Multivariate Poisson models based on comonotonic and counter-monotonic shocks

Juliana Schulz

Doctor of Philosophy

Department of Mathematics and Statistics
McGill University
Montréal, Québec, Canada

November 2018

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Abstract

In this thesis, two novel bivariate Poisson models are introduced: one encompassing positive dependence and the other for negative dependence. In the case of positively correlated margins, the proposed bivariate Poisson model formulation relies on the notion of comonotonic shocks. The proposed construction yields a bivariate Poisson model for positive associations with a fully flexible dependence structure; it is parametrized in terms of the marginal means and a dependence parameter which takes on values in the unit interval. The dependence parameter regulates the strength of the association between the margins and can accommodate any possible correlation ranging from zero (independence) to an upper bound representing the strongest possible correlation between arbitrary Poisson margins.

The model formulation in the case of negative dependence is analogous to that for positively correlated margins. The proposed bivariate Poisson model for negative dependence relies on the notion of counter-monotonic shocks and yields a fully flexible model for negatively correlated Poisson margins. Similar to the case of positive dependence, the strength of the dependence in the proposed counter-monotonic shock model is regulated by the dependence parameter, with correlation ranging from a lower bound, which represents the smallest possible correlation between arbitrary Poisson margins, to zero.

In addition to the two bivariate models, a multivariate extension of the comonotonic shock Poisson model is defined. Although the formulation of the proposed multivariate Poisson model implies certain restrictions on the covariance structure in dimensions greater than 2, the proposed formulation is nonetheless useful for modeling positively correlated count data.

For all three proposed models, several distributional properties are established and various estimation techniques are described in detail. These techniques are validated through several simulations studies. Furthermore, data illustrations are used to highlight the utility of each of the proposed multivariate Poisson models. Altogether, the proposed classes of multivariate Poisson distributions are each based on stochastic representations that are interpretable and intuitive. Moreover, in the bivariate setting the model constructions yield a fully flexible dependence structure.

Résumé

Dans cette thèse, deux nouveaux modèles de Poisson bivariés sont proposés : le premier induit une dépendance positive et l'autre une dépendance négative entre les variables. Dans le cas de marges corrélées positivement, la formulation du modèle de Poisson bivarié proposée repose sur la notion de choc comonotone. Cette construction conduit à un modèle de Poisson bivarié pour les associations positives avec une structure de dépendance entièrement flexible ; elle est paramétrée en termes des moyennes marginales et d'un paramètre de dépendance à valeurs dans l'intervalle unité. Le paramètre de dépendance contrôle le degré d'association entre les marges et peut prendre en compte toute corrélation possible entre zéro (indépendance) et une borne supérieure qui représente la corrélation la plus forte possible entre des marges de lois de Poisson.

La formulation du modèle en cas de dépendance négative est analogue à celle où les marges sont corrélées positivement. Le modèle de Poisson bivarié proposé pour la dépendance négative repose sur la notion de choc anti-monotone et s'avère totalement flexible pour les marges de Poisson corrélées négativement. Comme dans le cas de la dépendance positive, le degré de dépendance dans le modèle à choc anti-monotone est régulé par le paramètre de dépendance dont la corrélation varie entre la borne minimale et zéro, où la première de ces valeurs représente la plus petite corrélation possible entre des variables de lois de Poisson.

En plus des deux modèles bivariés, une généralisation multivariée du modèle de choc comonotone de Poisson est proposée. Bien que la formulation de ce modèle de Poisson multivarié limite en partie la structure de covariance en dimension supérieure à 2, elle n'en est pas moins utile pour modéliser des données de dénombrement corrélées positivement.

Pour les trois modèles proposés, plusieurs propriétés stochastiques sont établies et diverses techniques d'estimation sont décrites en détail. Ces techniques sont validées au moyen de diverses études de simulation. Des illustrations concrètes permettent en outre de mettre en évidence l'utilité de chacun des modèles de Poisson multivariés proposés. En somme, ces nouvelles classes de lois de Poisson multivariées reposent sur des représentations stochastiques interprétables et intuitives. De plus, dans le cas bivarié, la structure de dépendance des modèles est totalement flexible.

Acknowledgments

I would like to begin by thanking my supervisor, Professor Christian Genest, for his invaluable guidance and support throughout my doctoral studies. His insight greatly contributed to my thesis and I am very grateful. I would also like to thank Professor Mhamed Mesfioui (Université du Québec à Trois-Rivières), who collaborated with us on this project. His knowledge and contributions brought many improvements to the work. I am also very grateful towards Professor Russell Steele for our many insightful discussions.

I would also like to express my deepest gratitude towards my husband, parents, family and friends for their patience, support and encouragement throughout my studies.

Finally I would like to acknowledge the financial support I received throughout my PhD studies, notably from the Fonds de recherche du Québec – Nature et Technologies PhD Research Fellowship, the Schulich Graduate Fellowship as well as the funding I received through Dr. Genest's research grants from the Canada Research Chair Program, the Natural Sciences and Engineering Research Council of Canada, the Canadian Statistical Sciences Institute, and the Fonds de recherche du Québec – Nature et Technologies.

Contributions to original knowledge

Chapter 3:

Chapter 3 consists primarily of an extensive review of the theoretical implications of comonotonicity. Section 3.5 contains original work in that, to our knowledge, expressing the distribution of comonotonic Poisson pairs in terms of paired-ordered statistics is novel.

Chapter 4:

Chapter 4 consists entirely of original scholarship. In this chapter, the proposed comonotonic shock bivariate Poisson model is defined, distributional properties are derived and estimation techniques are described. In addition, extensive simulations are performed along with a novel application to rainfall data.

Chapter 5:

Chapter 5 consists entirely of original scholarship. This chapter is devoted to describing the proposed bivariate Poisson model for negative dependence, including its distributional properties as well as various approaches for estimation. Several simulations and a data application are also included in this chapter.

Chapter 6:

Chapter 6 consists entirely of original scholarship. This chapter proposes a multivariate extension to the bivariate Poisson model for positive dependence described in Chapter 4. Similarly to the aforementioned chapter, distributional properties and estimation techniques are described for the proposed multivariate Poisson model. In addition, several simulations are performed and a data application is included.

Contributions of authors

Chapter 3

I was responsible for the entirety of Chapter 3. The original portion of this chapter is predominantly contained in Section 3.5. In this section, I derived an alternative formulation for the joint distribution of comonotonic Poisson random pairs. This allowed to write the log-likelihood in terms of pair-ordered statistics and establish the maximum likelihood estimators in the case of comonotonic Poisson pairs. The latter was shown to differ from the marginal maximum likelihood estimators.

Chapter 4

In Chapter 4, the proposed bivariate Poisson model for positive dependence is defined. This is joint work done with Professors Christian Genest and Mhamed Mesfioui. We developed the model formulation together. I then derived several of the distributional properties. The proof of the PQD ordering given in Lemma 4.1 was derived jointly with Professor Genest, as was the work for the method of moments estimation approach. I described both likelihood-based estimation techniques (maximum likelihood and inference functions for margins estimation). I also performed both the simulations and data illustration.

Chapter 5

Chapter 5 introduces the bivariate Poisson model for negative dependence. Again, this is joint work done with Professors Genest and Mesfioui. The division of the work is similar to that for Chapter 4, as much of the properties and techniques are analogous to the model for positive dependence.

Chapter 6

In Chapter 6 a multivariate extension to the bivariate Poisson model for positive dependence is proposed. I did this work in its entirety.

Contents

List of Figures

List of Tables

1

Introduction

In statistics there are a myriad of well established univariate distributions. However, with a few exceptions, there are in general no uniquely defined multivariate analogues to these univariate families. In modern practical applications, there is a considerable demand for multivariate families of distributions that can accurately model correlated data. To this end, many advances in the field of multivariate statistics have attempted to fill this need.

Certainly, there are several techniques that can be used to define multivariate models. In particular, the concept of copulas greatly revolutionized the field of multivariate statistics in that it provides a flexible framework for building multivariate models with arbitrary margins. Although copula models remain valid constructions for count data, many difficulties arise when they are applied to discrete margins. Mixture models also prove to be another useful tool for defining multivariate models, although such constructions often lead to very complex models.

In the case of the Poisson distribution, there have been many proposals for defining its multivariate version, an overview of which will be given in Chapter 2. Focusing on the bivariate setting, the most popular definition of a bivariate Poisson model is derived from the notion of a common shock and relies on the property of infinite divisibility inherent in the univariate Poisson distribution. This construction allows to model correlated Poisson random variables but greatly restricts the strength of the dependence. As will be detailed in Chapter 2, this classical bivariate Poisson model only allows for positive dependence and, moreover, does not accommodate the full spectrum of possible positive correlation that an arbitrary pair of Poisson random variables can exhibit.

In this thesis, a novel bivariate Poisson model is introduced. The model construction relies on the notion of comonotonicity and provides a fully flexible model framework for positively correlated Poisson random variables. The proposed family stems from an interpretable and intuitive stochastic representation. Moreover, the proposed bivariate Poisson model extends naturally to higher dimensions. Using an analogous construction based on the concept of counter-monotonicity, the proposed framework can also be adapted to define a bivariate Poisson model for negatively dependent margins.

This thesis is organized as follows. Chapter 2 provides a review of several proposals found in the statistical literature for defining a multivariate Poisson distribution. The classical bivariate Poisson model, in particular, is discussed in great detail.

In Chapter 3, the notions of comonotonicity and counter-monotonicity are described. The differences between the Fréchet–Hoeffding bounds in the case of continuous margins and discrete margins are discussed, along with the difficulties that arise in estimation. Finally, the particular case of the upper Fréchet–Hoeffding boundary distribution with Poisson margins is examined in detail.

