On the Identifiability of Tensor Ranks via Prior Predictive Matching

Eliezer da Silva
BCAM – Basque Center for Applied Mathematics
University of Coimbra, CISUC/LASI, DEI

Arto Klami
Department of Computer Science
University of Helsinki

Diego Mesquita
School of Applied Mathematics
Getúlio Vargas Foundation

Iñigo Urteaga
BCAM — Basque Center for Applied Mathematics
IKERBASQUE — Basque Foundation for Science

Abstract

Selecting the latent dimensions (ranks) in tensor factorization is a central challenge that often relies on heuristic methods. This paper introduces a rigorous approach to determine rank identifiability in probabilistic tensor models, based on prior predictive moment matching. We transform a set of moment matching conditions into a log-linear system of equations in terms of marginal moments, prior hyperparameters, and ranks; establishing an equivalence between rank identifiability and the solvability of such system. We apply this framework to four foundational tensor-models, demonstrating that the linear structure of the PARAFAC/CP model, the chain structure of the Tensor Train model, and the closed-loop structure of the Tensor Ring model yield solvable systems, making their ranks identifiable. In contrast, we prove that the symmetric topology of the Tucker model leads to an underdetermined system, rendering the ranks unidentifiable by this method. For the identifiable models, we derive explicit closed-form rank estimators based on the moments of observed data only. We empirically validate these estimators and evaluate the robustness of the proposal.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

1 INTRODUCTION

Tensor factorizations have become indispensable tools for analyzing multi-way data across diverse contexts and fields such as neuroscience (Cong et al., 2015; Mosayebi and Hossein-Zadeh, 2020), bioinformatics and genomics (Alter et al., 2003), chemometrics (Bro and Kiers, 2003), computational social science (Schein et al., 2015), network science (Contisciani et al., 2022), recommender systems (Bhargava et al., 2015; Gopalan et al., 2015), supervised learning (Stoudenmire and Schwab, 2016), compression of large neural networks (Novikov et al., 2015), and trace back to early work by Hitchcock (1927). These methods decompose a high-dimensional tensor into a set of lower-dimensional factors, revealing latent structures and facilitating data interpretation and prediction (Chu and Ghahramani, 2009; Kolda and Bader, 2009; Cichocki et al., 2009).

A fundamental challenge in the design and application of tensor models is the selection of their latent dimensions, i.e., the **ranks** (r_p), which dictate the model's complexity and its ability to capture the observed statistics of the data, without overfitting. In contrast to simple matrix factorizations, most tensor models have multiple ranks, i.e., $p \geq 1$, depending on the specific form of the factorization. The determination of the tensor ranks is, in general, a NP-Complete problem for tensors over $\mathbb{Q}$ (Håstad, 1990), NP-Hard for tensors over $\mathbb{R}$ (Hillar and Lim, 2013) and hard to approximate in general (Swernofsky, 2018).

Bayesian approaches to tensor factorization (Chu and Ghahramani, 2009; Zhao et al., 2015; Rai et al., 2014) provide a principled alternative to exhaustive search, enabling uncertainty quantification, prior incorporation, and even automatic rank inference via non-parametric methods (Porteous et al., 2008) or automatic relevance determination (Zhao et al., 2015).

However, parametric models often demand user-specified ranks or strong priors, complicating deployment in real-world scenarios where domain knowledge is scarce. For matrix factorization models, the prior specification hurdle can be mitigated through prior predictive checks or empirical Bayes techniques (da Silva et al., 2023; Wang and Stephens, 2021). That is, we can determine the matrix rank by comparing observed data with data simulated from the model under alternative ranks. This established technique for matrix factorization inspires an analogous development for tensor factorization we study here.

The fundamental question we ask here is *when can a prior predictive moment matching approach identify the ranks of tensor factorization models*. To answer the question, we need to establish a range of tools for connecting the moments of the prior predictive distribution with the ranks of the tensor model and an approach to use them for characterizing the rank identifiability of the model.

In this work, we formulate a strategy for analyzing the identifiability of ranks in probabilistic tensor factorization models, by analyzing a log-linear system of equations derived from matching model-specific prior predictive moments with low-order moments of the observed tensor-data. We show that the algebraic structure of this system, which is determined by the chosen tensor decomposition, dictates whether the ranks can be uniquely identified by the method of moment-matching. We formalize this analysis by transforming the moment equations into a log-linear system, where identifiability is equivalent to the solvability of the system with respect to the unknown ranks.

Fundamentally, this framework establishes a link between a tensor factorization model's structure and its rank identifiability. We provide a clear theoretical characterization and distinction between different tensor models, demonstrating that rank identifiability is a direct consequence of the interaction topology of their latent factors. In addition, we showcase how ranks can be empirically estimated from observed tensor-data according to the prior predictive matching method.

Our specific contributions are:

1. We establish a general, prior predictive based framework for analyzing rank identifiability in probabilistic tensor models, by examining the rank of a log-linear system of moment equations.
2. We apply this framework to four foundational tensor decompositions: **Tucker** (Tucker, 1966), **PARAFAC/CP** (Hitchcock, 1927; Harshman et al., 1970; Carroll and Chang, 1970), **Tensor Train** (**TT**, Oseledets, 2011), and **Tensor Ring** (**TR**, Zhao et al., 2016).
3. We prove that the ranks of the standard Tucker model are not identifiable with this method due to symmetries in its moment structure leading to a degenerate, unsolvable system of equations.
4. We prove that the ranks of the PARAFAC/CP, TT and TR models are identifiable from the first and second moments and derive explicit closed-form estimators for them.
5. We present and evaluate a complete and robust pipeline for the estimation of ranks from observed tensor-data in identifiable models.

2 A GENERAL FRAMEWORK FOR RANK IDENTIFIABILITY

We hereby introduce the notation and assumptions to establish a general framework for rank identifiability in probabilistic tensor factorization models.

Notation. Let $\boldsymbol{Y} \in \mathbb{R}^{N_1 \times \cdots \times N_M}$ be an observed tensor with M modes. A *mode* of an order-M tensor is one of its M dimensions. Let $\mathbf{i} = (i_m)_{m \in [M]} \in \bigtimes_{m=1}^{M} [N_m]$ be the multi-index from the observed tensor domain, where $[N_m] = \{1, \ldots, N_m\}$. Let $\mathbf{r} = (r_p)_{p \in [R]}$ be the multi-rank vector of size $R \leq M$. Let $\beta \in \mathcal{K} = \bigtimes_{p=1}^{R} [r_p]$ be the latent multi-index over the rank domains. For multi-indices $\mathbf{i}, \mathbf{j}$, the shared-mode set is $S(\mathbf{i}, \mathbf{j}) = \{k : i_k = j_k\}$. We use bold sans-serif for tensors ($\boldsymbol{\eta}$) and vectors/matrices ($\boldsymbol{\theta}^{(p)}$). All latent factor multi-indices $\mathbf{i}_p(\mathbf{i}, \beta)$ are formed by combining some components of $\mathbf{i}$ and the latent index β, and each factor is probabilistically defined as $\theta^{(p)}_{\mathbf{i}_p(\mathbf{i},\beta)} \sim \pi_p(\mu_p, \sigma_p^2)$, i.i.d. across indices with $p \in \mathcal{P}$ indexing and grouping the factors.

Definition 2.1 (The Probabilistic Tensor-Model)**.** *We assume that elements in tensor $\boldsymbol{Y} \in \mathbb{R}^{N_1 \times \cdots \times N_M}$ are probabilistically generated from a rate tensor $\boldsymbol{\eta} \in \mathbb{R}^{N_1 \times \cdots \times N_M}$, whose structure is defined by a (model-dependent) tensor decomposition with latent factors of ranks $\mathbf{r} = (r_1, \ldots, r_R)$ of the following generic form:*

$$\eta_{\mathbf{i}} = \sum_{\beta \in \mathcal{K}} \prod_{p \in \mathcal{P}} \theta^{(p)}_{\mathbf{i}_p(\mathbf{i},\beta)}, \tag{1}$$

with latent indices defined by $\mathbf{i}_p(\mathbf{i}, \beta)$ and parameters $\theta^{(p)}_{\mathbf{i}_p(\mathbf{i},\beta)} \sim \pi_p(\mu_p, \sigma_p^2)$, drawn independently from a factor p-specific location-scale prior distribution $\pi_p(\cdot)$ with up-to-second order moments (μ_p, σ_p^2).

Tensor observations $Y_{\mathbf{i}}$ are conditionally independent given the rate tensor $\boldsymbol{\eta}$, with conditional moments $\mathbb{E}[Y_{\mathbf{i}}|\boldsymbol{\eta}] = \eta_{\mathbf{i}}$ and $\mathrm{Var}(Y_{\mathbf{i}}|\boldsymbol{\eta}) = \phi(\eta_{\mathbf{i}})$ for some model-dependent positive function $\phi(\cdot)$.

This probabilistic tensor formulation allows for different observation models (e.g. Poisson with $\phi(\eta) = \eta$, or Gaussian with $\phi(\eta) = \sigma_Y^2$ variance), and it generalizes existing matrix factorization frameworks (da Silva et al., 2023), setting the stage for the prior predictive moment matching analysis. Commonly used factorization models such as Tucker, PARAFAC/CP, TR and TT, are all instances of this general formulation [1].

Observable data and moments of interest. The second moments of the observable tensor-data can be described by two key quantities: the total covariance and the pure interaction terms, which form the key building blocks for the analysis that follows.

Definition 2.2 (Total Covariance and Pure Interaction Terms)**.** *Let* $\mathbf{i} = (i_1, \ldots, i_M)$ *and* $\mathbf{j} = (j_1, \ldots, j_M)$ *be two multi-indices from the tensor's domain. Let* $S(\mathbf{i}, \mathbf{j}) = \{k \mid i_k = j_k\}$ *be the set of indices they share.*

The ***total observable covariance****,* C_S*, is the theoretical covariance between two multi-index tensor entries* $\mathbf{i}$ *and* $\mathbf{j}$*:*

$$C_S = Cov(Y_{\mathbf{i}}, Y_{\mathbf{j}}) \ .$$

Given a factorization structure, this value depends only on the set of shared indices $S(\mathbf{i}, \mathbf{j})$*. For entries sharing no indices, we assume* $C_\emptyset = 0$*.*

The ***pure interaction term****,* v_S*, represents the variance arising uniquely from the interaction of factors in* $S(\mathbf{i}, \mathbf{j})$*, and induces the definition of the total covariance in terms of their additive contribution over the non-empty subsets of* $S(\mathbf{i}, \mathbf{j})$*:*

$$C_S = \sum_{\emptyset \neq S' \subseteq S} v_{S'} \ .$$

These pure interaction terms can be also defined via the principle of inclusion-exclusion from the observable covariances:

$$v_S = \sum_{S' \subseteq S} (-1)^{|S| - |S'|} C_{S'} .$$

Connecting theoretical and observable moments. For many tensor decomposition models, the theoretical moments of the rate tensor $\boldsymbol{\eta}$ can be expressed as polynomials: i.e., as products of the unknown parameters (ranks and prior moments).

Proposition 2.3 (Polynomial Structure of Rate Moments)**.** *For a probabilistic tensor model congruent with Definition 2.1, every pure interaction term* v_S *is a monomial in the unknown model hyperparameters* (r_p, μ_p, σ_p^2)*. Therefore, the marginal mean* $\mathbb{E}[\boldsymbol{\eta}]$*, the variance* $\mathrm{Var}(\boldsymbol{\eta})$*, and the covariance terms* $\mathrm{Cov}(\eta_{\mathbf{i}}, \eta_{\mathbf{j}})$ *can be expressed as polynomials of* $\{r_p, \mu_p, \sigma_p^2\}$*.*

Proof. Each rate element $\eta_{\mathbf{i}}$ is a sum-product of the latent factor elements $\{\theta^{(p)}_{\mathbf{i}_p(\mathbf{i},\beta)}\}$. Applying total expectation we obtain $\mathbb{E}[\boldsymbol{Y}_{\mathbf{i}}] = \sum_{\beta \in \mathcal{K}} \prod_p \mu_p$, with the summation index depending on the rank dimensions, and resulting in the monomial given by products of ranks $\{r_p\}$ and means $\{\mu_p\}$. Applying total covariance and conditional independence, the computation leads to moments $\mathbb{E}[\eta_{\mathbf{i}} \eta_{\mathbf{j}}]$, which resolve into a double sum-product over latent indexes, with σ_p^2 appearing for factors with the same latent index, μ_p^2 for independent factors, and powers of r_p from interactions of repeated sums and multiplications, leading to a polynomial on prior parameters $\{\mu_p, \sigma_p^2\}$ and ranks $\{r_p\}$. □

2.1 The Log-Linear System of Prior Predictive Moment Equations

The insight unveiled by Proposition 2.3 is that observable moments (like covariances C_S) are polynomials of the model parameters. More precisely, the individual terms of these polynomials are the **pure interaction terms** v_S, which are **monomials**. This monomial (multiplicative) structure is ideal for linearization. Namely, by taking the logarithm of the equation for each pure interaction term, we transform the system of prior predictive moments into a linear one, amenable to standard algebraic analysis.

Definition 2.4 (The Log-Linear System of Rate Moments)**.** *The log-linear system derived from the monomial equations of pure interaction terms is given by* $\mathbf{A}\mathbf{x} = \mathbf{b}$*, where* $\mathbf{x}$ *is the vector of the logarithms of the unknown parameters, i.e.,* $\log(r_p), \log(\mu_p^2), \log(\sigma_p^2)$*;* $\mathbf{b}$ *is the vector of the logarithms of the (estimable) moment terms, i.e.,* $\log(E[Y]), \log(v_S)$*; and* $\mathbf{A}$ *is the integer-valued* ***design matrix****, where each entry* A_{ij} *is the exponent of the* j*-th unknown parameter in the monomial expression for the* i*-th moment term.*

While the invertibility of the full matrix $\mathbf{A}$ determines if *all* parameters are identifiable, our primary interest is in the ranks of the tensors. The identifiability of the ranks hinges on whether they can be isolated from the other unknown parameters. This leads to a more precise, and relevant, principle of rank identifiability:

Proposition 2.5 (Principle of Rank Identifiability)**.** *The ranks* $\{r_p\}$ *of a tensor model are identifiable from the second moments of tensor-data if and only if the unknown prior-moment parameters can be algebraically eliminated from the full log-linear system of moments* $\mathbf{A}\mathbf{x} = \mathbf{b}$*.*

[1] If $\mathbf{i}_p(\mathbf{i}, \beta)$ is formed by taking the p-th components of $\mathbf{i}$ and β respectively, for $p > 0$, and for $p = 0$ define a special $\mathbf{i}_0(\mathbf{i}, \beta) = \beta$, we obtain Tucker. While doing the same but defining β as a one dimensional index shared by all, without a special $p = 0$ index, we obtain PARAFAC/CP.

