Stochastic Models in Probability and Statistics
Vol. 3, No. 1, 2026, pp. 45–52. Research Paper

Quasi-identifiability of multipath change-point models with general link functions

Masoud Majidizadeh*

Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

Email(s): m_majidizadeh@sbu.ac.ir

Abstract. The nonsingularity of the Fisher information matrix is a fundamental requirement for standard asymptotic inference. The framework developed by Asgharian (2014) shows that, for identifiable and smooth models, the set of singular parameter values is negligible. This result was previously established for multipath change-point (MCP) models in which the change hazard is specified through a proportional odds ratio structure. In this work, we substantially broaden that setting by proving the quasi-identifiability of MCP models under a general class of strictly monotone link functions, encompassing the probit, logit, and complementary log-log models as special cases. Our main result demonstrates that, for this broad family of models, the injectivity of the link function is sufficient to guarantee quasi-identifiability under standard regularity conditions. Consequently, the general singularity theory applies directly, implying that the Fisher information matrix is nonsingular for almost all parameter values. This, in turn, justifies standard likelihood-based inference for a considerably wider and more flexible class of models.

Keywords: Fisher information matrix; Link function; Multipath change-point models; Quasi-identifiability; Singularities.

1 Introduction

The multipath change-point (MCP) model, introduced in various forms by authors such as Joseph and Wolfson (1993), provides a flexible framework for analyzing longitudinal data where subjects may experience a change in regime at different, unobserved times. Conceptually, it belongs to the broad family of finite mixture models, which are a standard tool for modeling latent heterogeneity in longitudinal and survival data (e.g., McLachlan et al. (2019); Titterington et al. (1985)). Statistical inference in such models, particularly for the effects of covariates on the change-point process, often relies on the asymptotic properties of the maximum likelihood estimator (MLE). A fundamental prerequisite for establishing these properties, such as asymptotic normality and efficiency, is the nonsingularity of the Fisher information matrix (FIM) at the true parameter value.

The MCP model has evolved considerably since its inception. The foundational framework and the properties of the MLE were first established in the seminal works of Joseph and Wolfson (1992) and

*Corresponding author
Received: 24 October 2025 / Accepted: 24 June 2026
DOI: 10.22067/smps.2026.96138.1055

Joseph and Wolfson (1993). This initial framework was subsequently extended to accommodate more complex data structures, such as correlated observations between paths (Joseph et al., 1996). A pivotal extension, which forms the basis for the current work, was the formal incorporation of covariates into the model by Asgharian and Wolfson (2001), who investigated the modeling challenges and established the consistency of the MLE. The model continues to be an active area of research, with recent advancements focusing on modern statistical challenges such as simultaneous variable selection in high-dimensional settings (Shohoudi et al., 2016). This progression highlights the increasing complexity of the model, underscoring the need for a rigorous understanding of the theoretical underpinnings required for valid inference, particularly in the presence of covariates.

While verifying non-singularity is often straightforward in simple, well-posed models (Wang, 1993), it becomes a formidable challenge in more complex and empirically vital frameworks. Two such areas are of particular interest: finite mixture models and change-point models. Finite mixture models, whose identifiability was first systematically studied by Teicher (1963), are notorious for exhibiting singularities when components overlap, leading to a loss of rank in the information matrix (Chen, 1995; Lindsay, 1995; Manole and Ho, 2020; McLachlan et al., 2019). For complex mixture models like the MCP, direct verification of FIM nonsingularity for all possible parameter values is often intractable due to the intricate dependencies within the likelihood function. Indeed, the potential for a singular information matrix is a well-known theoretical challenge in the general mixture model literature, often arising from component redundancy or overfitting (Kim and Lindsay, 2015). A powerful alternative approach was developed by Asgharian (2014), who established a general theory concerning the singularities of the FIM. The central result is that if a model satisfies certain smoothness conditions and, crucially, is identifiable, then the set of parameter values for which the FIM is singular constitutes a nowhere dense set of Lebesgue measure zero. This effectively means that singularities are rare and do not pose a practical obstacle to inference for almost all data-generating processes. To illustrate this theory, Asgharian (2014) successfully applied it to an MCP model where the hazard of change, $\pi_k(\mathbf{z})$, was specified using a proportional odds ratio structure, which is closely related to the logistic link function.