The proposed bivariate Poisson model for positive dependence is presented in Chapter 4. Several distributional properties are explored as well as various estimation techniques. The proposed bivariate model is validated through various simulation studies and further applied to environmental data.

The model is then adapted to define a bivariate Poisson model for negative dependence in Chapter 5. The outline of this chapter is similar to the previous: distributional properties are described in detail and several estimation approaches are discussed. The model is then tested through various simulations and an application to sports data is provided.

Finally, Chapter 6 extends the model for positive dependence to higher dimensions. Distributional properties and estimation techniques are explored in detailed and the model is tested in various simulations. Furthermore, a data application in the trivariate setting is provided. To end, concluding remarks are found in Chapter 7.

2

Literature Review

This chapter provides an extensive summary of the classical bivariate Poisson model, including its distributional properties and estimation techniques. The natural multivariate extension to the classical bivariate Poisson model is subsequently discussed in detail. A survey of variants to the common shock method for constructing a multivariate Poisson model is also presented, followed by a brief overview of some alternative models for correlated count data that have been proposed in the literature.

2.1 The classical bivariate Poisson model

2.1.1 Introduction and motivation

A random pair (X_1, X_2) is said to arise from a bivariate Poisson distribution if both X_1 and X_2 are marginally univariate Poisson random variables. Although there is no unique way to characterize the joint distribution of such a pair (X_1, X_2), the most commonly used specification results from the so-called trivariate reduction technique, generally attributed to Campbell (1934). This construction makes use of the fact that the univariate Poisson distribution is infinitely divisible and thereby generates a pair of dependent Poisson random variables by convoluting a set of independent Poisson variables with a common shock variable. Following the parametrization considered by Holgate (1964), let Y_1, Y_2 and Z be mutually independent Poisson random variables with means $\lambda_1 - \xi$, $\lambda_2 - \xi$ and ξ, respectively, where $(\lambda_1, \lambda_2) \in (0, \infty)^2$ and $\xi \in [0, \min(\lambda_1, \lambda_2)]$. A pair (X_1, X_2) of correlated Poisson random variables is then generated by setting

$$X_1 = Y_1 + Z, \quad X_2 = Y_2 + Z. \tag{2.1}$$

This model will be referred to as the classical bivariate Poisson model, or equivalently, the common shock model. Note that in the literature, however, it is commonly referred to simply as the bivariate Poisson model. The formulation outlined in (2.1) is intuitive and interpretable, which surely contributed to its popularity.

Several other authors have considered different approaches to derive the classical bivariate Poisson distribution. For example, Kawamura (1973) and Marshall and Olkin (1985) consider the classical bivariate Poisson model as the limiting distribution of a bivariate Binomial distribution, while Papageorgiou (1983) and AlMuhayfith et al. (2016) work with conditional distributions.

It is clear from the construction given in (2.1) that the source of dependence in the model is attributed to the common shock variable Z. As is often pointed out in the literature, the main drawback of the classical model is that it only allows for positive dependence. Indeed, since $\text{cov}(X_1, X_2) = \text{var}(Z)$, the correlation implied by the classical model is given by

$$\text{corr}(X_1, X_2) = \frac{\xi}{\sqrt{\lambda_1 \lambda_2}}, \tag{2.2}$$

which is non-negative and, for fixed marginal rates λ_1 and λ_2, is increasing in ξ. As shown, e.g., by Griffiths et al. (1979), it is, however, possible for a pair (X_1, X_2) of Poisson random variables to exhibit negative dependence.

By convention, let a Poisson distribution with rate zero, denoted $\mathcal{P}(0)$, be a point mass at 0. Clearly X_1 and X_2 are independent when $\xi = 0$, i.e., when $Z \equiv 0$, and exhibit the highest degree of association (permissible in this model specification) when ξ attains its maximum, namely $\min(\lambda_1, \lambda_2)$. Thus, in model (2.1), the maximum correlation is

$$\frac{\min(\lambda_1, \lambda_2)}{\sqrt{\lambda_1, \lambda_2}}. \tag{2.3}$$

However, Griffiths et al. (1979) show that for an arbitrary pair of Poisson random variables with fixed marginal parameters λ_1 and λ_2, the maximum correlation is given by

$$\rho_{\max}(\lambda_1, \lambda_2) = \frac{1}{\sqrt{\lambda_1 \lambda_2}} \left[-\lambda_1 \lambda_2 + \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \min\{\bar{G}_{\lambda_1}(i), \bar{G}_{\lambda_2}(j)\} \right], \tag{2.4}$$

where $\mathbb{N} = \{0, 1, \ldots\}$ and G_λ and $\bar{G}_\lambda$ respectively denote the cumulative distribution function and corresponding survival function of a Poisson random variable with mean λ. Similarly, the minimum possible correlation is

$$\rho_{\min}(\lambda_1, \lambda_2) = \frac{1}{\sqrt{\lambda_1 \lambda_2}} \left[-\lambda_1 \lambda_2 - \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \min\{0, G_{\lambda_1}(i) + G_{\lambda_2}(j) - 1\} \right]. \tag{2.5}$$

Thus, for an arbitrary pair of Poisson random variables with marginal means λ_1 and λ_2, the classical bivariate Poisson model constricts the correlation to lie in the interval $[0, \min(\lambda_1, \lambda_2/\sqrt{\lambda_1, \lambda_2})]$ while the full range of permissible correlations is in fact the widened interval given by $[\rho_{\min}, \rho_{\max}]$.

Note that in the trivial case where the marginal parameters coincide, the range of permissible correlation values implied by the classical bivariate Poisson model agrees with that given by Griffiths et al. (1979). Indeed, setting $\lambda_1 = \lambda_2 = \lambda$ implies that $X_1 =_d X_2$, i.e., X_1 and X_2 have the same distribution, and the maximum possible correlation reduces to 1, which is equal to $\rho_{\max}(\lambda, \lambda)$.

Several alternative definitions of both bivariate and multivariate Poisson distributions have been proposed in the literature, an overview of which will be given in subsequent sections. Some of these specifications allow for a more flexible dependence structure, wherein negative dependence can be accommodated. In this thesis, yet another bivariate Poisson model is proposed, along with a multivariate extension. The proposed model construction relies on the convolution of independent Poisson random variables with a comonotonic shock. Unlike the classical model wherein a single shock variable induces the dependence, the proposed bivariate Poisson formulation uses a pair of perfectly positive dependent (comonotonic) Poisson random variables. This extension allows for the components to exhibit the full range of possible positive dependence, from 0 to $\rho_{\max}$. An analogous representation of a bivariate Poisson model exhibiting negative dependence is also possible using this method by considering a counter-monotonic shock rather than a comonotonic one.

In building up towards the specification of a more flexible bivariate Poisson model, the remainder of this section is dedicated to outlining the main distributional properties of the classical common shock model as well as various estimation approaches that have been suggested in the literature.

2.1.2 Model properties

The joint probability mass function, which will be denoted as $h_{\Lambda,\xi}$ for $\Lambda = (\lambda_1, \lambda_2)$, can be derived by conditioning on the underlying shock variable. For any $x_1, x_2 \in \mathbb{N}$, it follows that

$$\begin{aligned}
h_{\Lambda,\xi}(x_1, x_2) &= \Pr(X_1 = x_1, X_2 = x_2) \\
&= \sum_{z=0}^{\min(x_1,x_2)} \Pr(X_1 = x_1, X_2 = x_2 \mid Z = z) \Pr(Z = z) \\
&= \sum_{z=0}^{\min(x_1,x_2)} \Pr(Y_1 = x_1 - z) \Pr(Y_2 = x_2 - z) \Pr(Z = z) \\
&= \sum_{z=0}^{\min(x_1,x_2)} g_{\lambda_1-\xi}(x_1 - z) g_{\lambda_2-\xi}(x_2 - z) g_{\xi}(z) \\
&= \sum_{z=0}^{\min(x_1,x_2)} \frac{e^{-(\lambda_1-\xi)}(\lambda_1 - \xi)^{x_1-z} e^{-(\lambda_2-\xi)}(\lambda_2 - \xi)^{x_2-z} e^{-\xi}\xi^z}{(x_1 - z)!(x_2 - z)!z!} \\
&= e^{-(\lambda_1+\lambda_2-\xi)} \sum_{z=0}^{\min(x_1,x_2)} \frac{(\lambda_1 - \xi)^{x_1-z}(\lambda_2 - \xi)^{x_2-z}\xi^z}{(x_1 - z)!(x_2 - z)!z!},
\end{aligned}$$

where g_λ denotes the probability mass function of a Poisson random variable with mean λ. The corresponding cumulative distribution function (CDF), denoted $H_{\Lambda,\xi}$, is established in a similar way, or by directly working with the probability mass function, viz.

$$\begin{aligned}
H_{\Lambda,\xi}(x_1, x_2) &= \Pr(X_1 \leqslant x_1, X_2 \leqslant x_2) \\
&= \sum_{z=0}^{\min(x_1,x_2)} \Pr(X_1 \leqslant x_1, X_2 \leqslant x_2 \mid Z = z) \Pr(Z = z) \\
&= \sum_{z=0}^{\min(x_1,x_2)} \Pr(Y_1 \leqslant x_1 - z) \Pr(Y_2 \leqslant x_2 - z) \Pr(Z = z) \\
&= \sum_{z=0}^{\min(x_1,x_2)} G_{\lambda_1-\xi}(x_1 - z) G_{\lambda_2-\xi}(x_2 - z) g_\xi(z).
\end{aligned}$$

It is well known that the univariate Poisson distribution has the simple recurrence relation:

$$g_\lambda(x) = \frac{\lambda}{x} g_\lambda(x-1), \quad x \in \{1, 2, \ldots\}.$$