Consequently, it yields a non-trivial ***reduced system*** *of the log-ranks* $\mathbf{x}_r$ *alone: i.e.,* $\mathbf{A}_{red}\mathbf{x}_r = \mathbf{b}_{red}$*, where the reduced design matrix* $\mathbf{A}_{red}$ *has full rank. A full-rank* $\mathbf{A}_{red}$ *guarantees a unique solution for the ranks, meaning they occupy a parameter subspace that can be uniquely determined from the data's moments.*

Propositions 2.3 and 2.5 establish a direct link between the model-dependent, prior predictive moments and the moments of the observed tensor-data. This connection is formalized through an identifiability analysis of a reduced, rank-specific log-linear system.

However, we note that the utility of this link for rank estimation is fully model-dependent. We first present in Section 3 a cautionary example where the dependencies between prior hyperparameters and observed second moments are not identifiable for the popular Tucker tensor model. In contrast, we highlight in Section 4 several models where identifiability holds, enabling the design and implementation of a novel method for estimating tensor ranks directly and exclusively from observed moments.

3 A CAUTIONARY TALE: THE TUCKER MODEL

We now examine the standard Tucker decomposition and show that its symmetric "hub-and-spoke" Carlsson and Jia (2013) interaction topology precludes forming a solvable reduced system of equations for the Tucker model's ranks.

3.1 Model Definition and Its Moments

Definition 3.1 (The Tucker model)**.** *For an order-M tensor, with ranks* $\mathbf{r} = (r_1, \ldots, r_M)$ *the rate* $\boldsymbol{\eta}$ *is defined by the Tucker decomposition as*

$$\eta_{i_1 \ldots i_M} = \sum_{k_1=1}^{r_1} \cdots \sum_{k_M=1}^{r_M} G_{k_1 \ldots k_M} \prod_{p=1}^{M} \theta^{(p)}_{i_p,k_p} \ .$$

This model, introduced by Tucker (1966), is widely used for its interpretability and versatility in multi-way analysis.

The ranks to be estimated $(r_1, \ldots, r_M)$ are the dimensions of the core tensor G. As per Definition 2.1 the elements of G and each factor $\theta^{(p)}$ are i.i.d. random variables with moments (μ_G, σ_G^2) and (μ_p, σ_p^2), respectively.

Lemma 3.2 (Moment Structure of the Tucker Model)**.** *For the order-M Tucker model, the squared mean and the pure interaction terms involving the core* (v_G)*, a single factor* (v_p)*, or both* $(v_{G,p})$ *obey —see detailed derivations in Appendix A.2:*

$$\begin{aligned}
(E[Y])^2 &= \mu_G^2 \prod_{q=1}^{M} (r_q^2 \mu_q^2) \ , \\
v_G &= \sigma_G^2 \prod_{q=1}^{M} (r_q^2 \mu_q^2) \ , \\
v_p &= \mu_G^2 (r_p \sigma_p^2) \prod_{q \neq p} (r_q^2 \mu_q^2) \ , \\
v_{G,p} &= \sigma_G^2 (r_p \sigma_p^2) \prod_{q \neq p} (r_q^2 \mu_q^2) \ .
\end{aligned}$$

3.2 Rank Identifiability Analysis

We now prove that the symmetric topology of the Tucker model leads to a degeneracy in its moment equations by showing that the reduced design matrix for the ranks is the zero matrix. Hence, this structure makes the ranks fundamentally unrecoverable from observed tensor-data.

Lemma 3.3 (An Identity Among Observables)**.** *In the log-linear system for the Tucker model moments, the following M independent relations hold for every mode* $p = 1, \ldots, M$*:*

$$\log v_p - 2 \log E[Y] = \log v_{G,p} - \log v_G \ . \qquad (2)$$

Proof. We show that both sides of the equation reduce to the same expression. Using the monomials from Lemma 3.2, the left-hand side is:

$$\begin{aligned}
\log v_p - 2 \log E[Y] &= \log \left[\mu_G^2 r_p \sigma_p^2 \prod_{q \neq p} (r_q^2 \mu_q^2) \right] \\
&\quad - 2 \log \left[\mu_G \prod_{q=1}^{M} (r_q \mu_q) \right] \\
&= \log \left(\frac{\mu_G^2 r_p \sigma_p^2 \prod_{q \neq p} (r_q^2 \mu_q^2)}{\mu_G^2 \prod_{q=1}^{M} (r_q^2 \mu_q^2)} \right) \\
&= \log \left(\frac{1}{r_p} \frac{\sigma_p^2}{\mu_p^2} \right) \ .
\end{aligned}$$

An identical calculation shows that $\log v_{G,p} - \log v_G$ also simplifies to $\log \left(\frac{1}{r_p} \frac{\sigma_p^2}{\mu_p^2} \right)$, demonstrating the identity in Equation 2. □

Theorem 3.4 (Non-identifiability of Tucker ranks)**.** *The Tucker model ranks* $(r_1, \ldots, r_M)$ *are* ***not identifiable*** *directly from the second moments of the tensor-data.*

Proof. We prove that the reduced design matrix $\mathbf{A}_{\text{red}}$ is the zero matrix.

The identity established in Lemma 3.3 provides M linear dependencies among the rows of the full design matrix $\mathbf{A}$. Because these M identities express every standard-basis vector $\mathbf{e}_p$ (the column that selects $\log r_p$) as a linear combination of columns that are free of any log-rank variable, the M columns corresponding to the log-ranks are redundant. Consequently, the reduced design matrix $\mathbf{A}_{\text{red}}$ of Proposition 2.5 is the $0 \times M$ zero matrix, and the ranks are unidentifiable. □

4 EXEMPLARY TALES: IDENTIFIABLE RANKS WITH CLOSED-FORM EXPRESSION

In contrast to the analysis in the previous section, we present below the general solutions for the ranks of the PARAFAC/CP, Tensor Train, and Tensor Ring models for an arbitrary order M tensor, showcasing how to establish identifiability according to the principles formalized in Proposition 2.5.

4.1 PARAFAC/CP Model

The PARAFAC or Canonical Polyadic (CP) model's rank is identifiable due to a unique property of its second moments: the additivity of its variance components. Precisely, its total observable covariance decomposes into a clean sum of pure, monomial interaction terms, because the model is defined as a sum of r independent rank-one tensors. This structure provides the necessary constraints for a solvable rank system.

4.1.1 Model Definition and Its Moments

Definition 4.1 (The PARAFAC/CP model)**.** *For an order-M tensor, the rate $\boldsymbol{\eta}$ of a PARAFAC/CP decomposition of rank r follows:*

$$\eta_{\mathbf{i}} = \eta_{i_1 \dots i_M} = \sum_{k=1}^{r} \prod_{p=1}^{M} \theta_{i_p,k}^{(p)} \ .$$

The PARAFAC/CP decomposition, originally proposed by Hitchcock (1927) , with later formulations by Harshman et al. (1970) and Carroll and Chang (1970), assumes a sum of rank-1 tensors and is widely used in multiple fields given its simple structure and applicability. The variance calculation of the CP model results in a simple and regular form for its pure interaction terms, as formalized below, a fundamental property leading to a closed-form expression for the rank r.

Lemma 4.2 (Monomial Moment structure of PARAFAC/CP)**.** *For an order-M PARAFAC/CP model with rank r and factor moments (μ_k, σ_k^2), the squared mean and the pure-interaction terms obey*

$$\begin{aligned}
(E[Y])^2 &= r^2 \prod_{k=1}^{M} \mu_k^2 \ , \\
v_{\{p\}} &= r\sigma_p^2 \prod_{k \neq p} \mu_k^2 \ , \\
v_{\{p,q\}} &= r\sigma_p^2\sigma_q^2 \prod_{k \neq p,q} \mu_k^2 \qquad (p \neq q) \ .
\end{aligned}$$

Proof. The rate is $\eta_{\mathbf{i}} = \sum_{s=1}^{r} \prod_{k=1}^{M} \theta_{i_k,s}^{(k)}$. Because all factor entries across different components s are independent, the variance of the sum is the sum of the variances. The pure interaction term v_S is the component of the total variance arising from the variance of factors in the set S and the mean of factors not in S. For $v_{\{p\}}$, this is:

$$\begin{aligned}
v_{\{p\}} &= \mathrm{Var}_{\text{pure}}(\eta_{\mathbf{i}} \mid \text{mode } p) \\
&= \sum_{s=1}^{r} \mathrm{Var}(\theta_{i_p,s}^{(p)}) \prod_{k \neq p} \left(E[\theta_{i_k,s}^{(k)}] \right)^2 \\
&= r\sigma_p^2 \prod_{k \neq p} \mu_k^2 \ .
\end{aligned}$$

The derivations for other pure terms follow the same principle of summing variances over r independent components. □

4.1.2 Rank Identifiability Analysis

The simple monomial structure of the moments leads directly to a closed-form solution for the rank.

Theorem 4.3 (Closed-Form Rank Estimator for PARAFAC/CP)**.** *For any two distinct modes $p \neq q$, the rank r is identifiable and given by:*

$$r = \frac{v_{\{p,q\}} \cdot (E[Y])^2}{v_{\{p\}} \cdot v_{\{q\}}} \ . \tag{3}$$

Proof. We substitute the monomials from Lemma 4.2 into the expression:

$$\begin{aligned}
\frac{v_{\{p,q\}} \cdot (E[Y])^2}{v_{\{p\}} \cdot v_{\{q\}}} &= \frac{\left[r\sigma_p^2\sigma_q^2 \prod_{k \neq p,q} \mu_k^2\right] \cdot \left[r^2 \prod_{k=1}^{M} \mu_k^2\right]}{\left[r\sigma_p^2 \prod_{k \neq p} \mu_k^2\right] \cdot \left[r\sigma_q^2 \prod_{k \neq q} \mu_k^2\right]} \\
&= \frac{r^3 \left(\prod_{k \neq p,q} \mu_k^2\right) \left(\mu_p^2\mu_q^2 \prod_{k \neq p,q} \mu_k^2\right)}{r^2 \left(\mu_q^2 \prod_{k \neq p,q} \mu_k^2\right) \cdot \left(\mu_p^2 \prod_{k \neq p,q} \mu_k^2\right)} \\
&= r.
\end{aligned}$$

Every nuisance parameter cancels, yielding the rank itself. Hence, r is identifiable and the formula is exact. □

4.2 The Tensor Train Model

The Tensor Train (TT) model's topology is an open, asymmetric chain of interactions (Oseledets, 2011). This sequential structure breaks the symmetries that cause non-identifiability in the Tucker model, providing a sufficient number of distinct algebraic constraints to solve for all of the model's internal ranks.

4.2.1 Model Definition and Its Moments

Definition 4.4 (The Tensor-Train model)**.** *For an order-M tensor, the rate $\boldsymbol{\eta}$ of a Tensor Train decomposition with TT-ranks $(r_1, \ldots, r_{M-1})$ follows*

$$\eta_{i_1 \ldots i_M} = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} \theta^{(1)}_{1,i_1,k_1} \theta^{(2)}_{k_1,i_2,k_2} \cdots \theta^{(M)}_{k_{M-1},i_M,1}.$$

The ranks to be estimated are the internal bond dimensions $(r_1, \ldots, r_{M-1})$, which we refer to as *interior ranks*. A full derivation of the general TT moments involves a series of (repeated) algebraic calculations, which we present in Appendix A.3.

However, their algebraic structure contains the necessary asymmetry for rank identification. The key is that pure interaction terms involving non-contiguous modes (e.g., $\{p-1, p+2\}$) have a different dependence on the ranks than terms involving contiguous modes (e.g., $\{p, p+1\}$).

Lemma 4.5 (Monomial Moment Structure of Tensor–Train)**.** *The pure-interaction monomials for the TT model are functions of the ranks and prior moments. Their exact form depends on the chosen modes in S, but they all exhibit a crucial asymmetry. For an interior rank r_p, the relevant monomials for its identification are:*

$$v_{p+1} = r_p r_{p+1} r_{p+2} \mu_p^2 \sigma_{p+1}^2 \mu_{p+2}^2 \prod_{k \neq p,p+1,p+2} \mu_k^2 \,,$$

$$v_{p,p+1} = r_p r_{p+1} r_{p+2} \sigma_p^2 \sigma_{p+1}^2 \mu_{p+2}^2 \prod_{k \neq p,p+1,p+2} \mu_k^2 \,,$$

$$v_{p-1,p+2} = r_{p-1} r_p r_{p+1} r_{p+2} \sigma_{p-1}^2 \mu_p^2 \sigma_{p+2}^2 \prod_{k \neq p-1,p,p+2} \mu_k^2 \,,$$

$$v_{p-1,p,p+2} = r_{p-1} r_p r_{p+1} r_{p+2} \sigma_{p-1}^2 \sigma_p^2 \sigma_{p+2}^2 \prod_{k \neq p-1,p,p+2} \mu_k^2 \,.$$

Notably, the following identities hold:

$$\frac{v_{\{p,p+1\}}}{v_{\{p+1\}}} = \frac{\sigma_p^2}{\mu_p^2} \,, \tag{4}$$

$$\frac{v_{\{p-1,p,p+2\}}}{v_{\{p-1,p+2\}}} = \frac{1}{r_p} \frac{\sigma_p^2}{\mu_p^2} \,. \tag{5}$$

Informal Proof. The TT rate is a sum over products of core tensor entries. Each pure-interaction term is obtained by calculating the covariance, which involves a double summation over all internal rank indices. The exponents of the ranks in the final monomial count how many of these summations remain independent after constraints from the interaction set are applied. The chain-like structure ensures that non-contiguous interactions constrain the summations differently than contiguous ones, producing the necessary asymmetry —see Appendix A.3 for detailed derivations. □

4.2.2 Rank Identifiability Analysis

The asymmetry of the moment structure allows us to construct a solvable system for each interior rank.