A natural question arises: do these powerful theoretical guarantees hold for other common specifications of the change hazard? This paper answers this question affirmatively by moving beyond a single link function and addressing a general class of models. Our primary contribution is to establish the quasi-identifiability of an MCP model that employs a general strictly monotone link function for the hazard, which includes the probit, logit, and complementary log-log as prominent examples. This general result is significant as it allows us to directly invoke the theorems of Asgharian (2014) to conclude that the FIM is nonsingular almost everywhere in the parameter space. This work therefore substantially broadens the class of identifiable MCP models for which standard likelihood-based inference is theoretically sound.

2 The MCP model with a general link function

We follow the general framework of the MCP model as described in Asgharian (2014). The full data for a single subject consists of observations $\mathbf{x} = (x_1, \ldots, x_m)$ and a vector of time-invariant covariates $\mathbf{z}$. The model is formulated as a finite mixture, where the density of the observations is given by

$$f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^{m} \alpha_k(\mathbf{z}|\boldsymbol{\theta})\, g_{\mathbf{z},k}(\mathbf{x}|\gamma). \tag{1}$$

Here, $g_{\mathbf{z},k}(\mathbf{x}|\gamma)$ is the conditional density of the observations given that the change occurs at time k, and $\alpha_k(\mathbf{z}|\boldsymbol{\theta}) = P(\tau = k|\mathbf{Z} = \mathbf{z})$ is the mixing weight. The weights are constructed from the discrete-time hazard of change, $\pi_k(\mathbf{z}|\boldsymbol{\theta}) = P(\tau = k|\tau \geq k, \mathbf{Z} = \mathbf{z})$, as

$$\alpha_k(\mathbf{z}|\boldsymbol{\theta}) = \pi_k(\mathbf{z}|\boldsymbol{\theta}) \prod_{j=1}^{k-1} (1 - \pi_j(\mathbf{z}|\boldsymbol{\theta})), \tag{2}$$

for $k = 1, \ldots, m-1$, with $\alpha_m(\mathbf{z}|\boldsymbol{\theta}) = \prod_{j=1}^{m-1}(1 - \pi_j(\mathbf{z}|\boldsymbol{\theta}))$ to ensure they sum to one.

The central innovation of this paper lies in generalizing the specification of the hazard function. We model the hazard using a structure analogous to a generalized linear model (GLM), employing a general link function, G. The hazard is defined as

$$\pi_k(\mathbf{z}|\boldsymbol{\theta}) = G\left(\mathbf{z}^\top \boldsymbol{\beta} + \rho_k\right), \quad \text{for } k = 1, \ldots, m-1, \tag{3}$$

where

- $G : \mathbb{R} \to (0,1)$ is a known link function, assumed to be twice continuously differentiable and strictly monotone;
- $\mathbf{z} \in \mathbb{R}^p$ is the vector of covariates;
- $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients;
- ρ_k represents the baseline hazard component at time k, forming the vector $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{m-1}) \in \mathbb{R}^{m-1}$.

The complete parameter vector for the model is $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\rho}, \gamma)$.

Remark 1 (The class of admissible link functions)**.** *The requirement that the link function G be strictly monotone (and thus injective) is the key property that facilitates identifiability. This general framework encompasses many of the most common models for binary outcomes, making our results widely applicable. Prominent examples include:*

- ***The probit model:*** *$G(x) = \Phi(x)$, where Φ is the standard normal CDF.*
- ***The logit model:*** *$G(x) = \frac{e^x}{1+e^x}$, which corresponds to the proportional odds ratio structure used in Asgharian (2014).*
- ***The complementary log-log (cloglog) model:*** *$G(x) = 1 - \exp(-\exp(x))$.*

By proving our results for a general G, we simultaneously establish them for all these specific models and countless others that satisfy the monotonicity condition.