Since the common shock bivariate Poisson model formulation is based on the sum of independent Poisson random variables, this recurrence relation remains relevant. As shown, e.g., by Holgate (1964) and Kawamura (1985), for the classical bivariate Poisson distribution the following recurrence relations hold:

$$\begin{aligned}
x_1 h_{\Lambda,\xi}(x_1, x_2) &= (\lambda_1 - \xi) h_{\Lambda,\xi}(x_1 - 1, x_2) + \xi h_{\Lambda,\xi}(x_1 - 1, x_2 - 1) \\
x_2 h_{\Lambda,\xi}(x_1, x_2) &= (\lambda_2 - \xi) h_{\Lambda,\xi}(x_1, x_2 - 1) + \xi h_{\Lambda,\xi}(x_1 - 1, x_2 - 1).
\end{aligned} \tag{2.6}$$

The probability generating function and moment generating function of the classical bivariate Poisson model are easy to derive since both components X_1 and X_2 are the sums of independent Poisson random variables. The probability generating function is found to be

$$\begin{aligned}
\mathrm{E}(s_1^{X_1} s_2^{X_2}) &= \mathrm{E}(s_1^{Y_1+Z} s_2^{Y_2+Z}) = \mathrm{E}(s_1^{Y_1}) \mathrm{E}(s_2^{Y_2}) \mathrm{E}\{(s_1 s_2)^Z\} \\
&= \exp\{(\lambda_1 - \xi)(s_1 - 1) + (\lambda_2 - \xi)(s_2 - 1) + \xi(s_1 s_2 - 1)\}
\end{aligned}$$

and the moment generating function is

$$\begin{aligned}
M_{X_1,X_2}(t_1, t_2) &= \mathrm{E}(e^{t_1 X_1 + t_2 X_2}) = \mathrm{E}\{e^{t_1 Y_1 + t_2 Y_2 + (t_1+t_2)Z}\} \\
&= \mathrm{E}(e^{t_1 Y_1}) \mathrm{E}(e^{t_2 Y_2}) \mathrm{E}\{e^{(t_1+t_2)Z}\} \\
&= \exp\{(\lambda_1 - \xi)(e^{t_1} - 1) + (\lambda_2 - \xi)(e^{t_2} - 1) + \xi(e^{t_1+t_2} - 1)\}.
\end{aligned}$$

Working with the moment generating function, it can easily be seen that the classical bivariate

Poisson model is closed under convolution. Suppose that $(X_{11}, X_{12}) \sim H_{\Lambda^\dagger,\xi^\dagger}$ is independent of $(X_{21}, X_{22}) \sim H_{\Lambda^*,\xi^*}$. It then follows that $(X_{11}+X_{21}, X_{12}+X_{22})$ has moment generating function given by

$$\exp\{(\lambda_1^\dagger - \xi^\dagger + \lambda_1^* - \xi^*)(e^{t_1}-1) + (\lambda_2^\dagger - \xi^\dagger + \lambda_2^* - \xi^*)(e^{t_2}-1) + (\xi^\dagger + \xi^*)(e^{t_1+t_2}-1)\}.$$

Thus, the pair $(X_{11}+X_{21}, X_{12}+X_{22})$ has distribution $H_{\Lambda,\xi}$ where $\lambda_s = \lambda_s^\dagger + \lambda_s^*$, for $s \in \{1,2\}$ and $\xi = \xi^\dagger + \xi^*$.

The construction in (2.1) relies on a trivariate set of mutually independent random variables, wherein the source of dependence between the margins X_1 and X_2 is induced by the common shock variable Z. Surely, as the marginal mean of the common shock variable, ξ, increases, the degree of dependence increases accordingly. This statement will be made formal in the following lemma, which establishes the Positive Quadrant Dependence (PQD) ordering of the classical common shock model and will be denoted in terms of the inequality $\prec_{PQD}$. This property of the classical bivariate Poisson model is also shown in a similar manner by Bouezmarni et al. (2009).

Lemma 2.1 *Suppose* $(X_1, X_2) \sim H_{\Lambda,\xi}$ *and* $(X_1', X_2') \sim H_{\Lambda,\xi'}$. *Then* $\xi < \xi' \Rightarrow (X_1, X_2) \underset{PQD}{\prec} (X_1', X_2')$.

Proof. The proof follows from the fact that for any fixed $(x_1, x_2) \in \mathbb{N}^2$ the CDF $H_{\Lambda,\xi}(x_1,x_2)$ is an increasing function of the dependence parameter ξ. Differentiating $H_{\Lambda,\xi}(x_1,x_2)$ with respect to ξ yields the following:

$$\begin{aligned}
\frac{\partial}{\partial \xi} H_{\Lambda,\xi}(x_1,x_2) &= \frac{\partial}{\partial \xi} \sum_{z=0}^{\min(x_1,x_2)} G_{\lambda_1-\xi}(x_1-z) G_{\lambda-\xi}(x_2-z) g_\xi(z) \\
&= \sum_{z=0}^{\min(x_1,x_2)} \Bigg[\left\{ \frac{\partial}{\partial \xi} G_{\lambda_1-\xi}(x_1-z) \right\} G_{\lambda-\xi}(x_2-z) g_\xi(z) \\
&\quad + G_{\lambda_1-\xi}(x_1-z) \left\{ \frac{\partial}{\partial \xi} G_{\lambda-\xi}(x_2-z) \right\} g_\xi(z) \\
&\quad + G_{\lambda_1-\xi}(x_1-z) G_{\lambda-\xi}(x_2-z) \left\{ \frac{\partial}{\partial \xi} g_\xi(z) \right\} \Bigg].
\end{aligned}$$

For a fixed $n \in \mathbb{N}$, it is straightforward to show that

$$\frac{\partial}{\partial \xi} G_{\lambda_s-\xi}(n) = G_{\lambda_s-\xi}(n) - G_{\lambda_s-\xi}(n-1),$$

for $s \in \{1,2\}$, and

$$\frac{\partial}{\partial \xi} g_\xi(n) = -g_\xi(n) + g_\xi(n-1).$$

Putting everything together, one obtains

$$\begin{aligned}
\frac{\partial}{\partial \xi} H_{\Lambda,\xi}(x_1, x_2) = & \sum_{z=0}^{\min(x_1,x_2)} [G_{\lambda_1-\xi}(x_1 - z)G_{\lambda-\xi}(x_2 - z)g_\xi(z) \\
& - G_{\lambda_1-\xi}(x_1 - 1 - z)G_{\lambda-\xi}(x_2 - z)g_\xi(z) \\
& - G_{\lambda_1-\xi}(x_1 - z)G_{\lambda-\xi}(x_2 - 1 - z)g_\xi(z) \\
& + G_{\lambda_1-\xi}(x_1 - z)G_{\lambda-\xi}(x_2 - z)g_\xi(z - 1)] \\
= & H_{\Lambda,\xi}(x_1, x_2) - \sum_{z=0}^{\min\{x_1-1,x_2\}} G_{\lambda_1-\xi}(x_1 - 1 - z)G_{\lambda-\xi}(x_2 - z)g_\xi(z) \\
& - \sum_{z=0}^{\min\{x_1,x_2-1\}} G_{\lambda_1-\xi}(x_1 - z)G_{\lambda-\xi}(x_2 - 1 - z)g_\xi(z) \\
& + \sum_{z^*=0}^{\min\{x_1-1,x_2-1\}} G_{\lambda_1-\xi}(x_1 - 1 - z^*)G_{\lambda-\xi}(x_2 - 1 - z^*)g_\xi(z^*) \\
= & H_{\Lambda,\xi}(x_1, x_2) - H_{\Lambda,\xi}(x_1 - 1, x_2) - H_{\Lambda,\xi}(x_1, x_2 - 1) + H_{\Lambda,\xi}(x_1 - 1, x_2 - 1) \\
= & h_{\Lambda,\xi}(x_1, x_2) \geqslant 0.
\end{aligned}$$

Thus, $H_{\Lambda,\xi}$ is an increasing function of ξ and so for $\xi < \xi'$ one has that

$$H_{\Lambda,\xi}(i, j) \leqslant H_{\Lambda,\xi'}(i, j), \quad \forall (i, j) \in \mathbb{N}^2,$$

which is the desired result. □

2.1.3 Estimation

Let $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ be a random sample of size n from the classical bivariate Poisson distribution. It is then of interest to obtain estimates of the marginal parameters $\Lambda = (\lambda_1, \lambda_2)$ as well as the dependence parameter ξ. There are various estimation techniques that have been suggested, three of which will be detailed in what follows.

Estimation based on the method of moments

The method of moments estimation approach yields parameter estimates by solving the system of equations that sets the theoretical moments equal to the sample moments. For the common shock model, this reduces to the following:

$$\mathrm{E}(X_1) = \lambda_1 = \bar{X}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{i1}, \quad \mathrm{E}(X_2) = \lambda_2 = \bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} X_{i2},$$

and

$$\operatorname{cov}(X_1, X_2) = \xi = S_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2),$$

subject to the constraint $\xi \in [0, \min(\lambda_1, \lambda_2)]$.

Estimating the marginal Poisson rates by their respective sample means coincides with maximum likelihood estimation in the case of independence, i.e., when $\xi = 0$. Marginally, for $s \in \{1, 2\}$, $X_{s1}, \ldots, X_{sn}$ is a random sample from a univariate Poisson distribution. Standard maximum likelihood theory then implies that $\bar{X}_s$ is a consistent estimator for λ_s and

$$\sqrt{n}\,(\bar{X}_s - \lambda_s) \rightsquigarrow \mathcal{N}(0, \mathcal{I}^{-1})$$

where $\rightsquigarrow$ denotes convergence in distribution and $\mathcal{I}^{-1} = \lambda_s$ is the inverse of the Fisher Information, which can be consistently estimated by $\bar{X}_s$. Note that the asymptotic Normality of the estimator $\bar{X}_s$ also follows directly from the Central Limit Theorem.