Theorem 4.6 (Closed-Form Interior-Rank Estimator for TT)**.** *For any interior rank $1 < p < M-1$, r_p is identifiable and given by*

$$r_p = \frac{v_{\{p,p+1\}} \cdot v_{\{p-1,p+2\}}}{v_{\{p+1\}} \cdot v_{\{p-1,p,p+2\}}} \,. \tag{6}$$

Proof. The proof follows by dividing the two identities from Lemma 4.5. The unknown coefficient of variation term, σ_p^2/μ_p^2, cancels, resulting in the rank equation:

$$r_p = \frac{v_{\{p,p+1\}}/v_{\{p+1\}}}{v_{\{p-1,p,p+2\}}/v_{\{p-1,p+2\}}} = \frac{v_{\{p,p+1\}} \cdot v_{\{p-1,p+2\}}}{v_{\{p+1\}} \cdot v_{\{p-1,p,p+2\}}}.$$

Since the rank can be expressed purely in terms of observable moments, it is identifiable. □

Remark 4.7 (The Tensor Ring Model)**.** *A generalization of the TT, known as the Tensor Ring (TR) model, can be formed by starting from the TT and connecting its beginning and end, creating a closed loop with no limitations on the outer rank dimension. We prove the identifiability of such TR model, and present its corresponding rank estimation formulas in Appendix A.4.*

5 THE TENSOR-RANK ESTIMATION PIPELINE

We now present a complete, algorithmically specified, robust pipeline for applying the theoretical results of this work: it consists of estimating from observed tensor-data the pure-interaction terms $\{\hat{v}_S\}$ and the estimated mean $\hat{E}[Y]$, to replace them in the specific closed-form expression derived in Section 4 for the ranks of each identifiable model.

Recall that all rank estimation expressions are written as a division, i.e., the rank r_p is always calculated via an equation of the $\frac{\text{Num}_p}{\text{Den}_p}$ form, with Num_p the numerator and Den_p the denominator; where expressions are adjusted for each specific model.

Our tensor-rank estimation procedure is based on (1) computing an unbiased, single-pass estimation of the aforementioned moments; (2) a bootstrap procedure to provide a distribution of empirical (regularized) rank estimates; and (3) a summarization step to yield robust point and uncertainty-aware rank estimates.

Inner loop: ***The Unbiased Moment Estimation.*** The fundamental step is to compute unbiased estimates of pure interaction terms $\{\hat{v}_S\}$ from the observed tensor-data $\mathcal{Y}$:

1. **The Global Mean:** Compute the empirical global mean of the tensor, $\hat{E}[Y] = |\mathcal{Y}|^{-1} \sum_{\mathbf{i}} Y_{\mathbf{i}}$.
2. **Total Covariances:** For every non-empty pattern of shared indices $S \subseteq \{1, \ldots, M\}$, compute the sample covariance $\hat{C}_S$ by drawing a (large) number of pairs of entries $(\mathbf{a}_j, \mathbf{b}_j)$ that match exactly on the index set S, i.e.,
$$\hat{C}_S = (N_{\text{cov}} - 1)^{-1} \sum_{j=1}^{N_{\text{cov}}} (Y_{\mathbf{a}_j} - \hat{E}[Y])(Y_{\mathbf{b}_j} - \hat{E}[Y]) \ .$$
3. **Pure Interactions:** Calculate the pure interaction terms from the total covariances using the exact inclusion-exclusion principle:
$$\hat{v}_S = \sum_{S' \subseteq S} (-1)^{|S|-|S'|} \hat{C}_{S'} \ .$$

Outer loop: ***The Bootstrapped Distribution of Rank Estimates.*** To ensure robustness and quantify the estimation uncertainty, we repeat the inner moment estimation procedure described above on bootstrap replicates of the observed tensor-data.

For $b = 1, \ldots, B$:

1. A bootstrap sample $\boldsymbol{Y}^{(b)}$ is generated from the original tensor $\boldsymbol{Y}$, using a block bootstrap on tensor slices to preserve potential dependencies.
2. The complete "*Inner loop*" estimation procedure is applied to $\boldsymbol{Y}^{(b)}$ to obtain a set of bootstrap pure term estimates, $\{\hat{v}_S\}^{(b)}$.
3. These bootstrap moments are plugged into the appropriate closed-form formula described in Section 4 to yield the numerator $\text{Num}_p^{(b)}$ and denominator $\text{Den}_p^{(b)}$ for each bootstrapped replicate of the identifiable rank $r_p^{(b)}$:
   - PARAFAC/CP (order M, single rank r): for $p \neq q$, use Equation 3.
   - TT (order M, ranks $r_1, \ldots, r_{M-1}$): for interior ranks $1 < p < M-1$, use Equation 6.
4. Each bootstrap estimate is regularized to prevent numerical instability. For each replicate b, we propose a regularized rank
$$\hat{r}_p^{(b)} = \text{sign}(\text{Den}_p^{(b)}) \cdot \frac{\text{Num}_p^{(b)}}{|\text{Den}_p^{(b)}| + \varepsilon_p}.$$
The shrinkage term $\varepsilon_p = 1.96 \cdot \text{SE}(\{\text{Den}_p^{(j)}\}_{j=1}^B)$ is the data-driven standard error of the denominator, estimated from all bootstrap replicates.

The Final Rank Estimate. We provide point rank estimates with bootstrapped confidence intervals based on the collection of regularized estimates $\{\hat{r}_p^{(b)}\}_{b=1}^B$:

- **Point Estimate:** The median is used for its robustness to outliers, $\hat{r}_p^{\text{med}} = \text{median}\{\hat{r}_p^{(b)}\}$.
- **Confidence Interval:** The $(1-\alpha)$ percentile interval is computed from the bootstrap distribution: $[\hat{r}_p^{\text{low}}, \hat{r}_p^{\text{up}}] = \text{percentile}(\{\hat{r}_p^{(b)}\}, [\alpha/2, 1-\alpha/2])$.

6 Empirical Validation

We validate the theoretical findings for the identifiable models using simulation studies, confirming that the closed-form estimators accurately recover the true latent ranks when applied to finite tensor data.

We present results for the PARAFAC/CP model in Section 6.1, and the empirical studies for TT and TR in Appendix C. In addition, we present in Section 6.2 a comparison of the proposed rank estimation procedure and Bayesian inference alternatives on the real-world PINCAT MRI dataset (Candes et al., 2013)[2].

The codebase with the tensor rank estimation pipeline to replicate the experiments is available in https://github.com/iurteagalab/tensor-ranks.

6.1 Simulations of the PARAFAC/CP model

Generative Process. We generated synthetic tensor-data according to the PARAFAC/CP model, where elements of the latent factors $(\theta^{(p)})$ were drawn independently from Gamma distributions. The shape and scale parameters of the Gamma priors were chosen to control the factor-specific means and variances (μ_p, σ_p^2) while ensuring the resulting rate tensor $\boldsymbol{\eta}$ was strictly positive. The final observed tensor elements $Y_{\mathbf{i}}$ were then drawn from a Poisson distribution with the corresponding rate $\eta_{\mathbf{i}}$ and $Y_{\mathbf{i}} \sim \text{Poisson}(\eta_{\mathbf{i}})$.

Experimental Design. We study the precision of the constructed pipeline by generating PARAFAC/CP model data in a wide range of conditions: we vary the tensor order ($M \in \{3, 4\}$), the tensor dimensions ($N_p \in \{(100, 100, 100), (50, 50, 50, 50)\}$), and the ground-truth latent ranks ($r_p \in \{5, \ldots, 65\}$).

For each case, we aggregate results over multiple independent runs, i.e., realizations of both the data and the latents, using a set of values for the prior described

[2]Link to the dataset `https://candes.su.domains/software/SURE/data.html`

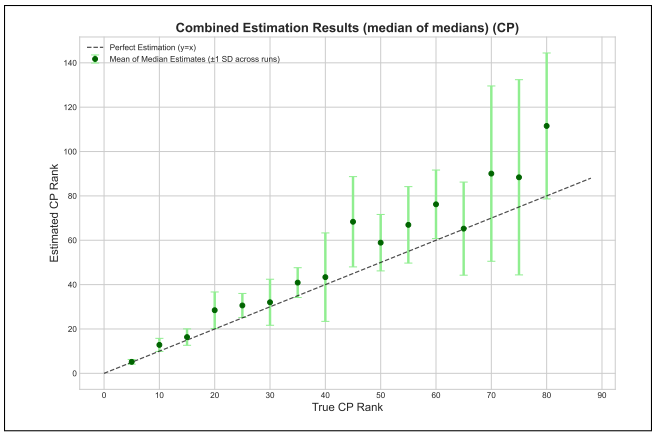

Figure 1: Combined estimation results for the PARAFAC/CP model. The dashed line represents perfect estimation ($y = x$). Green dots show the mean of median estimates across multiple runs, with error bars indicating ± 1 standard deviation.

in Appendix C. For each run, a median rank was estimated from a distribution of bootstrap replicates, and the plots show the mean of these median estimates.

Rank estimation results. We provide strong empirical support of the theoretical identifiability and estimation accuracy of the PARAFAC/CP model's ranks. Figure 1 shows the estimated CP rank versus the true CP rank, aggregated across all simulation runs. The mean of the estimated ranks (green dots) closely tracks the line of perfect estimation. For true ranks ranging from low values up to 90, the estimator correctly identifies the latent dimension and follows the linear trend.

Best performance (both in closer mean and reduced estimation uncertainty) is achieved for reduced (< 25) rank values. We observe a slight positive bias and an increase in estimation variance (error bars) as the true rank grows. This is expected, as higher-rank structures are more complex, and their corresponding moment-based signatures are more sensitive to the statistical noise inherent in finite data sampling. Overall, the results confirm that the estimator is effective and that the CP rank is practically identifiable.

Comparison with AIC/BIC Rank Estimation. To benchmark the efficacy of the proposed prior-predictive (PP) estimator, we evaluate it alongside information-criteria approaches (AIC/BIC) to rank estimation. Notably, the variational CP model coupled with AIC/BIC scoring rules consistently collapsed to the lowest available rank (often $R = 2$) in the grid search of our experiments, across the simulated CP datasets under various prior assumptions (Gamma, Uniform, Gaussian).

In contrast, the PP estimator consistently recovered the scale of the ground-truth rank across non-zero mean scenarios.[3]

6.2 Real-world tensor-data experiment.

We additionally assess rank estimation performance on the real-world PINCAT MRI dataset, by comparing our automatically determined rank ($\hat{r}_{\text{PP}} = 70$) with the rank values obtained by explicit Bayesian inference methods described by Zhang et al. (2022). Specifically, we compare our approach to the ranks obtained via BFCP and Hamiltonian Monte Carlo (HMC) methods using a hold-out of 10% of the data to measure the predictive performance.

Table 1 illustrates that our computationally inexpensive PP estimator matches the performance and uncertainty calibration of full Bayesian procedures.

Table 1: PINCAT MRI rank estimation and predictive performance against Bayesian baselines.

Method	Rank	Test RMSE	Test SD
Prior Predictive (ours)	70	0.6086	0.4876
BFCP (Zhang et al., 2022)	17	0.6087	0.4877
HMC (Zhang et al., 2022)	65	0.6086	0.4876

Robustness to Missing Data. A crucial aspect of real-world tensor analysis is the presence of missing entries. We stress-tested our PP estimator under Missing Completely At Random (MCAR) conditions, applying random masks incrementally (up to 50% data missingness) on the PINCAT dataset.

As the percentage of observed data decreases, the variance of the estimator increases (naturally matching the reduction in valid paired observations, N_{cov}). We observe in the results presented in Table 2 that the estimated rank shrinks gradually due to the reduced effective information: the PP estimator degrades smoothly and predictably, tracing the theoretical finite-sample boundaries.

Table 2: PINCAT MRI predictive and rank estimation performance under varying MCAR missingness.

Observed %	Test RMSE	Estimated Rank ($\hat{r}_{\text{PP}}$)
90%	0.6086	106.9
80%	0.6414	83.0
70%	0.6781	66.4
50%	0.7609	60.9

[3] If all factors have zero-mean priors, all pure interaction terms vanish and hence, tensor ranks are not identifiable from second-order moments.

7 DISCUSSION

We present a prior predictive moment matching analysis framework revealing that rank identifiability in probabilistic tensor models hinges on the latent factors' interaction topology.

More precisely, we demonstrate that linear (additive) structures like the PARAFAC/CP model yield monomial moments with clean algebraic cancellations, enabling closed-form estimators. Chain-like topologies (TT) introduce asymmetry via contiguous/non-contiguous interactions, providing independent equations per rank that also enable their empirical estimation. The cyclic structure in TR maintains solvability through invertible circulant matrices, despite its symmetry.

In contrast, multiplicative "hub-and-spoke" topologies (e.g., Tucker) induce linear dependencies, rendering the reduced system degenerate —we speculate that higher moments or correlation structure in the core tensor could render the models' ranks identifiable.

The presented topology-based lens offers a principled tensor model analysis toolbox: for identifiable models, low-order moments suffice for exact rank recovery; for others, heuristics like matricization (see Appendix A.2) implicitly reduce to CP-like problems.

Limitations of the presented work include the i.i.d. prior assumption —correlated prior-based moments remain log-linearizable, but they require an augmented parameterization, analogous to matrix results by (Zhu and Sato, 2025)— and the focus on second-order moments only —we hypothesize extensions to higher-order moments are possible and would broaden the applicability of the presented analysis.

Empirically, our pipeline is shown to be robust to moderate noise, but sensitive to extreme sparsity and data-missingness, which requires both further numerical considerations and robustness mechanisms. A natural next-step is the application of the proposed framework to other real-data sets (e.g., on fMRI tensors).

More ambitious future work might consider integration of the prior predictive moment matching approach into variational inference for scalable tensor ML, or exploring non-Euclidean topologies (e.g., hyperbolic embeddings).

A natural generalization of our framework is to consider generic tensor factorizations defined by arbitrary tensor networks (Cichocki et al., 2016). In this case, the latent ranks would correspond to the bond dimensions along network edges, and identifiability would depend on the network's topology (e.g., tree-like vs. loopy structures).

Acknowledgements

We would like to thank Caterina de Bacco, Aaron Schein, and Tomasz Kuśmierczyk for their insightful dialog and constructive feedback provided during the early stages of this research.

E. da Silva. and I. Urteaga were supported by "la Caixa" foundation's LCF/BQ/PI22/11910028 award, as well as the Basque Government through the BERC 2022-2025 program and the Ministry of Science and Innovation's BCAM Severo Ochoa accreditation CEX2021-001142-S/MICIN/AEI/10.13039/501100011033.

D. Mesquita acknowledges the support by the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (SEI-260003/020348/2025, SEI-260003/020694/2025) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (404336/2023-0, 305692/2025-9).

E. da Silva's work is funded by Portuguese national funds through FCT – Foundation for Science and Technology, I.P., within the scope of the research unit UID/00326 - Centre for Informatics and Systems of the University of Coimbra.