3 Main result: quasi-identifiability

The primary theoretical hurdle to applying the singularity framework of Asgharian (2014) is to establish model identifiability. For models incorporating covariates, the appropriate concept is quasi-identifiability. We first formally define this concept, following the original work.

Definition 1 (Quasi-identifiability, Asgharian (2014))**.** *A model parameterized by $\boldsymbol{\theta} \in \Theta$ is said to be quasi-identifiable if for any two distinct parameter vectors $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$, the set of covariates $\mathbf{z}$ for which $f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}')$ for almost all $\mathbf{x}$ has measure zero with respect to the covariate distribution.*

To construct the proof of our main theorem in a clear, modular fashion, we first establish a simple lemma. This lemma confirms the intuitive fact that if the final mixing weights α_k are identical, then the initial hazard probabilities π_k that generated them must also have been identical.

Lemma 1 (Equivalence of weights and hazards)**.** *Let the mixing weights $\{\alpha_k\}_{k=1}^{m}$ be generated from the hazard probabilities $\{\pi_k\}_{k=1}^{m-1}$ as in equation (2). Assume $\pi_k \in (0,1)$ for all k, as a hazard of 0 or 1 would imply a deterministic change process, collapsing the mixture structure. If two sets of hazards, $\{\pi_k\}$ and $\{\pi_k^*\}$, generate identical mixing weights $\alpha_k = \alpha_k^*$ for all $k = 1,\ldots,m$, then it must be that $\pi_k = \pi_k^*$ for all $k = 1,\ldots,m-1$.*

Proof. We proceed by induction on k. **Base Case (k=1):** By definition, $\alpha_1 = \pi_1$. Since $\alpha_1 = \alpha_1^*$, it immediately follows that $\pi_1 = \pi_1^*$.
Inductive Step: Assume that $\pi_j = \pi_j^*$ for all $j = 1,\ldots,k-1$. We need to show this implies $\pi_k = \pi_k^*$. The definition for α_k is $\alpha_k = \pi_k \prod_{j=1}^{k-1}(1-\pi_j)$. Since $\alpha_k = \alpha_k^*$, we have

$$\pi_k \prod_{j=1}^{k-1}(1-\pi_j) = \pi_k^* \prod_{j=1}^{k-1}(1-\pi_j^*).$$

By our inductive hypothesis, $\pi_j = \pi_j^*$ for all $j < k$, so the product terms $\prod(\cdot)$ on both sides are identical. Since $\pi_j < 1$, this product is non-zero. We can therefore divide both sides by this common term to obtain $\pi_k = \pi_k^*$. This completes the induction. □

We are now equipped to state and prove the main result of this paper.

Theorem 1 (General quasi-identifiability)**.** *Consider the MCP model defined by equations (1) and (3), using any strictly monotone, twice continuously differentiable link function G. Assume that:*

(i) The family of component densities $\{g_{\mathbf{z},k}(\cdot|\gamma)\}_{k=1}^{m}$ is identifiable with respect to γ and forms a linearly independent set.

(ii) The covariate vector $\mathbf{z}$ contains an intercept, and its distribution is not concentrated on a hyperplane.

Then, the model is quasi-identifiable with respect to the full parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\rho}, \gamma)$.