As shown in, e.g., Theorem 8 on p. 52 of Ferguson (1996), the sample covariance S_{12} is a consistent estimator of the theoretical covariance. Asymptotically, one has

$$\sqrt{n}\,(S_{12} - \xi) \rightsquigarrow \mathcal{N}\{0, \sigma^2(\xi, \lambda_1, \lambda_2)\},$$

where $\sigma^2(\xi, \lambda_1, \lambda_2) = \operatorname{var}\{(X_1 - \lambda_1)(X_2 - \lambda_2)\}$. Applying these results to the classical bivariate Poisson model, one then has that the method of moments estimate S_{12} for ξ is consistent and asymptotically Gaussian as long as $S_{12} \in [0, \min(\bar{X}_1, \bar{X}_2)]$. Moreover, it can easily be shown that the asymptotic variance $\sigma^2(\xi, \lambda_1, \lambda_2)$ simplifies to

$$\lambda_1\lambda_2 + \xi(\xi + 1).$$

Indeed, for $(X_1, X_2) \sim H_{\Lambda,\xi}$, write $(X_1 - \lambda_1)(X_2 - \lambda_2) = (\mathring{Y}_1 + \mathring{Z})(\mathring{Y}_2 + \mathring{Z})$, where $\mathring{Y}_1$, $\mathring{Y}_2$ and $\mathring{Z}$ denote the centred counterparts of Y_1, Y_2 and Z, respectively. Specifically, $\mathring{Y}_s = Y_s - (\lambda_s - \xi)$, for $s \in \{1, 2\}$, and $\mathring{Z} = Z - \xi$. Clearly, $\mathring{Y}_1$, $\mathring{Y}_2$ and $\mathring{Z}$ are mutually independent and each have expectation zero. One then has

$$\begin{aligned}
\operatorname{var}\{(X_1 - \lambda_1)(X_2 - \lambda_2)\} &= \operatorname{var}\{(\mathring{Y}_1 + \mathring{Z})(\mathring{Y}_2 + \mathring{Z})\} \\
&= \operatorname{var}\{\mathring{Y}_1\mathring{Y}_2 + \mathring{Y}_1\mathring{Z} + \mathring{Y}_2\mathring{Z} + \mathring{Z}^2\} \\
&= \mathrm{E}(\mathring{Y}_1^2)\mathrm{E}(\mathring{Y}_2^2) + \mathrm{E}(\mathring{Y}_1^2)\mathrm{E}(\mathring{Z}^2) + \mathrm{E}(\mathring{Y}_2^2)\mathrm{E}(\mathring{Z}^2) + \mathrm{E}(\mathring{Z}^4) - \{\mathrm{E}(\mathring{Z}^2)\}^2.
\end{aligned}$$

For $s \in \{1, 2\}$, $\mathrm{E}(\mathring{Y}_s^2) = \operatorname{var}(Y_s) = \lambda_s - \xi$. Additionally, $\mathrm{E}(\mathring{Z}^2) = \operatorname{var}(Z) = \xi$ and $\mathrm{E}(\mathring{Z}^4) = \xi(3\xi + 1)$. Then, upon simplification, one obtains the desired result, viz.

$$\operatorname{var}\{(X_1 - \lambda_1)(X_2 - \lambda_2)\} = \lambda_1\lambda_2 + \xi(\xi + 1).$$

Note that when S_{12} is significantly smaller than 0, it may be wiser to consider a different model altogether because the classical shock model cannot accommodate negative dependence.

Maximum likelihood estimation

Recall that the classical bivariate Poisson probability mass function has the form

$$h_{\Lambda,\xi}(x_1, x_2) = \sum_{z=0}^{\min(x_1,x_2)} g_{\lambda_1-\xi}(x_1 - z) g_{\lambda_2-\xi}(x_2 - z) g_\xi(z).$$

As was previously shown, for any $n \in \mathbb{N}$, one has

$$\frac{\partial}{\partial \xi} g_\xi(n) = -g_\xi(n) + g_\xi(n-1).$$

In a similar way, it can be shown that for $s \in \{1, 2\}$,

$$\frac{\partial}{\partial \xi} g_{\lambda_s-\xi}(n) = g_{\lambda-\xi}(n) - g_{\lambda-\xi}(n-1)$$
$$\frac{\partial}{\partial \lambda_s} g_{\lambda_s-\xi}(n) = -g_{\lambda-\xi}(n) + g_{\lambda-\xi}(n-1).$$

Accordingly, differentiating the probability mass function with respect to λ_1 yields the following:

$$\begin{aligned}\frac{\partial}{\partial \lambda_1} h_{\Lambda,\xi}(x_1, x_2) &= \sum_{z=0}^{\min(x_1,x_2)} \left\{ \frac{\partial}{\partial \lambda_1} g_{\lambda_1-\xi}(x_1 - z) \right\} g_{\lambda_2-\xi}(x_2 - z) g_\xi(z) \\ &= \sum_{z=0}^{\min(x_1,x_2)} g_{\lambda_1-\xi}(x_1 - 1 - z) g_{\lambda_2-\xi}(x_2 - z) g_\xi(z) \\ &\quad - \sum_{z=0}^{\min(x_1,x_2)} g_{\lambda_1-\xi}(x_1 - z) g_{\lambda_2-\xi}(x_2 - z) g_\xi(z),\end{aligned}$$

and hence

$$\begin{aligned}\frac{\partial}{\partial \lambda_1} h_{\Lambda,\xi}(x_1, x_2) &= \sum_{z=0}^{\min\{x_1-1,x_2\}} g_{\lambda_1-\xi}(x_1 - 1 - z) g_{\lambda_2-\xi}(x_2 - z) g_\xi(z) \\ &\quad - \sum_{z=0}^{\min(x_1,x_2)} g_{\lambda_1-\xi}(x_1 - z) g_{\lambda_2-\xi}(x_2 - z) g_\xi(z) \\ &= h_{\Lambda,\xi}(x_1 - 1, x_2) - h_{\Lambda,\xi}(x_1, x_2).\end{aligned}$$

In the same way, the derivative with respect to λ_2 yields

$$\frac{\partial}{\partial \lambda_2} h_{\Lambda,\xi}(x_1, x_2) = h_{\Lambda,\xi}(x_1, x_2 - 1) - h_{\Lambda,\xi}(x_1, x_2).$$

The partial derivative of $h_{\Lambda,\xi}(x_1, x_2)$ with respect to the dependence parameter gives

$$\frac{\partial}{\partial \xi} h_{\Lambda,\xi}(x_1, x_2) = \sum_{z=0}^{\min(x_1,x_2)} \left[\left\{ \frac{\partial}{\partial \xi} g_{\lambda_1-\xi}(x_1 - z) \right\} g_{\lambda_2-\xi}(x_2 - z) g_\xi(z) \right.$$
$$+ g_{\lambda_1-\xi}(x_1 - z) \left\{ \frac{\partial}{\partial \xi} g_{\lambda_2-\xi}(x_2 - z) \right\} g_\xi(z)$$
$$\left. + g_{\lambda_1-\xi}(x_1 - z) g_{\lambda_2-\xi}(x_2 - z) \left\{ \frac{\partial}{\partial \xi} g_\xi(z) \right\} \right],$$

which can be rewritten as

$$\frac{\partial}{\partial \xi} h_{\Lambda,\xi}(x_1, x_2) = \sum_{z=0}^{\min(x_1,x_2)} [\{g_{\lambda_1-\xi}(x_1 - z) - g_{\lambda_1-\xi}(x_1 - 1 - z)\} g_{\lambda_2-\xi}(x_2 - z) g_\xi(z)$$
$$+ g_{\lambda_1-\xi}(x_1 - z) \{g_{\lambda_2-\xi}(x_2 - z) - g_{\lambda_2-\xi}(x_2 - 1 - z)\} g_\xi(z)$$
$$+ g_{\lambda_1-\xi}(x_1 - z) g_{\lambda_2-\xi}(x_2 - z) \{g_\xi(z - 1) - g_\xi(z)\}]$$
$$= h_{\Lambda,\xi}(x_1, x_2) - h_{\Lambda,\xi}(x_1 - 1, x_2) + h_{\Lambda,\xi}(x_1, x_2) - h_{\Lambda,\xi}(x_1, x_2 - 1)$$
$$- h_{\Lambda,\xi}(x_1, x_2) + h_{\Lambda,\xi}(x_1 - 1, x_2 - 1)$$
$$= h_{\Lambda,\xi}(x_1, x_2) - h_{\Lambda,\xi}(x_1 - 1, x_2) - h_{\Lambda,\xi}(x_1, x_2 - 1) + h_{\Lambda,\xi}(x_1 - 1, x_2 - 1).$$