I. Urteaga also acknowledges the support of Grant RYC2023-045922-I funded by MICIU/AEI/10.13039/501100011033 and by ESF+.

References

O. Alter, P. O. Brown, and D. Botstein. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proceedings of the National Academy of Sciences*, 100(6):3351–3356, 2003.

P. Bhargava, T. Phan, J. Zhou, and J. Lee. Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor factorization on sparse user-generated data. In *WWW*, pages 130–140. ACM, 2015.

R. Bro and H. A. L. Kiers. A new efficient method for determining the number of components in parafac models. *Journal of Chemometrics*, 17(5):274–286, 2003. doi: https://doi.org/10.1002/cem.801.

E. J. Candes, C. A. Sing-Long, and J. D. Trzasko. Unbiased risk estimates for singular value thresholding and spectral estimators. *IEEE transactions on signal processing*, 61(19):4643–4657, 2013.

J. G. Carlsson and F. Jia. Euclidean hub-and-spoke networks. *Operations Research*, 61(6):1360–1382, 2013. doi: 10.1287/opre.2013.1219.

J. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way

generalization of eckart-young decomposition. *Psychometrika*, pages 283–319, 1970.

W. Chu and Z. Ghahramani. Probabilistic models for incomplete multi-dimensional arrays. In *AISTATS*, volume 5 of *JMLR Proceedings*, pages 89–96. JMLR.org, 2009.

A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, and D. P. Mandic. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends® in Machine Learning*, 9(4-5):249–429, 2016. ISSN 1935-8237. doi: 10.1561/2200000059.

F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi. Tensor decomposition of eeg signals: a brief review. *Journal of neuroscience methods*, 248:59–69, 2015.

M. Contisciani, F. Battiston, and C. De Bacco. Inference of hyperedges and overlapping communities in hypergraphs. *Nature communications*, 13(1):7229, 2022.

E. d. S. da Silva, T. Kuśmierczyk, M. Hartmann, and A. Klami. Prior specification for bayesian matrix factorization via prior predictive matching. *Journal of Machine Learning Research*, 24(67):1–51, 2023. URL `http://jmlr.org/papers/v24/21-0623.html`.

P. Gopalan, J. M. Hofman, and D. M. Blei. Scalable recommendation with hierarchical poisson factorization. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, UAI'15, 2015.

R. A. Harshman et al. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA working papers in phonetics*, 16(1):84, 1970.

J. Håstad. Tensor rank is np-complete. *J. Algorithms*, 11(4):644–654, 1990.

C. J. Hillar and L.-H. Lim. Most tensor problems are np-hard. *J. ACM*, 60(6), Nov. 2013. ISSN 0004-5411. doi: 10.1145/2512329. URL `https://doi.org/10.1145/2512329`.

F. L. Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927. doi: https://doi.org/10.1002/sapm192761164.

T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.

R. Mosayebi and G. Hossein-Zadeh. Correlated coupled matrix tensor factorization method for simultaneous eeg-fmri data fusion. *Biomed. Signal Process. Control.*, 62:102071, 2020.

A. Novikov, D. Podoprikhin, A. Osokin, and D. P. Vetrov. Tensorizing neural networks. In *NeurIPS*, pages 442–450, 2015.

I. V. Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

I. Porteous, E. Bart, and M. Welling. Multi-hdp: A non parametric bayesian model for tensor factorization. In *AAAI*, pages 1487–1490. AAAI Press, 2008.

P. Rai, Y. Wang, S. Guo, G. Chen, D. B. Dunson, and L. Carin. Scalable bayesian low-rank decomposition of incomplete multiway tensors. In *ICML*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 1800–1808. JMLR.org, 2014.

A. Schein, J. W. Paisley, D. M. Blei, and H. M. Wallach. Bayesian poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *KDD*, pages 1045–1054. ACM, 2015.

E. M. Stoudenmire and D. J. Schwab. Supervised learning with tensor networks. In *NeurIPS*, NIPS'16, page 4806–4814, 2016.

J. Swernofsky. Tensor rank is hard to approximate. In *APPROX-RANDOM*, volume 116 of *LIPIcs*, pages 26:1–26:9. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

W. Wang and M. Stephens. Empirical bayes matrix factorization. *J. Mach. Learn. Res.*, 22:120:1–120:40, 2021.

K. Zhang, C. Hawkins, and Z. Zhang. General-purpose bayesian tensor learning with automatic rank determination and uncertainty quantification. *Frontiers in Artificial Intelligence*, Volume 4 - 2021, 2022. doi: 10.3389/frai.2021.668353. URL `https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2021.668353`.

Q. Zhao, L. Zhang, and A. Cichocki. Bayesian CP factorization of incomplete tensors with automatic rank determination. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(9):1751–1763, 2015.

Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki. Tensor ring decomposition. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 2280–2284. IEEE, 2016. doi: 10.1109/EUSIPCO.2016.7760655.

Z. Zhu and I. Sato. Prior specification for exposure-based bayesian matrix factorization. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856. URL https://openreview.net/forum?id=o5R4Hv9XqC.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes] We include some proofs in the main text and other in the appendix.
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

On the Identifiability of Tensor Ranks via Prior Predictive Matching: Supplementary Materials

A THEORETICAL DETAILS AND PROOFS

This appendix provides complete, self-contained derivations that justify all claims used in the main text while preserving the main text's numbering and statements. We (i) clarify the definitions of observable covariances and pure interaction terms; (ii) give a general counting lemma that turns index-sharing patterns into monomial exponents; and (iii) supply full proofs for Tucker, TT, and TR.

A.1 Preliminaries: observation model, covariances, and pure terms

Notation. Let $\boldsymbol{Y} \in \mathbb{R}^{N_1 \times \cdots \times N_M}$ be an observed tensor with M modes. A *mode* of an order-M tensor is one of its M dimensions. Let $\mathbf{i} = (i_m)_{m\in[M]} \in \bigtimes_{m=1}^{M}[N_m]$ be the multi-index from the observed tensor domain, where $[N_m] = \{1, \ldots, N_m\}$. Let $\mathbf{r} = (r_p)_{p\in[R]}$ be the multi-rank vector of size $R \leq M$. Let the *complete* multi-index be $(i_1, \ldots, i_R) \in \bigtimes_{p=1}^{R}[r_p]$ for the product of *all* R latent rank domains $[r_p]$, and let $\mathbf{i}_\beta$ be a *derived* multi-index for latent domains $\beta \in \mathcal{K} \subset [R]$ representing a subset $\mathcal{K}$ of all possible latent dimensions. For multi-indices $\mathbf{i}, \mathbf{j}$, the shared-mode set is $S(\mathbf{i}, \mathbf{j}) = \{k : i_k = j_k\}$. We use bold sans-serif for tensors ($\boldsymbol{\eta}$) and vectors/matrices ($\boldsymbol{\theta}^{(p)}$). All latent factor multi-indices $\mathbf{i}_p(\mathbf{i}, \beta)$ are formed by combining some components of the observed tensor indices $\mathbf{i}$ and some components latent domains $\mathbf{i}_\beta$, and each factor is probabilistically defined as $\theta^{(p)}_{\mathbf{i}_p(\mathbf{i},\beta)} \sim \pi_p(\mu_p, \sigma_p^2)$, i.i.d. across indices with $p \in \mathcal{P}$ groups the factors/cores.

Observation model. Throughout, conditional on the rate tensor $\boldsymbol{\eta}$, entries are conditionally independent:

$$\mathbb{E}[Y_{\mathbf{i}} \mid \boldsymbol{\eta}] = \eta_{\mathbf{i}}, \qquad \mathrm{Var}(Y_{\mathbf{i}} \mid \boldsymbol{\eta}) = \phi(\eta_{\mathbf{i}}),$$

for some nonnegative function $\phi(\cdot)$. Hence, for *distinct* entries the *observable cross-covariance* equals the covariance of rates:

$$\mathrm{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}}) = \mathrm{Cov}(\eta_{\mathbf{i}}, \eta_{\mathbf{j}}) \quad \text{if } \mathbf{i} \neq \mathbf{j}.$$

Therefore, all results below are agnostic to the choice of ϕ for cross terms and only use $\mathbb{E}[Y] = \mathbb{E}[\eta]$.

Exact-sharing covariances. For two multi-indices $\mathbf{i}, \mathbf{j} \in \prod_{m=1}^{M}[N_m]$ and a subset $S \subseteq [M]$, we say the pair *shares exactly* S if $i_k = j_k$ for $k \in S$ and (i_k, j_k) are free (not constrained to be equal) for $k \notin S$. The *total observable covariance* for pattern S is

$$C_S \;:=\; \mathbb{E}\big[\mathrm{Cov}\big(Y_{\mathbf{i}}, Y_{\mathbf{j}} \,\big|\, \mathbf{i}, \mathbf{j} \text{ share exactly } S\big)\big]\,,$$

which depends only on S under the iid factor assumptions.

Pure interaction terms and inclusion–exclusion. Define v_S by the Moebius inversion (pure term) over the lattice of subsets:

$$C_S \;=\; \sum_{\emptyset \neq S' \subseteq S} v_{S'}, \qquad v_S \;=\; \sum_{S' \subseteq S} (-1)^{|S|-|S'|}\, C_{S'}. \tag{7}$$

In all derivations we work with *exact-sharing* C_S and pure terms v_S defined by equation 7.

Augmented interaction sets for Tucker. For the Tucker model we sometimes want to distinguish contributions that involve the core tensor G versus factor variances. For clarity, we allow *augmented* symbols such as $v_{\{G\}}$ and $v_{\{G,p\}}$ to denote the pure contribution that is proportional to σ_G^2 (respectively to $\sigma_G^2\sigma_p^2$) when *all physical modes are shared*. This does not change the observable definition in equation 7; it is a bookkeeping device that uniquely decomposes $C_{[M]}$ by variance sources.

A.1.1 A counting lemma for rank exponents

All models considered can be written as

$$\eta_{\mathbf{i}} \;=\; \sum_{\beta\in\mathcal{K}} \prod_{p\in\mathcal{P}} \theta^{(p)}_{\mathrm{i}_p(\mathbf{i},\beta)},$$

where β collects all *latent* indices (their cardinalities are products of the ranks $\{r\}$), and the product ranges over parameter groups p (factors/cores), each drawn iid with $\mathbb{E}[\theta^{(p)}] = \mu_p$ and $\mathrm{Var}(\theta^{(p)}) = \sigma_p^2$.

Lemma A.1 (Exponent counting)**.** *Fix a model topology (CP/TT/TR/Tucker), a sharing pattern $S \subseteq [M]$, and consider the second moment $\mathbb{E}[\eta_{\mathbf{i}}\eta_{\mathbf{j}}]$ with $(\mathbf{i},\mathbf{j})$ constrained to share exactly S. Write it as a double sum over two latent index tuples $(\beta,\beta') \in \mathcal{K}\times\mathcal{K}$. Then:*

1. *For each parameter group p, the factor $\theta^{(p)}$ contributes either a mean-square term μ_p^2 (if the two indices of $\theta^{(p)}$ are independent across (β,β') or across $(\mathbf{i},\mathbf{j})$) or a variance term σ_p^2 (if they coincide).*

2. *Each latent* link *(summation index) contributes an exponent of r^α where $\alpha \in \{0,1,2\}$ equals the number of* independent *surviving sums across (β,β') for that link after enforcing the constraints from S and from which factors are in variance mode.*

Consequently, every pure term v_S is a monomial *in $\{r\} \cup \{\mu_p, \sigma_p^2\}$ for CP/TT/TR; for Tucker, $v_{\{G\}}$, $v_{\{p\}}$, and $v_{\{G,p\}}$ are monomials when all physical modes are shared.*

Proof. Expand $\mathbb{E}[\eta_{\mathbf{i}}\eta_{\mathbf{j}}]$ into a double sum over latent indices for the two entries. Independence across different latent tuples makes the expectation factor across parameter groups p. Whether the same random variable appears twice (giving σ_p^2) or two independent copies (giving μ_p^2) is determined by index equalities induced by S and by whether the two latent index tuples coincide on the arguments of $\theta^{(p)}$. The number of independent sums that remain for each latent link determines the power of the corresponding rank. Subtracting lower-order $C_{S'}$ by inclusion–exclusion isolates the pure term v_S, which retains a single monomial. □

A.2 Theoretical analysis of the Tucker model

We work with the standard order-M Tucker rate

$$\eta_{i_1\ldots i_M} \;=\; \sum_{k_1=1}^{r_1}\cdots\sum_{k_M=1}^{r_M} G_{k_1\ldots k_M} \prod_{p=1}^{M} \theta^{(p)}_{i_p,k_p},$$

with iid G_k having mean μ_G and variance σ_G^2, and iid factor entries $\theta^{(p)}$ having moments (μ_p,σ_p^2).

Global mean. By independence,

$$\mathbb{E}[\eta_{i_1\ldots i_M}] = \sum_{k_1,\ldots,k_M} \mu_G \prod_{p=1}^{M}\mu_p = \Big(\prod_{p=1}^{M} r_p\Big)\,\mu_G \prod_{p=1}^{M}\mu_p.$$

Thus,

$$(E[Y])^2 \;=\; \Big(\prod_{p=1}^{M} r_p\Big)^2 \mu_G^2 \prod_{p=1}^{M}\mu_p^2. \tag{8}$$

Pure terms when all modes are shared. Consider $C_{[M]} = \mathrm{Cov}(Y_{\mathbf{i}}, Y_{\mathbf{j}})$ for pairs sharing all M physical indices (this is the only case in which core-variance and factor-variance mix coherently in Tucker). Decomposing by the variance sources we obtain four monomials:

Lemma A.2 (Tucker monomials for shared-all-modes)**.** *With the above conventions, the following hold:*

$$v_{\{G\}} = \sigma_G^2 \prod_{p=1}^{M} (r_p^2 \mu_p^2), \tag{9}$$

$$v_{\{p\}} = \mu_G^2 \left(r_p \sigma_p^2\right) \prod_{q \neq p} (r_q^2 \mu_q^2), \tag{10}$$

$$v_{\{G,p\}} = \sigma_G^2 \left(r_p \sigma_p^2\right) \prod_{q \neq p} (r_q^2 \mu_q^2), \tag{11}$$

and equation 8 for $(E[Y])^2$ *remains unchanged.*

Proof. Expand $\mathbb{E}[\eta_{\mathbf{i}} \eta_{\mathbf{j}}]$ as a double sum over the core indices (k, k') and factor indices for two copies. For the pure term $v_{\{G\}}$, variance arises from G with $k = k'$ (contributing σ_G^2). All factor groups are in mean mode and appear in *both* copies, leaving two independent summations per factor index; each contributes $r_q^2 \mu_q^2$. Thus $v_{\{G\}} = \sigma_G^2 \prod_q (r_q^2 \mu_q^2)$.

