one element. Therefore, the ordinal SEM model is identified.

The remaining parts can be easily proved through Corollary 3.13.   □

Corollary 3.18. *(Sufficient and necessary condition for (just-)identifiability of item factor analysis) An item factor analysis model (setting all of the error variances to be one) is identifiable/just-identifiable if and only if the corresponding factor analysis model (through correlation matrix) is identifiable/just-identifiable.*

Proof. Consider an m-dimensional item factor analysis with discriminant parameter vectors $\mathbf{a}^{(k)} \in \mathbb{R}^m$ and threshold vectors $\mathbf{b}^{(k)} \in \mathbb{R}^m$ for item k, $k = 1,\ldots,K$. Based on the formulas (6) and (13) in Takane and de Leeuw (1987), the model can be expressed as

$$\theta_{PM}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}, \mathbf{b}^{(1)},\ldots,\mathbf{b}^{(K)}\right) = \left(EC_K\left(\mathbf{0}, \mathbf{\Sigma}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right), \phi\right), \mathcal{A}\left(\mathbf{b}^{(1)},\ldots,\mathbf{b}^{(K)}\right)\right), \quad (4)$$

where $\mathcal{A}\left(\mathbf{b}^{(1)},\ldots,\mathbf{b}^{(K)}\right) = \left\{-\mathbf{b}^{(k)}\right\}_{k=1,\ldots,K}$ and $\mathbf{\Sigma}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right) = \left(\mathbf{a}^{(k)\top}\mathbf{a}^{(j)}\right) + \mathbf{I}_K$. We now check the conditions in Theorem 3.17. Because there is a one–one correspondence (through standardization and its inverse transformation) between $\mathbf{\Sigma}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right) = \left(\mathbf{a}^{(k)\top}\mathbf{a}^{(j)}\right) + \mathbf{I}_K$ and the FA model through correlation matrix, $\mathbf{P}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right) = \left(\mathbf{a}^{(k)\top}\mathbf{a}^{(j)}\right) + diag(q_1, q_2,\ldots,q_K)$, we can reparametrize the model as a PCM. Thus, it is identifiable/just-identifiable when the corresponding factor analysis model (through correlation matrix) is identifiable/just-identifiable.   □

Remark 3.19. Based on the formulas (6) and (13) in Takane and de Leeuw (1987), the model can be parameterized as an IRT or as an FA. The IRT parametrization corresponds to Equation (4), and the FA parametrization corresponds to the following equation

$$\theta_{PM}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}, \mathbf{b}^{(1)},\ldots,\mathbf{b}^{(K)}\right) = \left(EC_K\left(\mathbf{0}, \mathbf{\Sigma}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right), \phi\right), \mathcal{A}\left(\mathbf{b}^{(1)},\ldots,\mathbf{b}^{(K)}\right)\right),$$

where $\mathcal{A}\left(\mathbf{b}^{(1)},\ldots,\mathbf{b}^{(K)}\right) = \left\{-\mathbf{b}^{(k)}\right\}_{k=1,\ldots,K}$ and $\mathbf{\Sigma}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right) = \left(\mathbf{a}^{(k)\top}\mathbf{a}^{(j)}\right) + diag(\psi_1, \psi_2,\ldots,\psi_K)$, where ψ_k is the error variance corresponding to item k. Note that the FA parametrization also corresponds to the theta parametrization in MPLUS (Asparouhov & Muthén, 2020). Because $\mathbf{\Sigma}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right)$ corresponds to more than one $\mathbf{P}\left(\mathbf{a}^{(1)},\ldots,\mathbf{a}^{(K)}\right)$, the FA parametrization is not identifiable unless one sets $\psi_1 = \psi_2 = \cdots = \psi_K = 1$ (Takane & de Leeuw, 1987).

4. General Discussion

Two unsolved questions concerning the identifiability of PM or PCM have been raised: (a) Are PCM and/or PM with latent elliptical distributions identifiable? (b) If any of them is not identifiable, can we find the minimal identifiability constraints? For (a), we proved the just-identifiability of PCM and

non-identifiability of PM by generalizing Almeida and Mouchart's (2003a) argument. In particular, we proved the just-identification of the polychoric *t* correlation model based on the copula representation. For (b), we found the sets of identifiability constraints of PM using ECAI of Tsai (2000, 2003). Our results showed that PM of LS with GSC or EC is just-identified, and that PM of CJ with UVC or EC is just-identified. Moreover, all of GSC, EC, and UVC are minimal constraints for identification. We also showed that for CJ, the latent differences underlying the Thurstonian models are ratio scales.