An alternative formulation for the above partial derivatives can be derived using the recurrence relations given in (2.6). In particular, one has

$$\frac{\partial}{\partial \lambda_1} h_{\Lambda,\xi}(x_1, x_2) = h_{\Lambda,\xi}(x_1 - 1, x_2) - h_{\Lambda,\xi}(x_1, x_2)$$
$$= \left\{ \frac{x_1 h_{\Lambda,\xi}(x_1, x_2) - \xi h_{\Lambda,\xi}(x_1 - 1, x_2 - 1)}{(\lambda_1 - \xi)} \right\} - h_{\Lambda,\xi}(x_1, x_2),$$
$$\frac{\partial}{\partial \lambda_2} h_{\Lambda,\xi}(x_1, x_2) = h_{\Lambda,\xi}(x_1, x_2 - 1) - h_{\Lambda,\xi}(x_1, x_2)$$
$$= \left\{ \frac{x_2 h_{\Lambda,\xi}(x_1, x_2) - \xi h_{\Lambda,\xi}(x_1 - 1, x_2 - 1)}{(\lambda_2 - \xi)} \right\} - h_{\Lambda,\xi}(x_1, x_2),$$

$$\frac{\partial}{\partial \xi} h_{\Lambda,\xi}(x_1, x_2) = h_{\Lambda,\xi}(x_1, x_2) - h_{\Lambda,\xi}(x_1 - 1, x_2) - h_{\Lambda,\xi}(x_1, x_2 - 1) + h_{\Lambda,\xi}(x_1 - 1, x_2 - 1)$$
$$= h_{\Lambda,\xi}(x_1, x_2) - \{x_1/(\lambda_1 - \xi)\} h_{\Lambda,\xi}(x_1, x_2) + \{\xi/(\lambda_1 - \xi)\} h_{\Lambda,\xi}(x_1 - 1, x_2 - 1)$$
$$- \{x_2/(\lambda_2 - \xi)\} h_{\Lambda,\xi}(x_1, x_2) + \{\xi/(\lambda_2 - \xi)\} h_{\Lambda,\xi}(x_1 - 1, x_2 - 1)$$
$$+ h_{\Lambda,\xi}(x_1 - 1, x_2 - 1),$$

so that

$$\frac{\partial}{\partial \xi} h_{\Lambda,\xi}(x_1, x_2) = \left\{1 - \frac{x_1}{(\lambda_1 - \xi)} - \frac{x_2}{(\lambda_2 - \xi)}\right\} h_{\Lambda,\xi}(x_1, x_2) + \left\{1 + \frac{\xi}{(\lambda_1 - \xi)} + \frac{\xi}{(\lambda_2 - \xi)}\right\} h_{\Lambda,\xi}(x_1 - 1, x_2 - 1)$$

For a random sample $(X_{11}, X_{12}) \ldots, (X_{n1}, X_{n2})$ from the classical bivariate Poisson distribution, the log-likelihood has the form

$$\ell(\lambda_1, \lambda_2, \xi) = \sum_{i=1}^{n} \ln h_{\Lambda,\xi}(x_{i1}, x_{i2}).$$

Using the partial derivatives derived above, one obtains the following score equations:

$$\begin{aligned}\frac{\partial}{\partial \lambda_1} \ell(\lambda_1, \lambda_2, \xi) &= \sum_{i=1}^{n} \frac{1}{h_{\Lambda,\xi}(x_{i1}, x_{i2})} \left\{\frac{x_{i1} h_{\Lambda,\xi}(x_{i1}, x_{i2}) - \xi h_{\Lambda,\xi}(x_{i1} - 1, x_{i2} - 1)}{(\lambda_1 - \xi)} - h_{\Lambda,\xi}(x_{i1}, x_{i2})\right\} \\ &= \left(\frac{n\bar{x}_1}{\lambda_1 - \xi}\right) - n - \left(\frac{\xi}{\lambda_1 - \xi}\right) \sum_{i=1}^{n} \frac{h_{\Lambda,\xi}(x_{i1} - 1, x_{i2} - 1)}{h_{\Lambda,\xi}(x_{i1}, x_{i2})},\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial \lambda_2} \ell(\lambda_1, \lambda_2, \xi) &= \sum_{i=1}^{n} \frac{1}{h_{\Lambda,\xi}(x_{i1}, x_{i2})} \left\{\frac{x_2 h_{\Lambda,\xi}(x_{i1}, x_{i2}) - \xi h_{\Lambda,\xi}(x_{i1} - 1, x_{i2} - 1)}{(\lambda_2 - \xi)} - h_{\Lambda,\xi}(x_{i1}, x_{i2})\right\} \\ &= \left(\frac{n\bar{x}_2}{\lambda_2 - \xi}\right) - n - \left(\frac{\xi}{\lambda_2 - \xi}\right) \sum_{i=1}^{n} \frac{h_{\Lambda,\xi}(x_{i1} - 1, x_{i2} - 1)}{h_{\Lambda,\xi}(x_{i1}, x_{i2})},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial \xi} \ell(\lambda_1, \lambda_2, \xi) &= \sum_{i=1}^{n} \frac{1}{h_{\Lambda,\xi}(x_{i1}, x_{i2})} \left[\left\{1 - \frac{x_{i1}}{(\lambda_1 - \xi)} - \frac{x_{i2}}{(\lambda_2 - \xi)}\right\} h_{\Lambda,\xi}(x_{i1}, x_{i2}) \right. \\ &\quad \left. + \left\{1 + \frac{\xi}{(\lambda_1 - \xi)} + \frac{\xi}{(\lambda_2 - \xi)}\right\} h_{\Lambda,\xi}(x_{i1} - 1, x_{i2} - 1)\right] \\ &= n - \left(\frac{n\bar{x}_1}{\lambda_1 - \xi}\right) - \left(\frac{n\bar{x}_2}{\lambda_2 - \xi}\right) \\ &\quad + \left\{\frac{\xi}{\lambda_1 - \xi} + \frac{\xi}{\lambda_2 - \xi} + 1\right\} \sum_{i=1}^{n} \frac{h_{\Lambda,\xi}(x_{i1} - 1, x_{i2} - 1)}{h_{\Lambda,\xi}(x_{i1}, x_{i2})}.\end{aligned}$$

Introduce

$$\bar{R} = \frac{1}{n} \sum_{i=1}^{n} \frac{h_{\Lambda,\xi}(x_{i1} - 1, x_{i2} - 1)}{h_{\Lambda,\xi}(x_{i1}, x_{i2})}.$$

Then, the score equations for the common shock bivariate Poisson model reduce to

$$\frac{\partial}{\partial \lambda_1} \ell(\lambda_1, \lambda_2, \xi) = n \left\{ \left(\frac{\bar{x}_1}{\lambda_1 - \xi} \right) - 1 - \left(\frac{\xi}{\lambda_1 - \xi} \right) \bar{R} \right\} = 0, \tag{2.7}$$

$$\frac{\partial}{\partial \lambda_2} \ell(\lambda_1, \lambda_2, \xi) = n \left\{ \left(\frac{\bar{x}_2}{\lambda_2 - \xi} \right) - 1 - \left(\frac{\xi}{\lambda_2 - \xi} \right) \bar{R} \right\} = 0, \tag{2.8}$$

$$\frac{\partial}{\partial \xi} \ell(\lambda_1, \lambda_2, \xi) = n \left\{ 1 - \left(\frac{\bar{x}_1}{\lambda_1 - \xi} \right) - \left(\frac{\bar{x}_2}{\lambda_2 - \xi} \right) + \left(\frac{\xi}{\lambda_1 - \xi} + \frac{\xi}{\lambda_2 - \xi} + 1 \right) \bar{R} \right\} = 0. \tag{2.9}$$

Equations (2.7) and (2.8) imply that

$$\left(\frac{\xi}{\lambda_1 - \xi} \right) \bar{R} = \left(\frac{\bar{x}_1}{\lambda_1 - \xi} \right) - 1, \quad \left(\frac{\xi}{\lambda_2 - \xi} \right) \bar{R} = \left(\frac{\bar{x}_2}{\lambda_1 - \xi} \right) - 1.$$

Using this, (2.9) simplifies to $\bar{R} = 1$. Then, setting $\bar{R} = 1$ in equations (2.7) and (2.8) yields maximum likelihood estimates

$$\hat{\lambda}_1 = \bar{x}_1, \quad \hat{\lambda}_2 = \bar{x}_2.$$

The maximum likelihood estimate (MLE) for the dependence parameter is then found by solving for the value of ξ such that $\bar{R} = 1$, subject to the constraints that $\xi \in [0, \min(\bar{x}_1, \bar{x}_2)]$. This can be solve iteratively, using some numerical procedure.

Holgate (1964) shows that the covariance matrix for the maximum likelihood estimators $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\xi}$ is given by

$$\frac{1}{n} \begin{bmatrix} \lambda_1 & \xi & \xi \\ \xi & \lambda_2 & \xi \\ \xi & \xi & \frac{\xi^2(\lambda_1+\lambda_2-2\xi)-\xi^2+(\lambda_1-2\xi)(\lambda_2-2\xi)}{(\lambda_1\lambda_2-\xi^2)(Q-1)-(\lambda_1+\lambda_2-2\xi)} \end{bmatrix}$$

where

$$Q = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} \frac{h^2_{\Lambda,\xi}(x_1 - 1, x_2 - 1)}{h_{\Lambda,\xi}(x_1, x_2)}.$$

EM Algorithm

The stochastic representation of the classical bivariate Poisson model, given in (2.1), relies on the convolution of unobserved random variables. This construction makes the EM algorithm an appealing approach for determining the maximum likelihood estimates. This method was explored in the case of the bivariate Poisson model by Adamidis and Loukas (1994) and a multivariate common shock Poisson model by Karlis (2003). Here, we consider the simple case where iid pairs $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ are observed and the latent variables are taken to be the common shocks $Z_1, \ldots, Z_n$. Note that Adamidis and Loukas (1994) outline the EM algorithm procedure in the more general setting where missing values could also occur in the pairs (X_{i1}, X_{i2}) for certain $i \in \{1, \ldots, n\}$.