For $v_{\{p\}}$, variance arises from the single factor at mode p (giving $r_p \sigma_p^2$), while the core and all other factors contribute means in both copies: hence μ_G^2 and, for $q \neq p$, $r_q^2 \mu_q^2$. This yields equation 10.

For $v_{\{G,p\}}$, combine the two variance sources: G in variance and factor p in variance; the remaining factors contribute mean-squared terms across the two copies, producing equation 11. □

A Tucker identity and its consequence. Using equation 8–equation 11, we obtain

$$\begin{aligned} \log v_{\{p\}} - 2\log(E[Y]) &= \log\left[\mu_G^2 (r_p \sigma_p^2) \prod_{q \neq p} (r_q^2 \mu_q^2)\right] - 2\log\left[\left(\prod_q r_q\right) \mu_G \prod_q \mu_q\right] \\ &= \log\left(\frac{\sigma_p^2}{r_p \mu_p^2}\right), \end{aligned} \tag{12}$$

and similarly

$$\begin{aligned} \log v_{\{G,p\}} - \log v_{\{G\}} &= \log\left[\sigma_G^2 (r_p \sigma_p^2) \prod_{q \neq p} (r_q^2 \mu_q^2)\right] - \log\left[\sigma_G^2 \prod_q (r_q^2 \mu_q^2)\right] \\ &= \log\left(\frac{\sigma_p^2}{r_p \mu_p^2}\right). \end{aligned} \tag{13}$$

Hence we have the exact identity

$$\boxed{\log v_{\{p\}} - 2\log(E[Y]) = \log v_{\{G,p\}} - \log v_{\{G\}}} \tag{14}$$

which is the clarified version of Eq. (3.1) in the main text (Lemma 3.2/Lemma 3.3 there).

Theorem A.3 (Non-identifiability of Tucker ranks: full proof)**.** *Let* $\mathbf{Ax} = \mathbf{b}$ *be the log-linear system built from the monomials for* $(E[Y])^2$*,* $v_{\{G\}}$*,* $v_{\{p\}}$*,* $v_{\{G,p\}}$ *with* $p = 1, \ldots, M$*. Then the* M *columns corresponding to* $\log r_p$ *lie in the span of the columns that do not involve* $\log r_p$*; equivalently, the reduced design matrix for* $\mathbf{x}_r = (\log r_1, \ldots, \log r_M)$ *is the zero matrix. Thus the ranks* $(r_1, \ldots, r_M)$ *are not identifiable from second moments.*

Proof. Each identity equation 14 eliminates $\log r_p$ when forming the linear combination of rows

$$(\log v_{\{p\}}) - 2\log(E[Y]) - \left[(\log v_{\{G,p\}}) - \log v_{\{G\}}\right] = 0.$$

Hence the coordinate vector that picks $\log r_p$ is a combination of columns that correspond only to nuisance parameters (log-means and log-variances). Doing this for each p shows the reduced matrix has zero columns. □

A.3 Theoretical analysis of the Tensor Train model

This appendix provides complete, self-contained derivations for the TT model. We (i) formalize the monomial structure of pure terms v_S for arbitrary exact-sharing patterns S, (ii) derive the two key identities used to identify *interior* ranks, and (iii) provide boundary-mode analogues and closed-form estimators for *boundary* ranks.

Model. For an order-M TT, the rate is

$$\eta_{i_1 \dots i_M} \;=\; \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} \theta^{(1)}_{1,i_1,k_1}\, \theta^{(2)}_{k_1,i_2,k_2} \cdots \theta^{(M)}_{k_{M-1},i_M,1},$$

where entries of core p have moments (μ_p, σ_p^2) and are iid across indices. Exact-sharing covariances C_S and pure terms v_S follow Appendix A.1.

A.3.1 A general monomial formula for TT pure terms

Proposition A.4 (TT pure terms are monomials with explicit rank exponents)**.** *Let $S \subseteq \{1, \dots, M\}$ be an exact-sharing pattern. For each TT link $\ell \in \{1, \dots, M-1\}$, define*

$$\alpha_\ell(S) \;:=\; 2 - \mathbf{1}\{\ell \in S \text{ or } \ell+1 \in S\} \;\in \{1,2\}.$$

Then the pure term associated with S is the single *monomial*

$$\boxed{v_S \;=\; \Big(\prod_{\ell=1}^{M-1} r_\ell^{\alpha_\ell(S)}\Big)\Big(\prod_{p\in S} \sigma_p^2\Big)\Big(\prod_{p\notin S} \mu_p^2\Big).} \tag{15}$$

Proof. Expand $\mathbb{E}[\eta_{\mathbf{i}}\eta_{\mathbf{j}}]$ as a double sum over latent tuples $(k_1, \dots, k_{M-1})$ and $(k'_1, \dots, k'_{M-1})$. For a fixed S, modes in S contribute variance (σ_p^2), modes not in S contribute squared means (μ_p^2) after subtracting $\mathbb{E}[\eta_{\mathbf{i}}]\,\mathbb{E}[\eta_{\mathbf{j}}]$. A TT *link* ℓ is the pair (k_ℓ, k'_ℓ). If neither adjacent physical mode ℓ nor $\ell+1$ lies in S, the two copies keep independent sums over k_ℓ and k'_ℓ (yielding r_ℓ^2). If at least one of ℓ or $\ell+1$ lies in S, equality constraints collapse two sums into one (yielding r_ℓ). This gives the exponent $\alpha_\ell(S)$. Moebius inversion over subsets $T \subseteq S$ cancels all mixed patterns, leaving the single monomial equation 15. □

Consequences. Proposition A.4 immediately yields a catalogue of pure terms:

$$\text{Singleton } S=\{p\}: \quad v_{\{p\}} = \Big(\prod_{\ell\neq p-1,p} r_\ell^2\Big)\, r_{p-1}\, r_p \;\; \sigma_p^2 \prod_{k\neq p} \mu_k^2.$$

$$\text{Adjacent pair } S=\{p,p+1\}: \quad v_{\{p,p+1\}} = \Big(\prod_{\ell\notin\{p-1,p,p+1\}} r_\ell^2\Big)\, r_{p-1}\, r_p\, r_{p+1} \;\; \sigma_p^2\sigma_{p+1}^2 \prod_{k\notin\{p,p+1\}} \mu_k^2.$$

$$\text{Noncontiguous pair } S=\{p-1,p+2\}: \quad v_{\{p-1,p+2\}} = \Big(\prod_{\ell\notin\{p-2,p-1,p,p+1,p+2\}} r_\ell^2\Big)\, r_{p-2}\, r_{p-1}\, r_p^{\mathbf{2}}\, r_{p+1}\, r_{p+2} \times \sigma_{p-1}^2\sigma_{p+2}^2 \prod_{k\notin\{p-1,p+2\}} \mu_k^2.$$

$$\text{Triple } S=\{p-1,p,p+2\}: \quad v_{\{p-1,p,p+2\}} = \Big(\prod_{\ell\notin\{p-2,p-1,p,p+1,p+2\}} r_\ell^2\Big)\, r_{p-2}\, r_{p-1}\, r_p^{\mathbf{1}}\, r_{p+1}\, r_{p+2} \times \sigma_{p-1}^2\sigma_p^2\sigma_{p+2}^2 \prod_{k\notin\{p-1,p,p+2\}} \mu_k^2.$$

The highlighted exponents of r_p differ by one (2 vs. 1), which drives the $1/r_p$ factor used below.

A.3.2 Interior-mode identities and estimator

Lemma A.5 (Interior identities)**.** *For any interior mode p with $1 < p < M$:*

$$\frac{v_{\{p,p+1\}}}{v_{\{p+1\}}} = \frac{\sigma_p^2}{\mu_p^2}, \tag{16}$$

$$\frac{v_{\{p-1,p,p+2\}}}{v_{\{p-1,p+2\}}} = \frac{1}{r_p}\,\frac{\sigma_p^2}{\mu_p^2}. \tag{17}$$

Proof. Using equation 15: In equation 16, the link exponents coincide for every ℓ (both sets include $p+1$), so the rank parts cancel; the moment parts differ only at mode p, giving σ_p^2/μ_p^2. In equation 17, the moment parts differ only at p, again giving σ_p^2/μ_p^2, while the rank parts differ by exactly one power of r_p because $\alpha_p(\{p-1,p+2\}) = 2$ (neither p nor $p+1$ is in the set) but $\alpha_p(\{p-1,p,p+2\}) = 1$ (now $p \in S$). □

Theorem A.6 (Closed-form estimator for interior TT ranks)**.** *For $1 < p < M-1$,*

$$\boxed{r_p \;=\; \frac{v_{\{p,p+1\}}\, v_{\{p-1,p+2\}}}{v_{\{p+1\}}\, v_{\{p-1,p,p+2\}}}.} \tag{18}$$

Proof. Divide equation 16 by equation 17. □

A.3.2 Detailed TT moment calculations (interior and boundary cases)

This section expands Proposition A.4 for patterns used in practice and covers boundary modes $p = 1$ and $p = M-1$.

Interior patterns (full monomials) For $1 < p < M-1$:

$$\begin{aligned}
v_{\{p\}} &= \Big(\prod_{\ell\notin\{p-1,p\}} r_\ell^2\Big)\, r_{p-1}\, r_p \;\; \sigma_p^2 \prod_{k\neq p} \mu_k^2,\\
v_{\{p+1\}} &= \Big(\prod_{\ell\notin\{p,p+1\}} r_\ell^2\Big)\, r_p\, r_{p+1} \;\; \sigma_{p+1}^2 \prod_{k\neq p+1} \mu_k^2,\\
v_{\{p,p+1\}} &= \Big(\prod_{\ell\notin\{p-1,p,p+1\}} r_\ell^2\Big)\, r_{p-1}\, r_p\, r_{p+1} \;\; \sigma_p^2\sigma_{p+1}^2 \prod_{k\notin\{p,p+1\}} \mu_k^2,\\
v_{\{p-1,p+2\}} &= \Big(\prod_{\ell\notin\{p-2,p-1,p,p+1,p+2\}} r_\ell^2\Big)\, r_{p-2}\, r_{p-1}\, r_p^2\, r_{p+1}\, r_{p+2} \;\; \sigma_{p-1}^2\sigma_{p+2}^2 \prod_{k\notin\{p-1,p+2\}} \mu_k^2,\\
v_{\{p-1,p,p+2\}} &= \Big(\prod_{\ell\notin\{p-2,p-1,p,p+1,p+2\}} r_\ell^2\Big)\, r_{p-2}\, r_{p-1}\, r_p^1\, r_{p+1}\, r_{p+2} \;\; \sigma_{p-1}^2\sigma_p^2\sigma_{p+2}^2 \prod_{k\notin\{p-1,p,p+2\}} \mu_k^2.
\end{aligned}$$

From these, Lemma A.5 follows by cancellation.

Boundary patterns and identities

At the chain ends, the link r_1 sits between modes 1 and 2; the link r_{M-1} sits between $M-1$ and M. We need two identities per boundary rank: one giving σ_p^2/μ_p^2, and a second giving $(\sigma_p^2/\mu_p^2)/r_p$.

Left boundary ($p = 1$). For $M \geq 3$,

$$\frac{v_{\{1,2\}}}{v_{\{2\}}} = \frac{\sigma_1^2}{\mu_1^2}, \tag{19}$$

$$\frac{v_{\{1,3\}}}{v_{\{3\}}} = \frac{1}{r_1}\,\frac{\sigma_1^2}{\mu_1^2}. \tag{20}$$

Proof. The first ratio mirrors equation 16 with $p = 1$. For the second, note that including mode 3 constrains link r_2 in both numerator and denominator, while including mode 1 (only in the numerator) additionally constrains r_1, changing its exponent from 2 to 1. Moments differ only at mode 1. Hence the ratio equals $(\sigma_1^2/\mu_1^2)/r_1$.

Right boundary ($p = M - 1$). For $M \geq 3$,

$$\frac{v_{\{M-1,M\}}}{v_{\{M\}}} = \frac{\sigma_{M-1}^2}{\mu_{M-1}^2}, \tag{21}$$

$$\frac{v_{\{M-3,M-1\}}}{v_{\{M-3\}}} = \frac{1}{r_{M-1}} \frac{\sigma_{M-1}^2}{\mu_{M-1}^2}. \tag{22}$$

Proof. Symmetric to the left boundary with the appropriate shift (now link r_{M-1} is constrained iff mode $M-1$ or M is in S). The pair $\{M-3\}$ vs. $\{M-3, M-1\}$ plays the role of $\{3\}$ vs. $\{1, 3\}$.

Theorem A.7 (Closed-form estimators for boundary TT ranks)**.** *For $M \geq 3$, the boundary ranks are identifiable via*

$$\boxed{r_1 \;=\; \frac{v_{\{1,2\}}\, v_{\{3\}}}{v_{\{2\}}\, v_{\{1,3\}}}, \qquad r_{M-1} \;=\; \frac{v_{\{M-1,M\}}\, v_{\{M-3\}}}{v_{\{M\}}\, v_{\{M-3,M-1\}}}.} \tag{23}$$

Proof. Divide equation 19 by equation 20 (left boundary) and equation 21 by equation 22 (right boundary). □

Edge-case notes. (i) For $M = 3$, the identities remain valid with the natural interpretation of sets (e.g., $\{3\}$ and $\{1, 3\}$ are defined). (ii) For $M = 2$, TT reduces to a matrix factorization with a single rank and the CP formula applies.

Worked example (rank exponents table) For illustration, consider $M = 6$ and $p = 3$. The link exponents $\alpha_\ell(S)$ from Proposition A.4 are:

S	α_1	α_2	α_3	α_4	α_5
$\{3\}$	2	1	1	2	2
$\{4\}$	2	2	1	1	2
$\{3, 4\}$	2	1	1	1	2
$\{2, 5\}$	1	1	2	1	1
$\{2, 3, 5\}$	1	1	1	1	1

The highlighted entries show the drop $2 \to 1$ at link 3 when moving from $\{2, 5\}$ to $\{2, 3, 5\}$, producing the $1/r_3$ factor.

This completes the TT analysis with full monomial formulas, interior and boundary identities, and closed-form estimators for all TT ranks.