Identifiability is the premise of constructing consistent estimators. While researchers have advanced in proving the identifiability of several commonly used psychometric models (Fariña et al., 2019; Gu & Xu, 2019; Ouyang & Xu, 2022), the issue of non-identified models has not been fully addressed. ECAI provides an alternative by focusing on model hierarchy and can help researchers find identifiability constraints of non-identified general models. In particular, we showed in Theorem 3.9(i) that if the general model is identifiable, then the restricted model is identifiable. This theorem justifies the common practice that if the most general model among a set of model hierarchies is identifiable, then there is no need to prove the restricted models in the model hierarchy. For example, if a confirmatory factor analysis model is identifiable, then the model with more constraints is also identifiable.

The scales under PM with GSC or EC have more merits than PCM. Linking the results from two LS attitude tests might be more manageable. If the labels of the extremities of the anchors in the two LS tests are the same, then the two constructed polychoric models will have the same unit. We may assume that people have a consensus on understanding what the anchors mean. For example, the corresponding PM may have the same unit if two tests with the same LS anchor range from 1 (strongly disagree) to 7 (strongly agree). Alternatively, if two LSs have different anchors or different numbers of points, then these two scales might have different scale origins and units. Future studies should investigate how to equate two LS scales with different anchors or different numbers of points.

Traditionally, most researchers adopt PCM rather than PM to model LS. For researchers interested in modeling changes in longitudinal data, however, PCM does not meet the need to identify mean differences among different ages. Instead, researchers usually adopted PM with some scalar invariance assumptions to identify the mean differences (McArdle et al., 2015; Muthén, 1984). If longitudinal studies can adopt GSC, then the mean differences among different ages can be identified without further assumptions.

Identifiability has been an issue for Thurstonian models. Theorem 3.14(iv) provides a novel approach to the identifiability problem. If one adopts CJ items with more than two points, then PM is identifiable with EC. The intuition behind EC is that we define −1 and 1 as the points when a participant shows the strongest preference for one over another. Because EC can only apply to CJ with more than two points, we suggest researchers adopt CJ with more than two points for identifiability purposes. Similar to the LS case, if two CJ items have different numbers of points, they have two different units. Future studies should also investigate how to equate these two scales constructed by CJ items with different numbers of points.

Moreover, we have proved the necessary and sufficient conditions for the ordinal SEM model to be just-identifiable/identifiable. Specifically, we demonstrated that when the FA based on the correlation matrix is identifiable, the item factor analysis model within an IRT parametrization is also identifiable. However, the model within the FA parametrization remains unidentifiable unless all error variances are fixed at one. This result is consistent with the established literature about item factor analysis (Takane & de Leeuw, 1987; Wirth & Edwards, 2007). These theorems might help investigators to determine the identification of more advanced ordinal SEM models. Also, we note that these theorems release the normality assumption of ordinal SEM, providing more flexibility for statistical modeling.

We have focused on the identifiability issue of PCM and PM with latent elliptical distributions. Distributions not belonging to the family of elliptical distributions, such as the skewed normal distribution (Jin & Yang-Wallentin, 2017), need further investigation.