Let the complete data consist of the set of iid trivariate vectors (X_{i1}, X_{i2}, Z_i), while the observed data consists of the pairs $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$. For any $i \in \{1, \ldots, n\}$, one has

$$\begin{aligned}\Pr(X_1 = x_{i1}, X_2 = x_{i2}, Z = z_i) &= \Pr(X_1 = x_{i1}, X_2 = x_{i2} \mid Z = z_i) \Pr(Z = z_i) \\ &= g_{\lambda_1 - \xi}(x_{i1} - z_i) g_{\lambda_2 - \xi}(x_{i2} - z) g_{\xi}(z_i).\end{aligned}$$

It follows that the complete data log-likelihood is given by

$$\begin{aligned}\ell^C(\lambda_1, \lambda_2, \xi) &= \sum_{i=1}^{n} [-(\lambda_1 - \xi) + (x_{i1} - z_i) \ln(\lambda_1 - \xi) - \ln\{(x_{i1} - z_i)!\} \\ &\qquad - (\lambda_2 - \xi) + (x_{i2} - z_i) \ln(\lambda_2 - \xi) - \ln\{(x_{i2} - z_i)!\} \\ &\qquad\qquad - \xi + z_i \ln(\xi) - \ln(z_i!)] \\ &= -n(\lambda_1 + \lambda_2 - \xi) + n(\bar{x}_1 - \bar{z}) \ln(\lambda_1 - \xi) + n(\bar{x}_2 - \bar{z}) \ln(\lambda_2 - \xi) \\ &\qquad + \bar{z} \ln(\xi) - \sum_{i=1}^{n} \ln\{(x_{i1} - z_i)!(x_{i2} - z_i)!z_i!\}.\end{aligned}$$

Denote the parameter vector $(\lambda_1, \lambda_2, \xi)$ by Ψ and let $\Psi^{(k)}$ denote its estimated value at the kth iteration of the algorithm. Taking the conditional expectation of the complete data log-likelihood, given the observed data $\mathbf{X} = \{(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})\}$ and $\Psi^{(k)}$, the E-step yields

$$\begin{aligned}Q(\Psi; \Psi^{(k)}) &= \mathrm{E}\left\{\ell^C(\lambda_1, \lambda_2, \xi) \mid \mathbf{X}, \Psi^{(k)}\right\} \\ &= -n(\lambda_1 + \lambda_2 - \xi) + n\left\{\bar{x}_1 - \bar{q}(\Psi^{(k)})\right\} \ln(\lambda_1 - \xi) \\ &\qquad + n\left\{\bar{x}_2 - \bar{q}(\Psi^{(k)})\right\} \ln(\lambda_2 - \xi) + \bar{q}(\Psi^{(k)}) \ln(\xi) - R(\mathbf{X}, \Psi^{(k)}),\end{aligned}$$

where

$$\bar{q}(\Psi^{(k)}) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}(Z_i \mid x_{i1}, x_{i2}, \Psi^{(k)})$$

and

$$R(\mathbf{X}, \Psi^{(k)}) = \sum_{i=1}^{n} \mathrm{E}\left\{\ln\{(x_{i1} - z_i)!(x_{i2} - z_i)!z_i!\} \mid x_{i1}, x_{i2}, \Psi^{(k)}\right\}$$

is the remaining term, which does not depend on the unknown parameters λ_1, λ_2, ξ. The maximization step then leads to the following system of equations

$$\frac{\partial}{\partial \lambda_1} Q(\Psi; \Psi^{(k)}) = n\left[\frac{\{\bar{x}_1 - \bar{q}(\Psi^{(k)})\}}{(\lambda_1 - \xi)} - 1\right] = 0 \tag{2.10}$$

$$\frac{\partial}{\partial \lambda_2} Q(\Psi; \Psi^{(k)}) = n\left[\frac{\{\bar{x}_2 - \bar{q}(\Psi^{(k)})\}}{(\lambda_2 - \xi)} - 1\right] = 0, \tag{2.11}$$

and

$$\frac{\partial}{\partial \xi} Q(\Psi; \Psi^{(k)}) = n \left[1 - \frac{\{\bar{x}_1 - \bar{q}(\Psi^{(k)})\}}{(\lambda_1 - \xi)} - \frac{\{\bar{x}_2 - \bar{q}(\Psi^{(k)})\}}{(\lambda_2 - \xi)} + \frac{\bar{q}(\Psi^{(k)})}{\xi} \right] = 0. \tag{2.12}$$

The solution to (2.10) through (2.12) yields the following parameter updates:

$$\xi^{(k+1)} = \bar{q}(\Psi^{(k)}), \quad \lambda_1^{(k+1)} = \bar{x}_1, \quad \lambda_2^{(k+1)} = \bar{x}_2.$$

Thus, throughout the EM updates, the parameter estimates for the marginal parameters are held fixed at their respective MLEs, $\bar{x}_1$ and $\bar{x}_2$, while the estimate for the dependence parameter is updated according to $\bar{q}(\Psi^{(k)})$ until convergence. That is, the EM algorithm will continue to update the dependence parameter ξ until

$$\xi = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}(Z_i \mid x_{i1}, x_{i2}, \Psi),$$

wherein (λ_1, λ_2) is held fixed at the MLEs $(\bar{x}_1, \bar{x}_2)$.

Observe that the form of the EM updates ensure that at each iteration the estimate $\xi^{(k)}$ falls within the appropriate interval, namely $[0, \min(\lambda_1, \lambda_2)]$. Indeed, since for each $i \in \{1, \ldots, n\}$, one has $0 \leqslant Z_i \leqslant \min(X_{i1}, X_{i2})$, it follows that the expectation $\mathrm{E}(Z_i \mid x_{i1}, x_{i2}, \Psi)$ will also fall between 0 and $\min(X_{i1}, X_{i2})$. Accordingly, the average $\bar{q}(\Psi^{(k)})$ will also be positive and less than or equal to $\min(\bar{X}_1, \bar{X}_2)$. Thus, the form of the EM updates inherently reflects the parameter constraints on ξ.

Remark 2.1 It is interesting to note that for the classical bivariate Poisson model, the maximum likelihood estimates for λ_1 and λ_2 are the same in the case of independence. In fact, estimation of the two marginal means is unaffected by the strength of the dependence. Additionally, the maximum likelihood estimation approach results in the same estimates for λ_1 and λ_2 as those obtained via the method of moments.

Other approaches

There are several other approaches for parameter estimation in the classical bivariate Poisson model that have been suggested in the literature. A few of these techniques will be detailed in what follows.

Holgate (1964) explores the double zero proportion method wherein the dependence parameter is estimated using the observed proportion of $(0, 0)$ in a random sample. Let the latter quantity be denoted by ϕ. Following this approach, one has that the theoretical value of ϕ is equal to $\exp(-\lambda_1 - \lambda_2 + \xi)$. Thus, if the marginal parameter estimates are set equal to their respective marginal sample means, one obtains an estimate for ξ equal to $\bar{x}_1 + \bar{x}_2 + \ln \phi$.

Loukas et al. (1986) consider the even points estimation method, which is based on the relation that results from evaluating the probability generating function at the points $(1, 1)$ and $(-1, -1)$. The sum of the latter two expressions yields $1+\exp\{-2(\lambda_1+\lambda_2-2\xi)\}$, which can be equated to its empirical counterpart. Letting a/n denote the proportion of sample points of the form $(2x_1, 2x_2)$ and $(2x_1 + 1, 2x_2 + 1)$, the even points method sets $1 + \exp\{-2(\lambda_1 + \lambda_2 - 2\xi)\} = 2a/n$. Then, holding the marginal parameters fixed at their respective MLEs, an estimator for ξ is derived as $(\bar{X}_1 + \bar{X}_2)/2 + \ln\{2(a/n) - 1\}/4$.

As an alternative to the traditional method of moments approach, Papageorgiou and Kemp (1988) consider using conditional moments to estimate the dependence parameter. For the classical bivariate Poisson model, one has $\mathrm{E}(X_1 \mid X_2 = 0) = \lambda_1-\xi$ and similarly $\mathrm{E}(X_2 \mid X_1 = 0) = \lambda_2-\xi$. Setting λ_1 and λ_2 equal to $\bar{X}_1$ and $\bar{X}_2$, respectively, matching the theoretical conditional moments to those observed in the sample, two possible estimators of ξ are given by

$$\bar{x}_1 - \sum_{i=1}^{n} x_{i1}\mathbf{1}\{x_{i2} = 0\}, \quad \bar{x}_2 - \sum_{i=1}^{n} x_{i2}\mathbf{1}(x_{i1} = 0),$$

where $\mathbf{1}(\cdot)$ represents the indicator function. One could also consider taking the average of the two so that ξ is estimated by

$$\frac{1}{2}\left\{\bar{x}_1 + \bar{x}_2 - \sum_{i=1}^{n} x_{i1}\mathbf{1}(x_{i2} = 0) - \sum_{i=1}^{n} x_{i2}\mathbf{1}(x_{i1} = 0)\right\}.$$

2.1.4 Shortcomings of the model

As previously discussed, the main drawback of the classical bivariate Poisson model is its restricted range of permissible correlation. Notably, with the exception of the trivial case where the marginal means are equal, the common shock model will not reach the upper bound for the correlation determined by Griffiths et al. (1979) and thus cannot represent strong degrees of dependence. Moreover, the classical model cannot allow for negative association. Nonetheless, the classical bivariate Poisson model is intuitive and interpretable and certainly a useful model for correlated count data which exhibit a modest range of positive association.

2.2 Multivariate extension

The stochastic representation of the classical bivariate Poisson model given in (2.1) allows for a natural extension to higher dimensions via a multivariate reduction technique. In a similar way to what was done in two dimensions, a d-variate Poisson model can be constructed whereby dependence is introduced via a common shock variable. Consider a set of d mutually independent Poisson random variables $Y_1, \ldots, Y_d$ with respective means $\lambda_1-\xi, \ldots, \lambda_d-\xi$. Let $Z \sim \mathcal{P}(\xi)$ denote the common shock variable, which is taken to be independent of the random vector $(Y_1, \ldots, Y_d)$. As before, one obtains the d-variate classical Poisson model by adding the common shock variable to each component of $(Y_1, \ldots, Y_d)$, viz.

$$X_1 = Y_1 + Z, \quad \ldots, \quad X_d = Y_d + Z. \tag{2.13}$$

Similarly to the bivariate model, the dependence parameter is constrained to fall in the interval $[0, \min(\lambda_1, \ldots, \lambda_d)]$.