A.4 Theoretical analysis of the Tensor Ring model

We analyze the Tensor Ring (TR) model in full detail. Despite its cyclic symmetry, the first/second moments determine all ranks via a short linear system.

A.4.1 Model and notation

Definition A.8 (Tensor Ring (TR))**.** *For an order-M tensor with TR ranks $(r_1, \ldots, r_M)$ (indices modulo M), the rate is*

$$\eta_{i_1 \ldots i_M} \;=\; \mathrm{Tr}\Big(\Theta_{i_1}^{(1)} \Theta_{i_2}^{(2)} \cdots \Theta_{i_M}^{(M)}\Big),$$

where $\Theta_{i_p}^{(p)} \in \mathbb{R}^{r_{p-1} \times r_p}$ has i.i.d. entries with moments (μ_p, σ_p^2), and $r_0 \equiv r_M$ closes the ring. Conditional on $\boldsymbol{\eta}$, the $Y_{\mathbf{i}}$ are independent with $\mathbb{E}[Y_{\mathbf{i}} \mid \boldsymbol{\eta}] = \eta_{\mathbf{i}}$.

Let the TR *latent indices* be $s_0, s_1, \ldots, s_{M-1}$ with $s_0 = s_M$ and $s_p \in [r_p]$. Expanding the trace gives

$$\eta_{\mathbf{i}} = \sum_{s_0, \ldots, s_{M-1}} \prod_{p=1}^{M} \theta_{s_{p-1},\, i_p,\, s_p}^{(p)}, \qquad s_0 = s_M, \tag{24}$$

where $\theta^{(p)}$ denotes an entry of $\Theta^{(p)}$. As in Appendix A.1, for exact-sharing $S \subseteq [M]$ we work with $C_S = \mathrm{Cov}(\eta_{\mathbf{i}}, \eta_{\mathbf{j}})$ and the pure terms v_S via inclusion–exclusion.

A.4.2 TR moments: monomials and rank exponents

We first obtain the three families of monomials used throughout: $(E[Y])^2$, singletons $v_{\{p\}}$, and adjacent pairs $v_{\{p,p+1\}}$.

Lemma A.9 (TR prior predictive moments)**.** *For any $M \geq 3$ (indices modulo M), the following hold:*

$$(E[Y])^2 = \Big(\prod_{k=1}^{M} r_k \mu_k\Big)^2, \tag{25}$$

$$v_{\{p\}} = \Big(\prod_{k \notin \{p-1,p\}} r_k^2\Big)\, r_{p-1}\, r_p^2 \;\; \sigma_p^2 \prod_{k \neq p} \mu_k^2, \tag{26}$$

$$v_{\{p,p+1\}} = \Big(\prod_{k=1}^{M} r_k\Big) \;\; \sigma_p^2 \sigma_{p+1}^2 \prod_{k \notin \{p,p+1\}} \mu_k^2. \tag{27}$$

Proof. Mean. From equation 24, $\mathbb{E}[\eta_{\mathbf{i}}] = \sum_s \prod_p \mu_p = \prod_p (r_p \mu_p)$, hence equation 25.

Singleton $v_{\{p\}}$. Consider $\mathrm{Cov}(\eta_{\mathbf{i}}, \eta_{\mathbf{j}})$ for pairs sharing all modes and isolate the pure contribution where only $\Theta^{(p)}$ is in variance (Appendix A.1). Expanding $\mathbb{E}[\eta_{\mathbf{i}}\eta_{\mathbf{j}}]$ yields a double sum over latent chains $(s_0, \ldots, s_{M-1})$ and $(t_0, \ldots, t_{M-1})$. Modes $k \neq p$ contribute μ_k^2; mode p contributes σ_p^2. The rank exponents count *independent* sums per link:

- Links not adjacent to p (i.e., $k \notin \{p-1, p\}$) are unconstrained in both copies $\Rightarrow$ 2 independent sums $\Rightarrow r_k^2$.
- The link $p-1$ is adjacent to the sole variance and becomes *constrained across copies* by equality of the index that flows into $\Theta^{(p)}$; only one joint sum remains $\Rightarrow r_{p-1}$.
- The link p (outgoing from $\Theta^{(p)}$) remains *unconstrained* across copies because the variance identification at p does not couple the outgoing indices in the two copies; this keeps two independent sums $\Rightarrow r_p^2$.

This gives equation 26.

Adjacent pair $v_{\{p,p+1\}}$. Now only $\Theta^{(p)}$ and $\Theta^{(p+1)}$ are in variance, all other modes in mean. In the ring, placing two consecutive variances *glues the two copies into a single cyclic loop*: every link becomes constrained across the copies, leaving exactly one independent sum per link. Hence each r_k appears to the first power, and the moment factor is $\sigma_p^2 \sigma_{p+1}^2 \prod_{k \notin \{p,p+1\}} \mu_k^2$, yielding equation 27. □

Discussion. The asymmetric exponents in equation 26 (power 1 on r_{p-1}, but power 2 on r_p) are a genuine ring effect: the variance at node p constrains the *incoming* link (shared across copies) but leaves the *outgoing* link with independent sums. In contrast, two adjacent variances collapse the two-copy ring globally, forcing single powers for *all* links in equation 27.

A.4.3 Rank–ratio identity and identifiability

Define the observable statistic

$$\xi_p \;:=\; \frac{(E[Y])^2 \; v_{\{p,p+1\}}}{v_{\{p\}} \; v_{\{p+1\}}}. \tag{28}$$

Proposition A.10 (Rank–ratio identity)**.** *For all p (indices modulo M),*

$$\boxed{\xi_p \;=\; r_p \, \frac{r_{p-1}}{r_{p+1}}.} \tag{29}$$

Proof. Substitute equation 25–equation 27 into equation 28. All μ and σ factors cancel. For ranks, the numerator contributes $\big(\prod_k r_k\big) \cdot \big(\prod_k r_k\big)$, while the denominator contributes $(\prod_{k \notin \{p-1,p\}} r_k^2) r_{p-1} r_p^2 \cdot (\prod_{k \notin \{p,p+1\}} r_k^2) r_p r_{p+1}^2$. After cancellation, only the factor $r_p r_{p-1} / r_{p+1}$ remains. □

Let $x_p = \log r_p$ and $\psi_p = \log \xi_p$. Taking logs in equation 29 gives the *circulant* system

$$x_p + x_{p-1} - x_{p+1} \;=\; \psi_p, \qquad p = 1, \ldots, M, \tag{30}$$

i.e., $\mathbf{C}\mathbf{x} = \boldsymbol{\psi}$ with the $M \times M$ circulant matrix whose first row is $(1, -1, 0, \ldots, 0, 1)$:

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 1 \\ 1 & 1 & -1 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ -1 & 0 & 0 & \cdots & 1 & 1 \end{bmatrix}. \tag{31}$$

Theorem A.11 (TR identifiability via DFT)**.** *For every integer $M \geq 3$, the matrix $\mathbf{C}$ in equation 31 is invertible. Hence $\mathbf{x} = \mathbf{C}^{-1}\boldsymbol{\psi}$ is uniquely determined from moments, and $r_p = \exp(x_p)$ are identifiable for all p.*

Proof. Let $\omega = e^{2\pi i/M}$. The eigenpairs of a circulant are obtained by the DFT: the kth eigenvalue is

$$\lambda_k \;=\; 1 + \omega^k - \omega^{-k} \;=\; 1 + 2i\,\sin(2\pi k/M), \qquad k = 0, \ldots, M-1.$$

For $M \geq 3$, $\sin(2\pi k/M) = 0$ iff $k \in \{0, M\}$; in either case $\lambda_k = 1 \neq 0$. Thus no λ_k vanishes and $\det \mathbf{C} \neq 0$. Therefore equation 30 has a unique solution $\mathbf{x}$. Exponentiating gives the ranks. □

Computation. Practically, solve $\mathbf{C}\mathbf{x} = \boldsymbol{\psi}$ by FFTs: compute $\widehat{\psi} = \mathrm{DFT}(\psi)$, divide componentwise by λ_k, and invert the DFT to obtain x.

A.4.4 Detailed TR moment calculations and edge cases

We briefly expand the counting in Lemma A.9.

Singleton $v_{\{p\}}$. Write the two copies of the ring with latent chains $(s_0, \ldots, s_M)$ and $(t_0, \ldots, t_M)$ (with $s_0 = s_M$, $t_0 = t_M$). The variance at p identifies the *same* entry of $\Theta^{(p)}$ across copies, which forces $s_{p-1} = t_{p-1}$ but leaves s_p and t_p independent. Propagating around the cycle:

- For links far from p (neither endpoint adjacent to p), both copies retain independent sums $\Rightarrow r_k^2$.
- At link $p-1$ the identification $s_{p-1} = t_{p-1}$ leaves one sum $\Rightarrow r_{p-1}$.
- At link p the two outgoing indices remain independent across copies $\Rightarrow r_p^2$.

Moment factors follow since only mode p contributes σ_p^2; all others contribute μ_k^2. This reproduces equation 26.

Adjacent pair $v_{\{p,p+1\}}$. Variance at p and $p+1$ identifies $\Theta^{(p)}$ and $\Theta^{(p+1)}$ across copies. In the ring, this creates a *single* joint cycle: starting from any link, equality constraints propagate all the way around, leaving exactly one independent sum per link $\Rightarrow \prod_k r_k$. The moment factor is $\sigma_p^2 \sigma_{p+1}^2 \prod_{k \notin \{p,p+1\}} \mu_k^2$, giving equation 27.

Small-M edge cases.

- $M = 2$: the TR degenerates into a matrix factorization with a single rank parameter; the CP formulas apply.
- $M = 3$: all formulas above hold; $\mathbf{C}$ has eigenvalues 1, $1 \pm i\sqrt{3}$ (all nonzero), so identifiability still holds.

Robustness to minor model variants. If the entries of $\Theta^{(p)}$ are centered ($\mu_p = 0$), the above uses $(E[Y])^2$ only via $\prod \mu_k$, which would be zero; in that case one can replace $(E[Y])^2$ with $\mathrm{Var}(\eta)$ and form analogous ratios using higher moments. Our main identifiability result relies on $\mu_p \neq 0$, as assumed in the main text.

B Computational and Statistical Properties of the Estimators

B.1 Covariances, interaction sets, and pure terms

We estimate ranks in two steps.

1. **Observable covariances.** For a sharing set $S \subseteq \{1, \dots, M\}$, the exact–sharing covariance $\widehat{C}_S$ is computed from pairs $(Y_{\mathbf{i}}, Y_{\mathbf{j}})$ with indices equal on S and different on the complement.

2. **Pure terms.** The pure interaction $\widehat{v}_S$ is obtained from $\{\widehat{C}_{S'} : S' \subseteq S\}$ via Möbius inversion:

$$\widehat{v}_S \;=\; \sum_{S' \subseteq S} (-1)^{|S|-|S'|}\, \widehat{C}_{S'}.$$

   These $\widehat{v}_S$ are the inputs to the closed-form estimators in CP/TT/TR.

CP (PARAFAC). With i.i.d. factor priors, $v_S = R \prod_{p \in S} \sigma_p^2 \prod_{p \notin S} \mu_p^2$ for any nonempty S. Hence

$$\widehat{R}_S \;=\; \frac{\widehat{v}_S}{\prod_{p \in S} \sigma_p^2 \prod_{p \notin S} \mu_p^2}, \quad \text{aggregate over } S \in \{\{p\}\} \text{ and optionally } \{p, p{+}1\}.$$

Pure terms needed:

$$\widehat{v}_{\{p\}} = \widehat{C}_{\{p\}}, \qquad \widehat{v}_{\{i,j\}} = \widehat{C}_{\{i,j\}} - \widehat{C}_{\{i\}} - \widehat{C}_{\{j\}}.$$

Tensor Train (TT). Let $r_1, \dots, r_{M-1}$ be bond ranks (open chain). Using the monomial formula for pure terms (Appendix A.3), the closed forms are:

$$\begin{aligned}
\text{Interior } (1 < p < M-1): \quad r_p &= \frac{v_{\{p,p+1\}}\; v_{\{p-1,p+2\}}}{v_{\{p+1\}}\; v_{\{p-1,p,p+2\}}}. \\
\text{Left boundary } (p = 1,\; M \geq 3): \quad r_1 &= \frac{v_{\{1,2\}}\; v_{\{3\}}}{v_{\{2\}}\; v_{\{1,3\}}}. \\
\text{Right boundary } (p = M-1,\; M \geq 4): \quad r_{M-1} &= \frac{v_{\{M-1,M\}}\; v_{\{M-3\}}}{v_{\{M\}}\; v_{\{M-3,M-1\}}}.
\end{aligned}$$

Special cases use triplets directly:

$$M{=}3: \; r_1 = \tfrac{v_{\{1,2\}} v_{\{3\}}}{v_{\{2\}} v_{\{1,3\}}}, \quad r_2 = \tfrac{v_{\{2,3\}} v_{\{1\}}}{v_{\{3\}} v_{\{1,2\}}}; \qquad M{=}4: \; \begin{cases} r_1 = \dfrac{v_{\{1,2\}}}{v_{\{2\}}} \cdot \dfrac{v_{\{3,4\}}}{v_{\{1,3,4\}}}, \\ r_2 = \dfrac{v_{\{2,3\}}}{v_{\{3\}}} \cdot \dfrac{v_{\{1,4\}}}{v_{\{1,2,4\}}}, \\ r_3 = \dfrac{v_{\{3,4\}}}{v_{\{2\}}} \cdot \dfrac{v_{\{1,2\}}}{v_{\{1,3,4\}}}. \end{cases}$$

Required exact–sharing sets are: all singletons; adjacent pairs; the gap-two pair $\{p{-}1, p{+}2\}$; the triplet $\{p{-}1, p, p{+}2\}$; and the boundary helpers $\{1, 3\}$, $\{M{-}3\}$, $\{M{-}3, M{-}1\}$. For $M \in \{3, 4\}$, include $\{1, 2, 3\}$ and $\{1, 3, 4\}$, $\{1, 2, 4\}$ as needed.

Tensor Ring (TR). Let $\psi_p = \log(E[Y]^2) + \log v_{\{p,p+1\}} - \log v_{\{p\}} - \log v_{\{p+1\}}$ (indices mod M). Then the circulant system $C \log r = \psi$ recovers all r_p. Only singletons and adjacent pairs are required.

Sampling kernel (all models). For each required set S, draw S_{cov} pairs per bootstrap with exact sharing pattern S, compute $\widehat{C}_S$, and obtain $\widehat{v}_S$ by Möbius inversion over the (small) downward-closed family that contains S.