We mention that the approach we proposed in this article is not confined to the identifiability issue of PM and/or PCM models. In Figure 1, we provide a flowchart for establishing model identifiability and finding identifiability constraints, in which each process is annotated with the corresponding theorem(s)
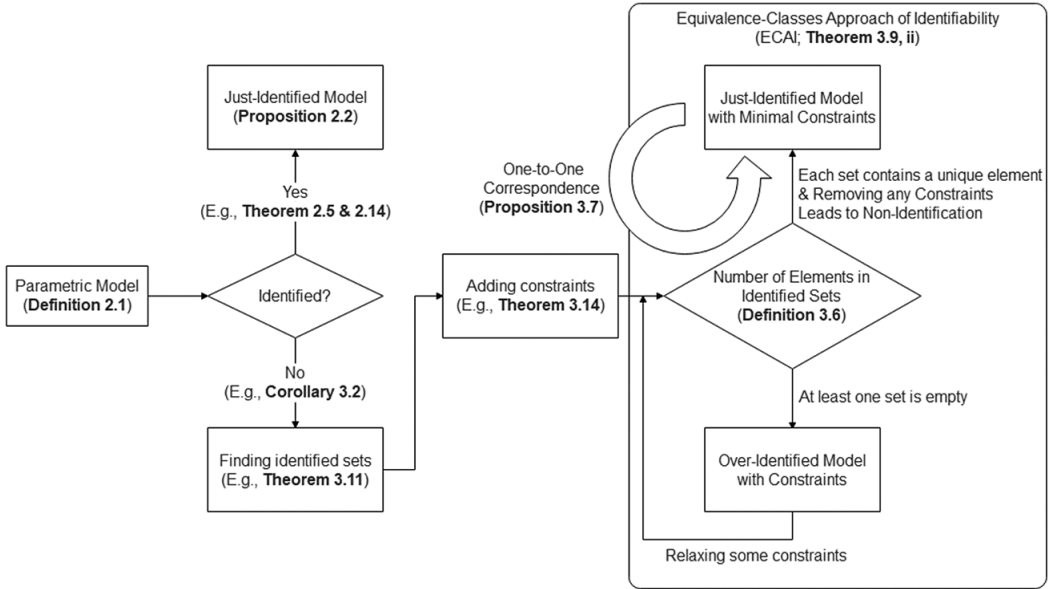
Figure 1. A flowchart for establishing model identifiability and finding identifiability constraints.
Note: In the flowchart, each process is annotated with the corresponding theorem(s) and example(s) in this article.

and example(s) in this article. The steps of this principled approach are as follows. (1) Select a parametric model. If it proves to be identifiable, then it is deemed just-identified. If, however, the model is not identifiable, we search for identifiability constraints through ECAI. (2) Upon applying some constraints, if each identified set contains at most one element, then the model is identified. If, on the other hand, at least one set is empty, then the model is over-identified, which implies a probable need to relax some constraints. (3) If each set contains a unique element and removing any constraint makes the model non-identifiable, then the model is just identified and these constraints are minimal constraints.

In this article, the concept of identifiability discussed is global identifiability. There may be some weaker conditions that might be useful. First, some commonly used latent variable models struggle to ascertain whether they are globally identifiable. In such cases, researchers can verify whether the model possesses the characteristics of local identifiability (see Bekker et al., 1994; Skrondal & Rabe-Hesketh, 2004). Second, we may conceive a weaker concept of identifiability from lower-order margins. In this study, the polychoric correlation models show "identifiability from bivariate distributions." In such a scenario, it may be possible to find consistent estimators of the parameters by only assuming bivariate distributions (without assumptions about higher-order joint distributions).[2] These estimators, which coincide with the limited-information estimator of the elliptical cases, may lead to robust estimators. These discussions potentially open a new direction for future research.

---

[2]We thank one of the reviewers for pointing this out.

References

Agresti, A. (1992). Analysis of ordinal paired comparison data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *41*, 287–297. https://doi.org/10.2307/2347562

Almeida, C., & Mouchart, M. (2003a, March 21). Identification of polychoric correlations: A copula approach. STAT Discussion Papers, Universite Catholique de Louvain.

Almeida, C., & Mouchart, M. (2003b, March 22). A note on a copula approach to polychoric correlations. STAT Discussion Papers, Universite Catholique de Louvain.