This formulation of a multivariate Poisson model is discussed by several authors. For example, Loukas and Papageorgiou (1991) and Loukas (1993) discuss the common shock approach in the trivariate Poisson distribution while Tsionas (1999) and Karlis (2003) consider the model in higher dimensions. As shown by the aforementioned authors, many of the distributional properties of the bivariate Poisson model extend easily to higher dimensions and so will only be briefly discussed here.

The probability mass function takes on a very similar form as the bivariate model. For a d-dimensional random vector $(X_1, \ldots, X_d)$ arising from the common shock multivariate Poisson distribution, the joint probability mass function is given by

$$h_{\Lambda,\xi}(x_1, \ldots, x_d) = \exp\left\{-\sum_{j=1}^{d} \lambda_j + (d-1)\xi\right\} \sum_{z=0}^{\min(x_1,\ldots,x_d)} \prod_{j=1}^{d} \frac{(\lambda_j - \xi)^{x_j - z}}{(x_j - z)!} \frac{\xi^z}{z!}$$

with corresponding cumulative distribution function given by

$$H_{\Lambda,\xi}(x_1, \ldots, x_d) = \sum_{z=0}^{\min(x_1,\ldots,x_d)} \prod_{j=1}^{d} G_{\lambda_j - \xi}(x_j - z) g_\xi(z).$$

Similarly to the bivariate setting, in the above Λ denotes the vector $(\lambda_1, \ldots, \lambda_d)$ of marginal parameters and ξ represents the mean of the unobserved common shock variable.

Focusing on the case where $d = 3$, Loukas and Papageorgiou (1991) establish recurrence relations, which are shown to be

$$\begin{aligned}
(x_1 + 1)h_{\Lambda,\xi}(x_1 + 1, x_2, x_3) &= (\lambda_1 - \xi)h_{\Lambda,\xi}(x_1, x_2, x_3) + \xi h_{\Lambda,\xi}(x_1, x_2 - 1, x_3 - 1),\\
(x_2 + 1)h_{\Lambda,\xi}(x_1, x_2 + 1, x_3) &= (\lambda_2 - \xi)h_{\Lambda,\xi}(x_1, x_2, x_3) + \xi h_{\Lambda,\xi}(x_1 - 1, x_2, x_3 - 1),\\
(x_3 + 1)h_{\Lambda,\xi}(x_1, x_2, x_3 + 1) &= (\lambda_3 - \xi)h_{\Lambda,\xi}(x_1, x_2, x_3) + \xi h_{\Lambda,\xi}(x_1 - 1, x_2 - 1, x_3).
\end{aligned}$$

Note that recurrence relations in higher dimensions are also provided by Kano and Kawamura (1991) and Karlis (2003).

These recurrence relations allow for simplifications in the score equations for maximum likelihood estimation, similarly to what was found in (2.7) through (2.9). The MLEs for the marginal parameters are found to be their respective marginal sample means while the dependence parameter estimator is found to be value of ξ such that $\bar{R} = 1$, where

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n}\frac{h_{\Lambda,\xi}(x_{i1}-1, x_{i2}-1, x_{i3}-1)}{h_{\Lambda,\xi}(x_{i1}, x_{i2}, x_{i3})}$$

subject to the constraint $\xi \in [0, \min(\lambda_1, \lambda_2, \lambda_3)]$. An expression for the asymptotic variance of the maximum likelihood estimators can also be derived; see Loukas and Papageorgiou (1991) for details.

Continuing with the trivariate model, a moment based estimation procedure is outlined in Loukas (1993). Once again, the marginal means are estimated by their respective sample means. Mixed moments are required for estimation of the dependence parameter. In particular, Loukas (1993) suggests matching the theoretical moment

$$\mathrm{E}\left\{(X_1-\lambda_1)(X_2-\lambda_2)(X_3-\lambda_3)\right\} = \xi$$

to its sample moment given by

$$\frac{1}{n}\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)(x_{i3}-\bar{x}_3),$$

subject to the constraint that $\xi \in [0, \min(\lambda_1, \lambda_2, \lambda_3)]$. The resulting covariance matrix for the moment-based estimators is detailed in Loukas (1993).

Loukas (1993) also outlines alternative estimation techniques, in particular, the method of zero frequency and the method of even points. The former is similar to the double zero proportion method used by Holgate (1964) wherein the theoretical probability of observing $(0, 0, 0)$ is matched to its empirical analogue to obtain an estimate of ξ, while holding $(\lambda_1, \lambda_2, \lambda_3)$ fixed at $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$. The latter approach is the multivariate extension of the even points method discussed for the bivariate model and consists of working with the relation that ensues by evaluating the probability generating function at $(1, 1, 1)$ and $(-1, -1, -1)$.

Obviously, as the dimension increases, estimation in the multivariate Poisson model becomes increasingly complex. Seemingly, the approaches outlined by Loukas and Papageorgiou (1991) and Loukas (1993) will become numerically infeasible in higher dimensions. As yet another alternative, Tsionas (1999) considers a Bayesian approach for estimation in the common shock multivariate Poisson model and provides a detailed numerical illustration using Gibbs sampling. Karlis (2003) suggests using the EM algorithm to obtain maximum likelihood estimates, which is a natural extension to the bivariate case previously discussed. Note that Karlis (2003) works with a different parametrization of the model and allows for an offset term. He also points out that a recursive relation occurs in the conditional distribution in the E-step, thus simplifying the computations.

In addition to inheriting the same drawbacks discussed in the bivariate model, the multivariate common shock construction leads to a restrictive covariance structure wherein all pairwise covariances must coincide. Indeed, for any $i \neq j$, the pair $(X_i, X_j) \subset (X_1, \ldots, X_d) \sim H_{\Lambda,\xi}$ has covariance $\text{cov}(X_i, X_j) = \xi$.

A more flexible definition of a multivariate Poisson model can be achieved by convoluting a d-dimensional vector comprised of independent Poisson components with a set of $2^d - d - 1$ independent shock variables. This produces a $2^d - 1$ parameter multivariate Poisson distribution which allows for a less restrictive covariance structure. This construction is examined by several authors. For example, Mahamunulu (1967) and Kawamura (1976) consider the trivariate case, while higher dimensions are examined by Karlis and Meligkotsidou (2005), to name a few.

Following the general notation of Karlis and Meligkotsidou (2005), let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector with Poisson margins. Then $\mathbf{X}$ is said to follow a multiple common shock multivariate Poisson distribution if one can write $\mathbf{X} = \mathbf{AY}$, where $\mathbf{Y} = (Y_1, \ldots, Y_k)$ is a vector of independent Poisson random variables such that $Y_r \sim \mathcal{P}(\nu_r)$ for each $r \in \{1, \ldots, k\}$, and $\mathbf{A}$ is a $d \times k$ matrix, where all elements are either 0 or 1 and no columns are repeated. Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_k)$ denote the parameter vector. Then, this extended definition of a multivariate Poisson model results in the following mean and covariance structure:

$$\mathrm{E}(\mathbf{X}) = \mathbf{A}\boldsymbol{\nu}, \quad \text{cov}(\mathbf{X}) = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top,$$

where $\boldsymbol{\Sigma} = \text{diag}(\nu_1, \ldots, \nu_\mathrm{k})$.

Consider, for example, the trivariate case. Letting $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_{12}, Y_{13}, Y_{23}, Y_{123})$, the extended construction leads to the following stochastic representation:

$$\begin{aligned} X_1 &= Y_1 + Y_{12} + Y_{13} + Y_{123}, \\ X_2 &= Y_2 + Y_{12} + Y_{23} + Y_{123}, \\ X_3 &= Y_3 + Y_{13} + Y_{23} + Y_{123}. \end{aligned}$$

A similar interpretation to what was discussed in the bivariate model follows in this more complex setting. The random vector (Y_1, Y_2, Y_3) consists of the independent base while $(Y_{12}, Y_{13}, Y_{23}, Y_{123})$ may be viewed as a set of common shock variables. If $\mathbf{Y}$ has mean vector $(\nu_1, \nu_2, \nu_3, \nu_{12}, \nu_{13}, \nu_{23}, \nu_{123})$, it follows that

$$\begin{aligned} \text{cov}(X_1, X_2) &= \text{cov}(Y_1 + Y_{12} + Y_{13} + Y_{123}, Y_2 + Y_{12} + Y_{23} + Y_{123}) \\ &= \text{var}(Y_{12}) + \text{var}(Y_{123}) \\ &= \nu_{12} + \nu_{123}. \end{aligned}$$

Similarly, one has that

$$\text{cov}(X_1, X_3) = \nu_{13} + \nu_{123}, \qquad \text{cov}(X_2, X_3) = \nu_{23} + \nu_{123}.$$

Thus, when $\nu_{12} = \nu_{13} = \nu_{23} = 0$ the extended trivariate model simplifies to the single common shock representation in (2.13). Moreover, it is easily seen that this representation allows for the pairwise covariances to differ.

Clearly, for an arbitrary dimension d, the extended multivariate Poisson construction allows for varying pairwise covariances as the common shock variables affect each margin differently. Despite the added flexibility, this extended model still restricts the range of permissible pairwise correlations and falls short of the upper bound $\rho_{\max}$ except in trivial cases. This follows from the fact that the univariate Poisson distribution is infinitely divisible, and thus any two components of $\mathbf{X}$ can be rewritten in the form of the bivariate Poisson construction given in (2.1).