Proof. Suppose for two parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, we have $f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}^*)$ for all covariates $\mathbf{z}$ in a set A of positive measure. From the model's mixture structure in (1), this implies

$$\sum_{k=1}^{m} \alpha_k(\mathbf{z}|\boldsymbol{\theta})\, g_{\mathbf{z},k}(\mathbf{x}|\gamma) = \sum_{k=1}^{m} \alpha_k(\mathbf{z}|\boldsymbol{\theta}^*)\, g_{\mathbf{z},k}(\mathbf{x}|\gamma^*).$$

By assumption (i), the component densities $\{g_{\mathbf{z},k}\}$ are linearly independent, which for a finite mixture implies the equality of both the component parameters and the mixing weights. Thus, for $\mathbf{z} \in A$, we must have $\gamma = \gamma^*$ and $\alpha_k(\mathbf{z}|\boldsymbol{\theta}) = \alpha_k(\mathbf{z}|\boldsymbol{\theta}^*)$ for all k.

By Lemma 1, the equality of the mixing weights implies the equality of the underlying hazard probabilities: $\pi_k(\mathbf{z}|\boldsymbol{\theta}) = \pi_k(\mathbf{z}|\boldsymbol{\theta}^*)$. Substituting our general model specification from (3) yields

$$G(\mathbf{z}^\top \boldsymbol{\beta} + \rho_k) = G(\mathbf{z}^\top \boldsymbol{\beta}^* + \rho_k^*).$$

The core assumption on our link function G is that it is **strictly monotone**, which guarantees it is **injective** (one-to-one). Therefore, the equality of the function's outputs implies the equality of its inputs:

$$\mathbf{z}^\top \boldsymbol{\beta} + \rho_k = \mathbf{z}^\top \boldsymbol{\beta}^* + \rho_k^*.$$

This can be rearranged to $\mathbf{z}^\top(\boldsymbol{\beta} - \boldsymbol{\beta}^*) = \rho_k^* - \rho_k$. This equation must hold for all $\mathbf{z} \in A$. If $\boldsymbol{\beta} \neq \boldsymbol{\beta}^*$, the left-hand side is a non-constant linear function of $\mathbf{z}$, while the right-hand side is a constant. By assumption (ii), the distribution of $\mathbf{z}$ is not concentrated on a hyperplane, meaning the set of $\mathbf{z}$ values for which a non-constant linear function equals a constant must have measure zero. This contradicts our premise that the equality holds on a set A of positive measure.

The only way to resolve this contradiction is to conclude that $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. Substituting this back into the linear equation immediately yields $\rho_k = \rho_k^*$. Combined with $\gamma = \gamma^*$, we have shown that $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, which completes the proof of quasi-identifiability. □

Example 1 (Importance of the theorem's assumptions)**.** To illustrate the necessity of the conditions in Theorem 1, consider a simple case with $m = 2$ (one potential change-point) and a single covariate z_1. For concreteness, we use a probit link, so the hazard is $\pi_1(\mathbf{z}) = \Phi(\beta_0 + \beta_1 z_1)$, where $\mathbf{z} = (1, z_1)^\top$. The model density is a two-component mixture

$$f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = \pi_1(\mathbf{z}) g_1(\mathbf{x}|\gamma) + (1 - \pi_1(\mathbf{z})) g_2(\mathbf{x}|\gamma).$$

Now, consider scenarios where the assumptions of the theorem are violated.

1. **Violation of covariate richness (Assumption ii):** Suppose the covariate z_1 is not sufficiently rich and can only take a single value, c. Then the linear predictor becomes a constant, $\eta = \beta_0 + \beta_1 c$, and so does the hazard, $\pi_1 = \Phi(\eta)$. In this case, countless combinations of β_0 and β_1 can produce the exact same value for η. For instance, $(\beta_0, \beta_1) = (\eta, 0)$ and $(\beta_0, \beta_1) = (\eta - c, 1)$ would yield the same hazard. The parameters β_0 and β_1 are not separately identifiable. This highlights why the covariates cannot be concentrated on a hyperplane (in this case, a single point).