Asparouhov, T., & Muthén, B. (2020). *IRT in Mplus*. Version 4. www.statmodel.com

Bekker, P. A., Merckens, A., & Wansbeek, T. J. (1994). *Identification, equivalent models, and computer algebra*. Academic Press. https://doi.org/10.1016/B978-0-12-084775-4.50003-2

Bockenholt, U., & Tsai, R.-C. (2001). Individual differences in paired comparison data. *British Journal of Mathematical and Statistical Psychology*, *54*(2), 265–277. https://doi.org/10.1348/000711001159555

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons. https://doi.org/10.1002/9781118619179

Bonett, D. G., & Price, R. M. (2005). Inferential methods for the tetrachoric correlation coefficient. *Journal of Educational and Behavioral Statistics*, *30*(2), 213–225. http://www.jstor.org/stable/3701350

Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, *18*(1), 36–52. https://doi.org/10.1037/a0030641

Brown, A., & Maydeu-Olivares, A. (2018). Ordinal factor analysis of graded-preference questionnaire data. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(4), 516–529. https://doi.org/10.1080/10705511.2017.1392247

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Thomson Learning.

Casper, W. C., Edwards, B. D., Wallace, J. C., Landis, R. S., & Fife, D. A. (2020). Selecting response anchors with equal intervals for summated rating scales. *Journal of Applied Psychology*, *105*(4), 390–409. https://doi.org/10.1037/apl0000444

Chang, Y.-W., Hsu, N.-J., & Tsai, R.-C. (2017). Unifying differential item functioning in factor analysis for categorical data under a discretization of a normal variant. *Psychometrika*, *82*(2), 382–406. https://doi.org/10.1007/s11336-017-9562-0

Dummit, D. S., & Foote, R. M. (2004). *Abstract algebra* (3rd ed.). Prentice Hall.

Fang, K., Kotz, S., & Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Chapman & Hall.

Fariña, P., González, J., & San Martín, E. (2019). The use of an identifiability-based strategy for the interpretation of parameters in the 1PL-G and Rasch models. *Psychometrika*, *84*(2), 511–528. https://doi.org/10.1007/s11336-018-09659-w

Gu, Y., & Xu, G. (2019). The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika*, *84*(2), 468–483. https://doi.org/10.1007/s11336-018-9619-8

Hofert, M. (2018). *Elements of copula modeling with R* (1st ed.). Springer Berlin Heidelberg.

Jin, S., & Yang-Wallentin, F. (2017). Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika*, *82*(1), 67–85. https://doi.org/10.1007/s11336-016-9512-2

Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). The Guilford Press.

Kolbe, L., Oort, F., & Jak, S. (2021). Bivariate distributions underlying responses to ordinal variables. *Psych*, *3*(4), 562–578. https://doi.org/10.3390/psych3040037

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement, Vol 3: Representation, axiomatization, and invariance*. Academic Press. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1990-97738-000&site=ehost-live&scope=site

McArdle, J. J., Petway, K. T., & Hishinuma, E. S. (2015). IRT for growth and change. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment*. (1st ed., pp. 435–456). Routledge/Taylor & Francis Group.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132. https://doi.org/10.1007/BF02294210

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. https://doi.org/10.1007/BF02296207

Ouyang, J., & Xu, G. (2022). Identifiability of latent class models with covariates. *Psychometrika*, *87*, 1343–1360. https://doi.org/10.1007/s11336-022-09852-y

Pearson, K. (1901). I. Mathematical contributions to the theory of evolution. – VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *195*(262–273), 1–47. https://doi.org/10.1098/rsta.1900.0022

Reckase, M. (2009). *Multidimensional item response theory*. Springer.

Rice, J. A. (2007 *). Mathematical statistics and data analysis* (3rd ed.). Thomson/Brooks/Cole.

Rodgers, J. L. (2019). Degrees of freedom at the start of the second 100 years: A pedagogical treatise. *Advances in Methods and Practices in Psychological Science*, *2*(4), 396–405. https://doi.org/10.1177/2515245919882050

Roscino, A., & Pollice, A. (2006). A generalization of the polychoric correlation coefficient. In S. Zani, A. Cerioli, M. Riani, & M. Vichi (Eds.), *Data analysis, classification and the forward search* (pp. 135–142). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-35978-8_16

Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores. *ETS Research Bulletin Series*, *1968*(1), i–169. https://doi.org/10.1002/j.2333-8504.1968.tb00153.x

Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). Springer New York. https://doi.org/10.1007/978-1-4757-2691-6_5

San Martín, E., & Quintana, F. (2002). Consistency and identifiability revisited. *Brazilian Journal of Probability and Statistics*, *16*(1), 99–106. http://www.jstor.org/stable/43601008