B.1.1 Complexity Analysis (Simplified).

Let $\mathcal{S}$ be the number of distinct sharing sets used, B bootstraps, and S_{cov} pairs per set. Pair generation and access are $O(1)$ amortized; Möbius is $O(1)$ per set since $|S| \leq 3$ here. The dominant cost is sampling:

$$\boxed{\text{Time} \;\approx\; \Theta(\mathcal{S}\, B\, S_{\text{cov}}), \qquad \text{Memory} \;\approx\; \Theta(\mathcal{S}\, B) \text{ (store-all) or } \Theta(\mathcal{S}) \text{ (stream).}}$$

Model-specific $\mathcal{S}$ (to estimate all ranks):

CP: $\mathcal{S} = M$ (singletons) or $2M$ (+ adjacent pairs).
TT: $\mathcal{S} = C(M-1)$, with small constant C (singletons, adjacents, gap-two, triplets).
TR: $\mathcal{S} = 2M$ (singletons + adjacent pairs) + one tiny linear solve $O(M \log M)$.

Thus all three scale *linearly* in M via $\mathcal{S}$; TR has the cheapest set family, TT adds a constant-factor overhead for gap-two/triplets, and CP is minimal with singletons.

B.2 Variance and SNR of the Pure-Term Estimators

Among the family of moment-based estimators derived in this work, the *pure-term estimator* for the Tensor Train (TT) structure is the only one that relies *exclusively on second-order moments*— specifically, on covariances between entries sharing particular subsets of tensor modes. Unlike estimators that exploit higher-order cross-moments or structured mean terms, this construction avoids the need for explicit fourth-order moment tensors, making it computationally efficient and broadly applicable. However, this same property also makes it particularly sensitive to the *signal-to-noise ratio (SNR)* of the empirical covariances, because all rank information is inferred from ratios of covariances that may have small expected magnitude and high sampling variability.

In low-SNR regimes, random fluctuations in the empirical covariances can dominate the ratio computations, leading to unstable or biased rank estimates even when the theoretical ratio identities are exact. This motivates the detailed SNR analysis that follows: we study how the prior mean μ and variance (via the coefficient of variation cv) of the latent factors jointly determine the effective SNR of the covariance-based estimator. The goal is to identify prior configurations that maximize the estimator's robustness without sacrificing the closed-form structure that makes it analytically tractable.

Let v_S be the TT pure interaction defined by exact–sharing Moebius inversion over covariances C_S. With per-mode prior mean μ_m, variance σ_m^2, second raw moment $m_{2,m} = \mu_m^2 + \sigma_m^2$ and fourth raw moment $m_{4,m}$, the TT population pure term is

$$v_S \;=\; \left(\prod_{\ell=1}^{M-1} r_\ell^{e_\ell(S)}\right)\left(\prod_{m\in S}\sigma_m^2\right)\left(\prod_{m\notin S}\mu_m^2\right), \quad e_\ell(S) = \mathbf{1}\{\ell \in S \text{ or } \ell+1 \in S\} + 1.$$

Let $\widehat{C}_{S'}$ be the exact–sharing covariance estimated from n paired samples and $\widehat{v}_S = \sum_{S'\subseteq S}(-1)^{|S'|-|S'|}\widehat{C}_{S'}$. Under independent paired sampling per S' and large n,

$$\mathrm{Var}(\widehat{C}_{S'}) \approx \frac{1}{n}\left(\prod_\ell r_\ell^{h_\ell(S')}\right)\left(\prod_{m\in S'} m_{4,m}\right)\left(\prod_{m\notin S'} m_{2,m}^2\right) - \frac{1}{n}\left(\prod_\ell r_\ell^{2e_\ell(S')}\right)\left(\prod_{m\in S'} m_{2,m}^2\right)\left(\prod_{m\notin S'} \mu_m^4\right),$$

where $h_\ell(S') = 2e_\ell(TS' \in \{2,4\}$ are the fourth–moment exponents for TT. Consequently,

$$\mathrm{Var}(\widehat{v}_S) \;\approx\; \sum_{S'\subseteq S}\frac{1}{n}(-1)^{2|S|-2|S'|}\left[\left(\prod_\ell r_\ell^{h_\ell(S')}\right)\left(\prod_{m\in S'} m_{4,m}\right)\left(\prod_{m\notin S'} m_{2,m}^2\right) - \left(\prod_\ell r_\ell^{2e_\ell(S')}\right)\left(\prod_{m\in S'} m_{2,m}^2\right)\left(\prod_{m\notin S'} \mu_m^4\right)\right].$$

A convenient leading–order SNR follows by keeping the dominant term in the bracket:

$$\mathrm{SNR}(S) \equiv \frac{v_S}{\sqrt{\mathrm{Var}(\widehat{v}_S)}} \;\approx\; \sqrt{n}\prod_{m\in S}\frac{\sigma_m^2}{\sqrt{m_{4,m}}}\prod_{m\notin S}\frac{\mu_m^2}{m_{2,m}} \;\times\; \prod_\ell r_\ell^{e_\ell(S)-\frac{1}{2}h_\ell(S)}.$$

For TT, $h_\ell(S) = 2e_\ell(S)$, so the rank factor cancels and

$$\boxed{\mathrm{SNR}(S) \approx \sqrt{n}\prod_{m\in S}\frac{\sigma_m^2}{\sqrt{m_{4,m}}}\prod_{m\notin S}\frac{\mu_m^2}{\mu_m^2+\sigma_m^2}\,.}$$

Gamma prior (equal across modes). If $\theta^{(m)} \sim \mathrm{Gamma}(\alpha,\theta)$ with mean μ and $\mathrm{cv}^2 = 1/\alpha$, then $m_2 = \mu^2(1+\mathrm{cv}^2)$ and $m_4/\mu^4 = 1 + 6\mathrm{cv}^2 + 3\mathrm{cv}^4$. Therefore

$$\mathrm{SNR}(S) \approx \sqrt{n}\,\mu^{|S|}\left(\frac{\mathrm{cv}^2}{\sqrt{1+6\mathrm{cv}^2+3\mathrm{cv}^4}}\right)^{|S|}\left(\frac{1}{1+\mathrm{cv}^2}\right)^{M-|S|}.$$

This exhibits the empirical trade–off observed in practice: increasing cv boosts the shared–mode factor but diminishes the non–shared factor; an intermediate cv maximizes SNR for fixed $|S|$ and M.

Graphical analysis. We can observe from Figure 2 that the SNR of the pure-interaction covariance estimator for $|S| = 1$ and $|S| = 2$ (different sizes of interaction sets) has a narrow band leading to joint high SNR. In most areas of the graph we have prior configurations that lead to low SNR, affecting the quality and variability of any estimator based on pure-interactin covariance values.

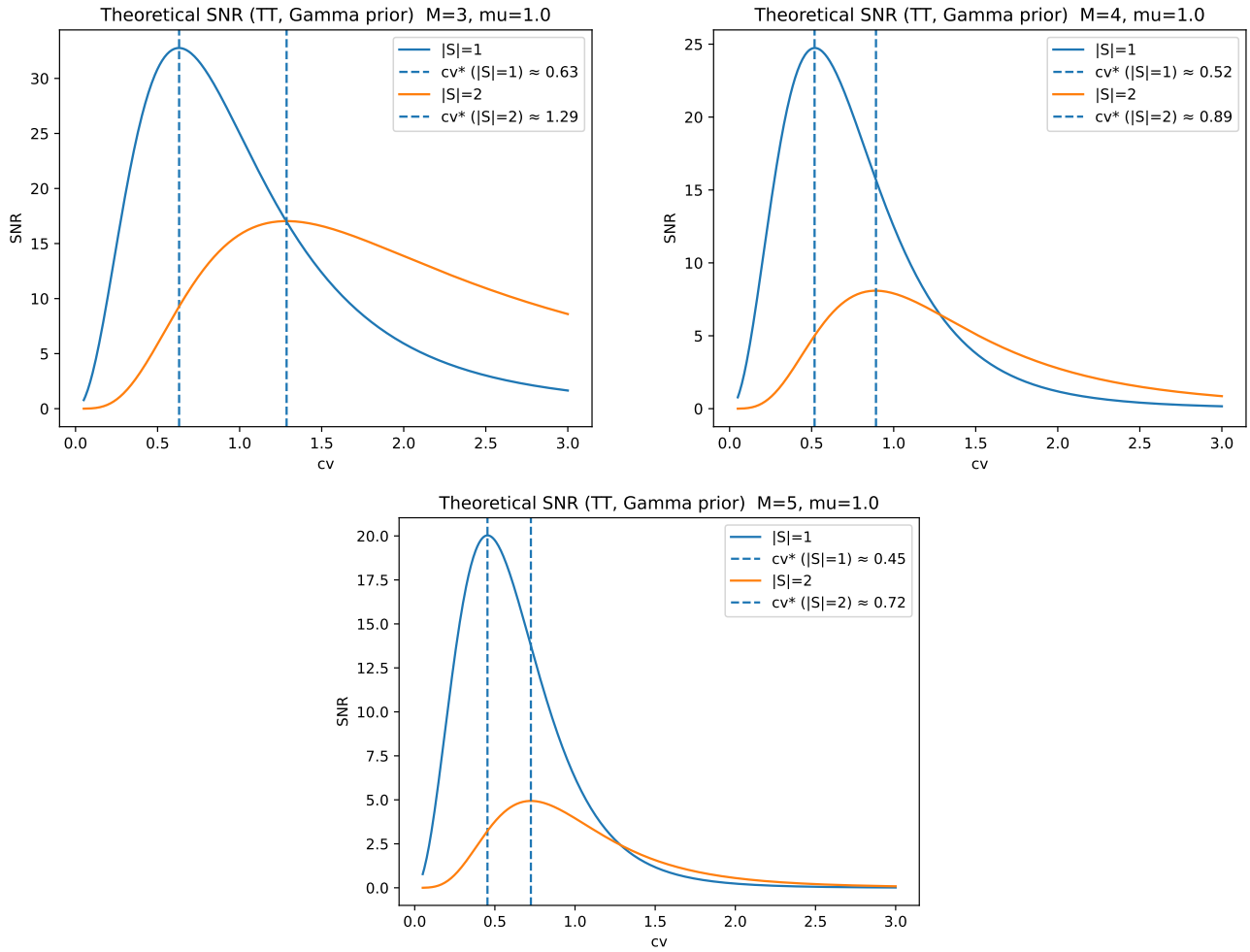

Figure 2: Theoretical SNR (TT, Gamma prior) for $M \in \{3, 4, 5\}$, $\mu = 1.0$. Vertical lines mark theoretical optimal region for cv*.

B.3 Bias Analysis of the Regularized (and Mean-Normalized) Rank Estimators

All closed-form rank estimators in this work are *ratios of estimated moments*, and hence inherit the well-known sensitivity of ratio estimators when denominators are small or noisy. We analyze the regularized estimator used throughout and the effect of the optional *mean normalization* (dividing each $\hat{v}_S$ by $(\widehat{E}[Y])^2$ before forming ratios).

Regularized ratio. For any target rank written as $r_p = \mathrm{Num}_p/\mathrm{Den}_p$, we use

$$\hat{r}_p^{\mathrm{reg}} \;=\; \mathrm{sign}\big(\widehat{\mathrm{Den}}_p\big)\,\frac{\widehat{\mathrm{Num}}_p}{\big|\widehat{\mathrm{Den}}_p\big| + \varepsilon_p}, \qquad \varepsilon_p \;=\; 1.96 \cdot \mathrm{SE}_{\mathrm{boot}}\big(\widehat{\mathrm{Den}}_p\big), \tag{32}$$

where $\mathrm{SE}_{\mathrm{boot}}$ is the bootstrap standard error computed *across* bootstrap replicates. The small additive ε_p stabilizes divisions by noisy denominators. Using $\mathrm{sign}(\cdot)$ preserves the correct orientation when sampling variation flips the estimated sign (rare but possible for small-S pure terms).

Mean normalization. Define mean-normalized pure terms $\tilde{v}_S := \hat{v}_S/\big((\widehat{E}[Y])^2 + \varepsilon_E\big)$ with a tiny $\varepsilon_E > 0$ (e.g. 10^{-12}). All CP/TT/TR rank identities are homogeneous in v_S, so replacing v_S by $\tilde{v}_S$ *does not change the population target.* In finite samples it reduces dispersion because we scale by a high-SNR quantity, $(\widehat{E}[Y])^2$. We therefore analyze both equation 32 with $(\widehat{\mathrm{Num}}, \widehat{\mathrm{Den}})$ built from $\hat{v}_S$ or from $\tilde{v}_S$.

Delta-method bias. Let $R = \widehat{\mathrm{Num}}/\widehat{\mathrm{Den}}$ and $f(x, y) = x/(y + \varepsilon)$ for a fixed $\varepsilon \geq 0$. A first-order delta-method expansion around $(\mu_X, \mu_Y) = (\mathbb{E}[\widehat{\mathrm{Num}}], \mathbb{E}[\widehat{\mathrm{Den}}])$ yields

$$\mathbb{E}[f(\widehat{\mathrm{Num}}, \widehat{\mathrm{Den}})] \approx \frac{\mu_X}{\mu_Y + \varepsilon} \; + \; \frac{1}{2}\Big(f_{xx}\mathrm{Var}(\widehat{\mathrm{Num}}) + 2f_{xy}\mathrm{Cov}(\widehat{\mathrm{Num}}, \widehat{\mathrm{Den}}) + f_{yy}\mathrm{Var}(\widehat{\mathrm{Den}})\Big),$$

with $f_x = 1/(\mu_Y + \varepsilon)$, $f_y = -\mu_X/(\mu_Y + \varepsilon)^2$, $f_{xx} = 0$, $f_{xy} = -1/(\mu_Y + \varepsilon)^2$, $f_{yy} = 2\mu_X/(\mu_Y + \varepsilon)^3$. Ignoring the mixed term (or if it is small by design—e.g., disjoint pairs), the dominant bias contribution is

$$\mathrm{Bias} \;\approx\; \mathbb{E}[\hat{r}_p^{\mathrm{reg}}] - r_p \;\approx\; -\,r_p \cdot \frac{\varepsilon_p}{\mu_Y} \;+\; r_p \cdot \frac{\mathrm{Var}(\widehat{\mathrm{Den}})}{(\mu_Y + \varepsilon_p)^2} \;-\; \frac{\mathrm{Cov}(\widehat{\mathrm{Num}}, \widehat{\mathrm{Den}})}{(\mu_Y + \varepsilon_p)^2}. \tag{33}$$

Thus the shrinkage term $-r_p\,\varepsilon_p/\mu_Y$ is *negative* and of the same order as the stochastic error; the variance term is positive and of order $1/N_{\mathrm{cov}}$; the cross-covariance term often reduces in absolute value under mean normalization and when we estimate v_S via ridge-regularized Moebius regression.