2. **Violation of component distinguishability (Assumption i):** Suppose the component densities are identical, i.e., $g_1(\cdot|\gamma) = g_2(\cdot|\gamma) = g(\cdot|\gamma)$. The mixture model then collapses, as the mixing weights become irrelevant:
$$f_{\mathbf{z}}(\mathbf{x}|\boldsymbol{\theta}) = \pi_1(\mathbf{z}) g(\mathbf{x}|\gamma) + (1 - \pi_1(\mathbf{z})) g(\mathbf{x}|\gamma) = g(\mathbf{x}|\gamma).$$
The resulting density is entirely independent of the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$. Consequently, these parameters are completely unidentifiable as they have no influence on the likelihood. This illustrates the necessity of linear independence for the component densities.

4 Implications for the information matrix

Having established the quasi-identifiability for a broad class of MCP models in Theorem 1, we are now positioned to directly apply the general singularity theory developed in Asgharian (2014). This theory connects model identifiability with the geometric and analytic properties of the set of singularities of the Fisher information matrix, $I(\boldsymbol{\theta})$. Our general result provides the key to unlock these powerful implications.

The following corollary is the principal outcome of our analysis. It confirms that for any model within our proposed general class, the set of parameter values where standard asymptotic theory might fail is negligible.

Corollary 1. *Consider the MCP model defined with any strictly monotone, twice continuously differentiable link function G as in Section 2. Assume that the conditions of Theorem 1 hold. Further, assume that the model satisfies the necessary smoothness and regularity conditions (e.g., A.1–A.5 and B.1–B.3 as specified in Asgharian (2014)). Let $\Lambda = \{\boldsymbol{\theta} \in \Theta : \det I(\boldsymbol{\theta}) = 0\}$ be the set of singularities of the Fisher information matrix. Then:*

1. *The set Λ is a nowhere dense subset of the parameter space Θ.*

2. *If, in addition, the component densities $g_{\mathbf{z},k}(\mathbf{x}|\gamma)$ are analytic functions of the parameters, the set Λ has Lebesgue measure zero.*

Proof. The proof is a direct application of the main results from Asgharian (2014). Theorem 1 of this paper establishes the quasi-identifiability of our general model, which is the central prerequisite for their Theorems 3 and 4. By assuming that the link function G is twice continuously differentiable and that the component densities satisfy standard smoothness conditions, the overall regularity conditions A.1–A.5 and B.1–B.3 of Asgharian (2014) are met. Specifically:

- Theorem 3 of Asgharian (2014) states that for any quasi-identifiable and sufficiently smooth model, the set of singularities Λ is nowhere dense.
- Theorem 4 of Asgharian (2014) strengthens this result, stating that if the model is also analytic, Λ has Lebesgue measure zero.

Since our general model satisfies the key conditions of quasi-identifiability and smoothness, the conclusions of the corollary follow directly by invoking their theorems. □

Remark 2 (Theoretical and practical implications)**.** *The conclusions of Corollary 1 are of profound theoretical importance. They provide a rigorous justification for the use of standard likelihood-based inferential procedures (e.g., Wald tests, likelihood ratio tests, and confidence intervals based on the inverse of the FIM) for a wide and versatile class of MCP models. The property of being nowhere dense implies that singularities are isolated, while having Lebesgue measure zero implies that the probability of randomly selecting a singular parameter is zero. Together, these results confirm that the pathologies associated with a singular FIM are confined to a negligible, thin set of parameter values.*

It is important, however, to distinguish this theoretical, asymptotic result from practical, finite-sample challenges. In applied settings, numerical issues akin to singularity can arise due to data structure, particularly the problem of complete or quasi-complete separation, where covariates perfectly predict the change-point time. In such cases, maximum likelihood estimates may diverge, and the observed Fisher information matrix can become ill-conditioned or numerically singular. Our results provide the theoretical foundation for asymptotic inference, but they do not eliminate the need for careful model diagnostics and an awareness of data-induced instabilities in finite samples.

5 Conclusion

In this paper, we addressed the fundamental issue of Fisher information matrix singularity for multipath change-point (MCP) models, building upon the theoretical framework of Asgharian (2014). Their work guaranteed the negligibility of singularities for identifiable models, a result they demonstrated for an MCP model with a proportional odds ratio (logistic) structure for the change hazard. We investigated whether these guarantees were an artifact of that specific choice or a more fundamental property of the modeling framework.