SAS Institute Inc. (2024). The IRT procedure. In *SAS/STAT® user's guide*. SAS Institute Inc.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Chapman & Hall/CRC. http://www.loc.gov/catdir/enhancements/fy0646/2004042808-d.html

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680. https://doi.org/10.1126/science.103.2684.677

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*(3), 393–408. https://doi.org/10.1007/BF02294363

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286. https://doi.org/10.1037/h0070288

Torgerson, W. S. (1952). Multidimensional scaling: I Theory and method. *Psychometrika*, *17*, 401–419. https://doi.org/10.1007/BF02288916

Tsai, R.-C. (2000). Remarks on the identifiability of Thurstonian ranking models: Case V, case III, or neither? *Psychometrika*, *65*(2), 233–240. https://doi.org/10.1007/Bf02294376

Tsai, R.-C. (2003). Remarks on the identifiability of Thurstonian paired comparison models under multiple judgment. *Psychometrika*, *68*(3), 361–372. https://doi.org/10.1007/BF02294732

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods*, *12*(1), 58–79. https://doi.org/10.1037/1082-989X.12.1.58

Zwislocki, J. J., & Goodman, D. A. (1980). Absolute scaling of sensory magnitudes: A validation. *Perception & Psychophysics*, *28*, 28–38. https://doi.org/10.3758/BF03204312

Appendix

The statements shown here are used in the proof of Theorem 2.14 for the identification of PCM with a latent multivariate t distribution.

Lemma A.1. *Suppose that X and Y are independent random variables with $X \sim Gamma(a_1, b), Y \sim Gamma(a_2, b)$, for $a_1, a_2, b > 0$. Then $X + Y \sim Gamma(a_1 + a_2, b)$.*

   Lemma A.2 is used to establish the monotonicity of the CDF of the product of a jointly distributed random variable.

Lemma A.2. *If X and Y are two independent, continuous random variables and $Z = XY$, then Z is a continuous random variable. Moreover, if X is supported on the positive values and Y is supported on the whole real line, then Z is supported on the whole real line, and the CDF of Z is strictly increasing and injective on the real line.*

Proof. Suppose that there exists z_0 such that $f(z_0) = \int_{-\infty}^{\infty} f_X(x) f_Y\left(\frac{z_0}{x}\right) \frac{1}{|x|} dx = 0$. Because $f_X(x) f_Y\left(\frac{z_0}{x}\right) \frac{1}{|x|} \geq 0$, we have $f_X(x) f_Y\left(\frac{z_0}{x}\right) \frac{1}{|x|} = 0$ almost everywhere. However, $f_X(x)$ and $f_Y\left(\frac{z_0}{x}\right)$ are nonzero almost everywhere, so z_0 cannot exist. Thus Z is supported on the whole real line. Now because for any $b > a$, where ε is the infimum of within $[a, b]$. Thus the distribution of Z is strictly increasing, and so the function is injective on the real line. □

Corollary A.3. *A t distribution is supported on the whole real line, and the CDF is strictly increasing and injective on the real line.*

Lemma A.4. *Let t_v be the CDF of a t distribution with degrees of freedom v. Then for any $v_2 > v_1 > 0$,*

$$\begin{cases} t_{v_1}(x) > t_{v_2}(x), x < 0 \\ t_{v_1}(x) < t_{v_2}(x), x > 0. \end{cases}$$

Therefore, $\partial_v F(x|v)$ and x have the same sign. More generally, let $F(\mathbf{x}|v)$ be the CDF of a multivariate t distribution $t_v(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $x = (x_1, x_2, \ldots, x_p)$. If $v_2 > v_1$, then

$$\begin{cases} F(\mathbf{x}|v_1) > F(\mathbf{x}|v_2), \mathbf{x} < \boldsymbol{\mu}, \\ F(\mathbf{x}|v_1) < F(\mathbf{x}|v_2), \mathbf{x} > \boldsymbol{\mu}. \end{cases}$$

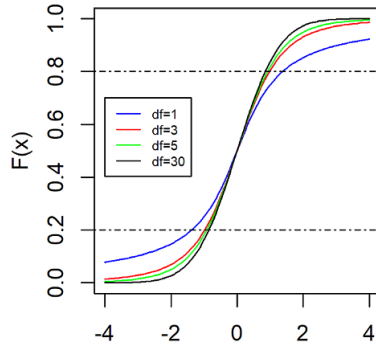**Figure A1.** The CDF of the *t* distribution with different degrees of freedom.
Note: Fixing the target probability $p < 0.5$, then the quantile is a strictly increasing function of v; if $p > 0.5$, then the quantile is a strictly decreasing function of v.