B.3.1 PARAFAC/CP

For CP, $r = \frac{v_{\{p,q\}}(E[Y])^2}{v_{\{p\}}v_{\{q\}}}$ with two singletons in the denominator. By the delta method in the *log domain*, write

$$\log \hat{r} = \big(\log \hat{v}_{\{p,q\}} - \log \hat{v}_{\{p\}} - \log \hat{v}_{\{q\}}\big) + \big(2\log \widehat{E}[Y]\big),$$

so $\mathrm{Var}(\log \hat{r})$ is a *sum* of variances/covariances of these log-terms. Translating back, the relative bias of equation 32 satisfies

$$\frac{|\mathrm{Bias}|}{r} = O\Big(\mathrm{CV}(\widehat{\mathrm{Den}})\Big) = O\Big(\tfrac{1}{\sqrt{N_{\mathrm{cov}}}}\Big),$$

where $\mathrm{CV}(\widehat{\mathrm{Den}}) = \sqrt{\mathrm{CV}^2(\hat{v}_{\{p\}}) + \mathrm{CV}^2(\hat{v}_{\{q\}}) + 2\,\rho\,\mathrm{CV}(\hat{v}_{\{p\}})\mathrm{CV}(\hat{v}_{\{q\}})}$ and ρ is the correlation between the singleton estimates (empirically small if pairs are sampled independently across S). Mean normalization divides all v_S by $(\widehat{E}[Y])^2$, shrinking both variance and covariance terms without changing the target.

B.3.2 Tensor Train

For interior bonds, $r_p = \frac{v_{\{p,p+1\}}v_{\{p-1,p+2\}}}{v_{\{p+1\}}v_{\{p-1,p,p+2\}}}$. The denominator is a product; in the *log domain* we have

$$\log \hat{r}_p = \big(\log \hat{v}_{\{p,p+1\}} + \log \hat{v}_{\{p-1,p+2\}}\big) - \big(\log \hat{v}_{\{p+1\}} + \log \hat{v}_{\{p-1,p,p+2\}}\big),$$

so $\mathrm{Var}(\log \hat{r}_p)$ is a sum of the *four* log-variances plus their cross-covariances. Exponentiating back implies

$$\frac{|\mathrm{Bias}|}{r_p} \;=\; O\Big(\sqrt{\sum_{S \in \mathcal{S}_p} \mathrm{Var}\big(\log \hat{v}_S\big)}\Big) \;=\; O\Big(\tfrac{1}{\sqrt{N_{\mathrm{cov}}}}\Big),$$

where $\mathcal{S}_p = \{\{p, p+1\},\ \{p-1, p+2\},\ \{p+1\},\ \{p-1, p, p+2\}\}$. Mean normalization replaces $\hat{v}_S$ by $\tilde{v}_S$, reducing both $\mathrm{Var}(\log \hat{v}_S)$ and cross-covariances in practice because $(\widehat{E}[Y])^2$ is high-SNR. The regularization bias term in equation 33 remains $O(1/\sqrt{N_{\mathrm{cov}}})$ since $\varepsilon_p \propto \mathrm{SE}(\widehat{\mathrm{Den}}_p)$.

B.3.3 Tensor Ring

For TR we estimate $\xi_p = \frac{(E[Y])^2 v_{\{p,p+1\}}}{v_{\{p\}}v_{\{p+1\}}}$ and solve the *linear* system $x_p + x_{p-1} - x_{p+1} = \psi_p$ with $x_p = \log r_p$ and $\psi_p = \log \hat{\xi}_p$. Hence

$$\widehat{\mathbf{x}} = \mathbf{C}^{-1}\widehat{\boldsymbol{\psi}}, \qquad \mathrm{Var}(\widehat{\mathbf{x}}) = \mathbf{C}^{-1}\mathrm{Var}(\widehat{\boldsymbol{\psi}})\big(\mathbf{C}^{-1}\big)^{\top}.$$

Each ψ_p is a *log-ratio* of estimated moments; its variance is reduced by mean normalization (replacing v_S by $\tilde{v}_S$), and any regularization is applied *componentwise* to the underlying denominators that define ξ_p (as in equation 32). Since $\mathbf{C}$ is fixed and well-conditioned for $M \geq 3$, the bias order remains $O(1/\sqrt{N_{\mathrm{cov}}})$.

B.3.4 General Guideline

Across CP/TT/TR, with or without mean normalization, and with denominator regularization equation 32, the induced bias satisfies

$$\frac{|\text{Bias}|}{r_p} \lesssim c \cdot \text{CV}\big(\widehat{\text{Den}}_p\big) = O\Big(\tfrac{1}{\sqrt{N_{\text{cov}}}}\Big),$$

for a constant c that depends smoothly on the moment mix and (weak) cross-covariances. In practice, the reduction in variance (and hence RMSE) from the stabilizer ε_p and from mean normalization outweighs the small $O(1/\sqrt{N_{\text{cov}}})$ bias. We therefore recommend (i) mean normalization of $\hat{v}_S$ by $(\widehat{E}[Y])^2$; (ii) bootstrap-based ε_p; and (iii) (optionally) least-squares Moebius inversion with a small ridge penalty when computing $\hat{v}_S$, which further reduces both variance and cross-covariances among the pure terms.

C ADDITIONAL EXPERIMENTS

Tensor Train (Figure 3). Across settings, TT rank recovery tracks the $y{=}x$ line for small and moderate true ranks, with dispersion increasing as ranks grow or the signal-to-noise ratio (SNR) drops (e.g., smaller Poisson means or more dispersed Gamma priors). This is expected for moment-based estimators: the pure terms v_S shrink in magnitude at low SNR, making inclusion–exclusion noisier and the denominators in the closed-form ratios closer to zero, which amplifies variance and occasionally induces slight upward bias. Interior bonds are typically better behaved than boundary bonds, which involve longer interaction sets. One practical mitigation that helped in our runs was *mean normalization* of all $\hat{v}_S$ by $(\widehat{E}[Y])^2$, which reduces spread without changing the target. Across panels, estimation accuracy improves as the effective SNR increases: low-mean or high-variance priors yield underestimation and wide dispersion, while moderate (μ, cv) combinations (as predicted by the SNR sweet-spot analysis in Appendix B.2) yield nearly unbiased recovery. The results confirm that the covariance-only TT estimator remains consistent when the prior mean and variance are chosen to keep $\text{SNR}(S) \gtrsim 5$ for the pure-term subsets involved in the ratio identities.

Tensor Ring (Figure 4). For TR, the median estimates follow the $y{=}x$ line across a wide range of ranks, with variance increasing at larger ranks and under lower SNR (smaller Poisson means / more dispersed Gamma priors). This behavior is consistent with our theory: each ξ_p is a log–ratio of estimated moments, so noise in singletons $v_{\{p\}}$ and adjacent pairs $v_{\{p,p+1\}}$ propagates through the circulant linear system for $\{\log r_p\}$. The cyclic coupling also means local errors can diffuse around the ring, slightly inflating uncertainty relative to TT at comparable ranks. In practice, *mean normalization* of $\hat{v}_S$ by $(\widehat{E}[Y])^2$ tighten the spread of $\widehat{\xi}_p$ and, consequently, the recovered ranks.

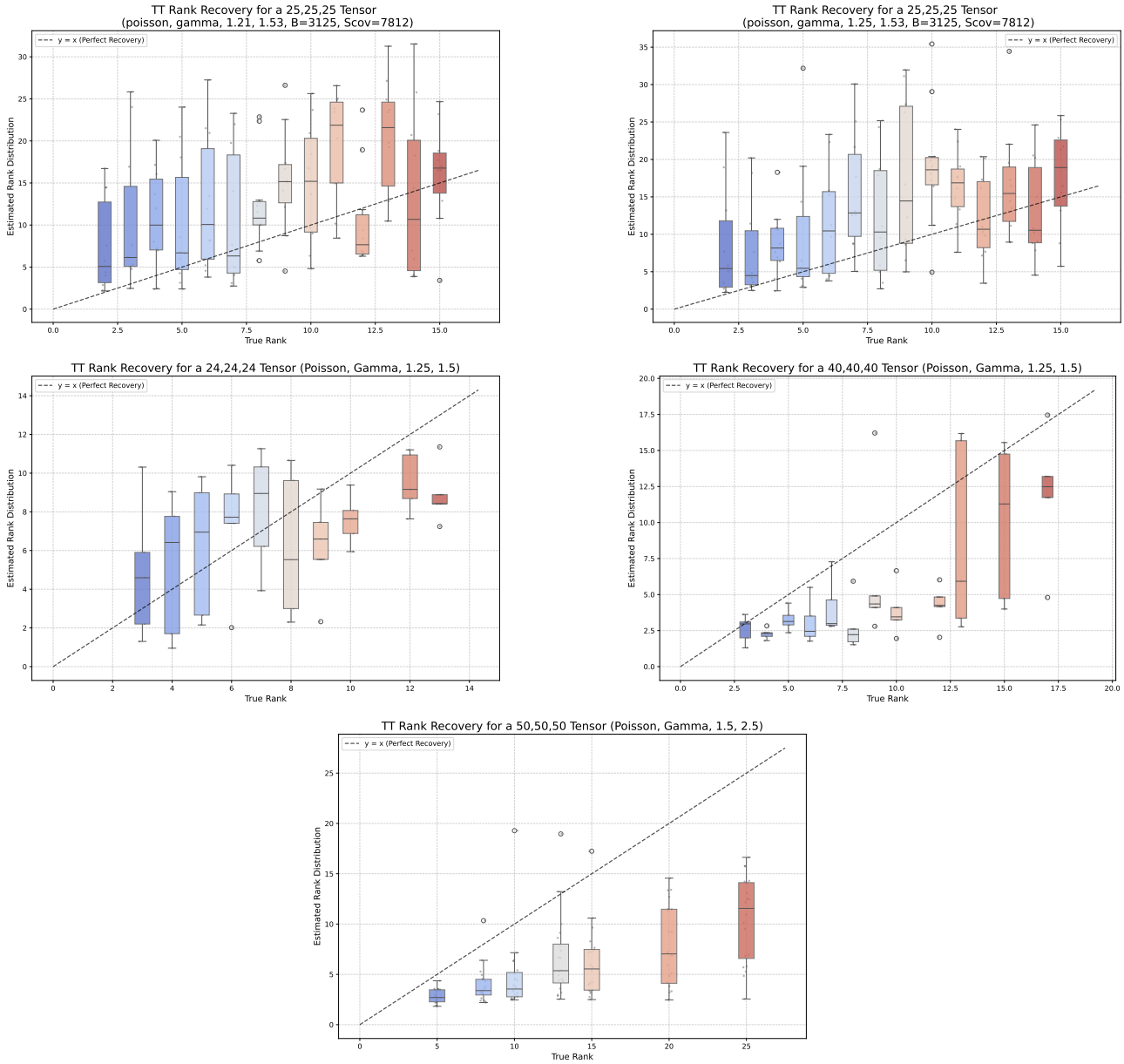

Figure 3: Tensor Train (TT) rank estimation under varying dimensions and Gamma–Poisson generative settings. Each panel reports median rank estimates (over 20 independent runs) grouped by true rank, with boxplots showing the distribution of empirical estimates. The dashed diagonal ($y{=}x$) indicates perfect recovery. All models are trained using the covariance-only TT estimator. **Top row:** TT tensors of size $(25, 25, 25)$ with Poisson likelihood and Gamma priors Gamma(α, θ) set to $(1.2, 1.5)$ (left) and $(1.25, 1.5)$ (right). **Middle row:** TT tensors of sizes $(24, 24, 24)$ and $(40, 40, 40)$ with Gamma$(1.25, 1.5)$ priors, showing improved stability with increasing tensor size. **Bottom:** TT tensor of size $(50, 50, 50)$ with Gamma$(1.5, 2.5)$ prior, corresponding to a higher mean and lower relative variance (lower cv).

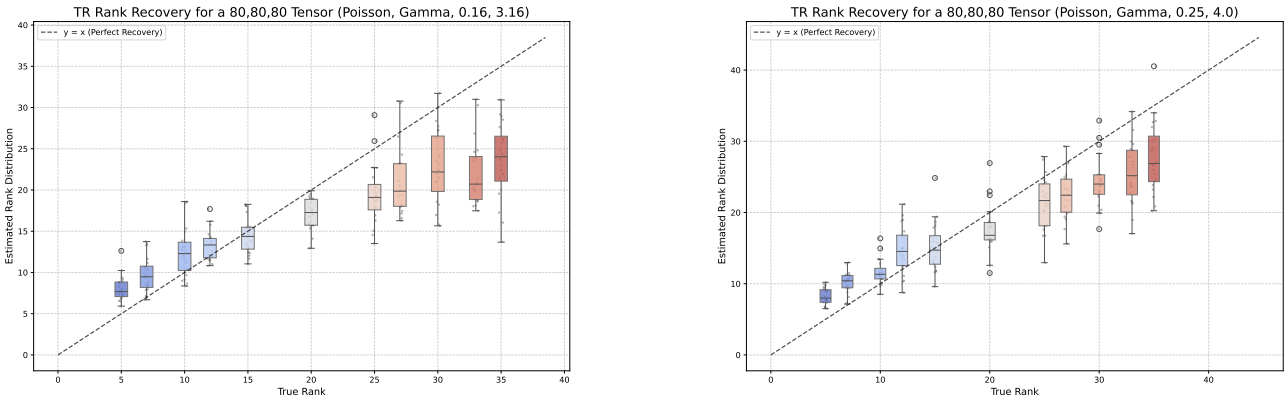

Figure 4: **Tensor Ring (TR) rank estimation on** $(80, 80, 80)$ **tensors with Gamma–Poisson generative models.** Each panel reports median rank estimates across 20 independent runs, grouped by true rank and visualized as boxplots. The dashed diagonal (y=x) denotes perfect recovery. Both experiments use a Poisson likelihood with Gamma priors Gamma(α, θ) given by $(0.16, 3.16)$ (left) and $(0.25, 4.0)$ (right). The shape and scale values correspond to different mean–variance tradeoffs, with the higher mean/lower variance prior (right) yielding improved accuracy and reduced dispersion, consistent with the expected gain in effective signal-to-noise ratio.