Our main contribution was to unify and generalize this result by establishing, under standard smoothness conditions, the quasi-identifiability of MCP models for a broad class of strictly monotone link functions. By focusing on the core mathematical property required for identifiability—the injectivity of the link function—our main theorem (Theorem 1) simultaneously covers the probit, logit, complementary log-log, and numerous other model specifications under a single, coherent proof.

The principal implication, presented in Corollary 1, is that for this entire class of models, the set of FIM singularities is both nowhere dense and of Lebesgue measure zero. This finding provides a rigorous and general theoretical justification for the application of standard likelihood-based inference to a versatile set of MCP models. Our work not only broadens the family of models for which inference is known to be theoretically sound but also reinforces the robustness of the singularity framework by demonstrating its wide applicability. While this asymptotic result provides a strong foundation, we also acknowledge that it must be distinguished from finite-sample challenges, where issues like complete data separation can still lead to numerical instability.

Future research could extend this line of inquiry in several directions, such as investigating the identifiability of models with more complex structures, including time-varying covariates or multiple change-points; exploring the challenges posed by non-monotone or data-driven link functions; or developing robust inferential methods that are less sensitive to the practical problem of data separation in finite samples. In essence, this work demonstrates that the validity of standard inference in these complex models hinges on a simple and verifiable property of the chosen link function: its monotonicity. This insight provides both theoretical assurance and practical guidance to researchers working with multipath change-point models.

Acknowledgements

I am deeply grateful to Professor Masoud Asgharian (McGill University) for his continuous invaluable guidance, which has significantly shaped both this work and my broader research endeavors. We would like to thank the Editor-in-Chief, Associate Editor and anonymous reviewers for their thorough evaluation of the manuscript and for their valuable comments and suggestions.

References

Asgharian, M. (2014). On the singularities of the information matrix and multipath change-point problems. *Theory of Probability and Its Applications*, **58**, 546–561.

Asgharian, M. and Wolfson, D. B. (2001). Covariates in multipath change-point problems: Modelling and consistency of the MLE. *Canadian Journal of Statistics*, **29**, 515–528.

Chen, J. (1995). Optimal rate of convergence for finite mixture models. *The Annals of Statistics*, **23**, 221–233.

Joseph, L., Vandal, A. C., and Wolfson, D. B. (1996). Estimation in the multipath change point problem for correlated data. *The Canadian Journal of Statistics*, **24**, 37–53.

Joseph, L. and Wolfson, D. B. (1992). Estimation in multi-path change-point problems. *Communications in Statistics-Theory and Methods*, **21**, 897–913.

Joseph, L. and Wolfson, D. B. (1993). Maximum likelihood estimation in the multi-path change-point problem. *Annals of the Institute of Statistical Mathematics*, **45**, 511–530.

Kim, D. and Lindsay, B. G. (2015). Empirical identifiability in finite mixture models. *Annals of the Institute of Statistical Mathematics*, **67**, 745–772.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry, and Applications.* NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Mathematical Statistics; Penn. State University.

Manole, T. and Ho, N. (2020). Uniform convergence rates for maximum likelihood estimation under two-component Gaussian mixture models. *arXiv preprint arXiv:2006.00704.*

McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, **6(**, 355–378.

Shohoudi, A., Khalili, A., Wolfson, D. B., and Asgharian, M. (2016). Simultaneous variable selection and de-coarsening in multi-path change-point models. *Journal of Multivariate Analysis*, **143**, 15–429.

Teicher, H. (1963). Identifiability of finite mixtures. *The Annals of Mathematical Statistics*, **34**, 1265–1269.

Titterington, D. M., Smith, A. F., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons, New York.

Wang, X. (1993). Non-singularity of fisher information for autoregressive moying-average processes. *Journal of Time Series Analysis*, **14**, 547–548.