Proof. By the definition of *t* distribution, if $X \sim t_{v_1}$, then $X = \frac{Z}{\sqrt{S/v_1}}$ for some $S \sim \chi^2_{v_1}$ and $Z \sim \mathcal{N}(0,1)$. Without loss of generality, for any $x < 0$,

$$
\begin{aligned}
t_{v_1}(x) &= P\left(\frac{Z}{\sqrt{S/v_1}} \leq x \middle| v_1 \right) \\
&= P\left(\frac{Z\sqrt{v_1}}{x} \geq \sqrt{S} \middle| v_1 \right) \\
&= \int_{-\infty}^{\infty} P\left(S \leq \frac{z^2 v_1}{x^2} \middle| Z = z \right) dF_Z(z) \\
&= \int_{-\infty}^{\infty} P\left(S \leq \frac{z^2 v_1}{x^2} \right) dF_Z(z) \qquad (S \perp Z) \\
&> \int_{-\infty}^{\infty} P\left(S \leq \frac{z^2 v_2}{x^2} \right) dF_Z(z) \\
&\geq \int_{-\infty}^{\infty} P\left(S + S_1 \leq \frac{z^2 v_2}{x^2} \right) dF_Z(z) \qquad \left(S_1 \sim \chi^2_{v_2 - v_1}, S_1 \perp S; S_1 \text{ is nonnegative} \right) \\
&= \int_{-\infty}^{\infty} P\left(S_2 \leq \frac{z^2 v_2}{x^2} \right) dF_Z(z) \qquad \left(S_2 := S_1 + S; \text{By Lemma A.1}, S_2 \sim \chi^2_{v_2} \right) \\
&= P\left(\frac{Z}{\sqrt{S_2/v_2}} \leq x \middle| v_2 \right) \qquad (S_2 \perp Z) \\
&= t_{v_2}(x)
\end{aligned}
$$

Likewise, for any $x > 0, t_{v_1}(x) < t_{v_2}(x)$. The multivariate case can be proved using a similar method. $\square$

Proposition A.5 shows that the quantile of the *t* distribution is an implicit function of the degrees of freedom. Figure 1 demonstrates this property.

Proposition A.5. *Let q be the u-quantile of the t distribution, i.e., $u = t_v(q)$. Given u, if $0 < u < 0.5$ then $q(v)$ is strictly increasing and invertible on $q < 0$; if $0.5 < u < 1$ then $q(v)$ is strictly decreasing and invertible on $q > 0$. In other words, for any $v_2 > v_1 > 0$,*

$$
\begin{cases}
t^{-1}_{v_2}(u) > t^{-1}_{v_1}(u), & 0 < u < 0.5 \\
t^{-1}_{v_2}(u) < t^{-1}_{v_1}(u), & 0.5 < u < 1.
\end{cases}
$$

Proof. Fix $0.5 < u < 1$, and let q_1 and q_2 be the u quantile of t distributions with degrees of freedom v_1 and v_2, respectively. The uniqueness of the quantiles is guaranteed by Corollary A.3. Also, based on the property of t distributions, $q_1, q_2 > 0$. Then, based on Lemma A.4, for any $v_2 > v_1, u = t_{v_1}(q_1) = t_{v_2}(q_2) > t_{v_1}(q_2)$. By Corollary A.3, $q_2 < q_1$, therefore, $q(v)$ is strictly increasing. Similarly, for $0 < u < 0.5, q(v)$ is strictly decreasing. Also, for any fixed values of u and v, there is a value q such that $u = t_v(q)$. Thus $q(v)$ is invertible on both $q < 0$ and $q > 0$. $\qquad\square$