For example, consider once again the case where $d = 3$. Focusing on the first two components, one has

$$X_1 = Y_1 + Y_{12} + Y_{13} + Y_{123} = Y_1^* + Z^*,$$
$$X_2 = Y_2 + Y_{12} + Y_{23} + Y_{123} = Y_2^* + Z^*,$$

where

$$Y_1^* = Y_1 + Y_{13}, \quad Y_2^* = Y_2 + Y_{23}, \quad Z^* = Y_{12} + Y_{123}.$$

Since the Poisson distribution is closed under convolution, $Y_1^* \sim \mathcal{P}(\nu_1 + \nu_{13})$, $Y_2^* \sim \mathcal{P}(\nu_2 + \nu_{23})$ and $Z^* \sim \mathcal{P}(\nu_{12} + \nu_{123})$. Moreover, Y_1^*, Y_2^* and Z^* are independent as they are each sums of non-overlapping independent random variables. Thus, the pair (X_1, X_2) follows the classical bivariate Poisson distribution and accordingly its covariance is restricted to $[0, \min(\lambda_1, \lambda_2)]$, where $\lambda_1 = \nu_1 + \nu_{12} + \nu_{13} + \nu_{123}$ and $\lambda_2 = \nu_2 + \nu_{12} + \nu_{23} + \nu_{123}$.

2.3 Alternative multivariate Poisson models

Thus far, only the common shock, or multivariate reduction, technique for building a multivariate Poisson model has been discussed. Certainly, there are alternative ways to generate a multivariate Poisson distribution. This section will give a brief overview of alternative methods proposed in the literature, including copula models, conditional models, mixture models as well as some more general models that include the Poisson as a special case.

2.3.1 Copula models

Copula models provide a flexible framework for building multivariate models with arbitrary margins. Consider the bivariate case: suppose X_1 and X_2 are Poisson random variables with respective marginal distribution functions G_{λ_1} and G_{λ_2}. Given a class (C_ϕ) of copulas indexed by a dependence parameter ϕ, a joint distribution of (X_1, X_2) can then be constructed by setting, for all $x_1, x_2 \in \mathbb{N}$,

$$\Pr(X_1 \leqslant x_1, X_2 \leqslant x_2) = C_\phi \left\{ G_{\lambda_1}(x_1), G_{\lambda_2}(x_2) \right\}.$$

The choice of copula dictates the nature of the dependence in the model while the dependence parameter ϕ will regulate the strength of the dependence. Indeed, for a suitable choice of copula model, this approach can allow for both positive and negative associations between the margins. Moreover, in the case of the upper and lower Fréchet–Hoeffding boundary copulas, this construction will imply a correlation reaching $\rho_{\max}$ and $\rho_{\min}$, respectively.

This general approach towards constructing correlated Poisson random variables is discussed by several authors, notably Van Ophem (1999), Nikoloulopoulos and Karlis (2009), Smith and Khaled (2012), Pfeifer and Nešlehová (2004) and Panagiotelis et al. (2012), to name a few. Although the copula construction for generating correlated discrete random variables is valid, several issues arise as a result of the discontinuities in the marginal distribution functions. In particular, there is no unique copula representation, inference becomes more difficult, and many of the nice properties that follow for continuous margins no longer hold true. An in-depth examination of copula models for discrete data is given in Genest and Nešlehová (2007).

Along the same lines as copula models, Lakshminarayana et al. (1999) derive a bivariate Poisson model wherein the joint probability mass function is written as the product of the marginal probability mass functions (PMFs) and a multiplicative factor that induces dependence. The proposed construction allows for greater flexibility than the classical bivariate Poisson model as both positive and negative correlation are possible. The authors consider a particular form for the multiplicative factor and derive some properties of the resulting bivariate Poisson distribution, including the range of possible correlation values. In this specific case, the proposed bivariate model falls short of the extremal values of correlation given by Griffiths et al. (1979).

In a similar manner, Nelsen (1987) proposes deriving bivariate models with discrete margins by writing the joint probability mass function as a convex linear combination of the upper and lower Fréchet–Hoeffding boundary distributions along with the product of the marginal PMFs (i.e., the joint distribution under independence). This general construction allows for both positive and negative correlation. Moreover, when the model parameters are chosen such that the bivariate distribution is at the boundaries (i.e., either the upper or lower Fréchet–Hoeffding bound), the implied correlation will reach $\rho_{\max}$ and $\rho_{\min}$, respectively. Griffiths et al. (1979) also provide a few specific examples of bivariate Poisson distributions exhibiting negative correlation.

2.3.2 Mixture models

Another way to construct a multivariate Poisson model is through mixture models. Karlis and Xekalaki (2005) provide an overview of various Poisson mixture models in both the univariate and multivariate setting. Focusing on the bivariate case, the authors provide several definitions of a mixed bivariate Poisson distribution. For example, one may consider the classical common shock model wherein each parameter is proportional to some mixing random variable denoted by a. Following this definition, conditional on a the pair (X_1, X_2) follows the classical bivariate Poisson distribution with parameters $(a\lambda_1, a\lambda_2, a\xi)$. The choice of distribution for a will then dictate the unconditional distribution of the pair (X_1, X_2) and regulate the dependence between the margins. A more general bivariate Poisson mixture model results when the parameters λ_1, λ_2, ξ follow some multivariate distribution.

There are numerous papers which explore specific mixture models. For example, Sarabia and Gómez-Déniz (2011) consider a multivariate Poisson-Beta distribution, Gómez-Déniz et al. (2012) define a multivariate Poisson-Lindley distribution and Sellers et al. (2016) propose a bivariate Conway–Maxwell-Poisson distribution. Karlis and Meligkotsidou (2007) consider a finite mixture of multivariate Poisson distributions which results as the special case where the mixing distribution is multinomial. In the latter paper, parameter estimation is carried out using the EM algorithm.

Mixture models allow for greater flexibility in the covariance structure through the mixing variable. For example, in the univariate case, a mixture Poisson model can allow for over-dispersion. In addition, for suitable choices of mixing distributions, one may construct a multivariate Poisson model with negative dependence.

2.3.3 General multivariate reduction technique

In conjunction with the common shock method and mixture models, Lai (1995) discusses constructing bivariate distributions using a generalized trivariate reduction technique. In particular, the author develops a modified structured mixture model in terms of three mutually independent random variables Y_1, Y_2, Y_3 and a pair (I_1, I_2) of correlated Bernoulli variables, where by design each component of (Y_1, Y_2, Y_3) is independent of the random pair (I_1, I_2). A dependent random vector is then generated by setting

$$X_1 = Y_1 + I_1 Y_3, \quad X_2 = Y_2 + I_2 Y_3.$$

The classical bivariate Poisson model is then the special case where Y_1, Y_2, Y_3 each follow a Poisson distribution and the Bernoulli random pair is such that $\Pr(I_1 = 1, I_2 = 1) = 1$.

This approach towards generating correlated pairs allows for greater flexibility than the classical common shock construction as it allows for two sources of dependence: the common shock variable Y_3 and the dependence induced by the correlated indicator variables (I_1, I_2). Lai (1995) derives the implied correlation between X_1 and X_2 as well as the minimum and maximum correlation permissible in the model. Note that while the structured mixture model method allows for both positive and negative association between the margins, the implied bounds for the correlation do not reach those given by Griffiths et al. (1979) except in trivial cases.

In a similar manner, Cuenin et al. (2016) use a multivariate reduction technique to construct a multivariate Tweedie distribution. Since the univariate Tweedie distribution is closed under convolution, the authors suggest building a multivariate Tweedie model wherein the margins can be written as linear combinations of a set of independent univariate Tweedie random variables. The classic multivariate Poisson model is a special case of the proposed multivariate Tweedie model, where the parameter $p = 1$.

Cuenin et al. (2016) also extend the model to the case of negative dependence by working with sums of independent components along with both comonotonic and counter-monotonic Tweedie random variables. In the bivariate Poisson case (i.e., $p = 1$), this construction reduces to setting $X_1 = Y_{11} + Y_{12}$ and $X_2 = Y_{22} + Y_{21}$, where (Y_{11}, Y_{22}) are independent Poisson random variables, with respective means $\lambda_1 - \lambda_{12}$ and $\lambda_2 - \lambda_{12}$, which are also independent of (Y_{12}, Y_{21}). The latter random pair consists of counter-monotonic Poisson random variables with $(Y_{12}, Y_{21}) = (G^{-1}_{\lambda_{12}}(U), G^{-1}_{\lambda_{12}}(1-U))$. Indeed, in this proposed bivariate Poisson model, setting $\lambda_{12} = \lambda_1 = \lambda_2 = \lambda$ yields $\text{corr}(X_1, X_2) = \rho_{\min}$. In Chapter 5, we propose a bivariate Poisson model with negative dependence that resembles the model formulation considered by Cuenin et al. (2016) in that the construction relies on the use of counter-monotonic Poisson random variables. In particular, our proposed model coincides with that of Cuenin et al. (2016) in the special case where $\lambda_1 = \lambda_2$.

Note that in higher dimensions, the notion of negative dependence becomes more complex. This results from the fact that the concept of counter-monotonicity does not readily extend to dimensions greater than 2. Consider the trivariate Poisson model resulting from the construction of Cuenin et al. (2016). In terms of independent $\mathcal{U}(0, 1)$ random variables U_1, U_2, U_3, U, one has

$$\begin{aligned} X_1 &= Y_{11} + Y_{12} + Y_{13} = G^{-1}_{\lambda_1-\lambda_{12}-\lambda_{13}}(U_1) + G^{-1}_{\lambda_{12}}(U) + G^{-1}_{\lambda_{13}}(U), \\ X_2 &= Y_{22} + Y_{21} + Y_{23} = G^{-1}_{\lambda_2-\lambda_{12}-\lambda_{23}}(U_2) + G^{-1}_{\lambda_{12}}(1-U) + G^{-1}_{\lambda_{23}}(U), \\ X_3 &= Y_{33} + Y_{31} + Y_{32} = G^{-1}_{\lambda_3-\lambda_{13}-\lambda_{23}}(U_3) + G^{-1}_{\lambda_{13}}(1-U) + G^{-1}_{\lambda_{23}}(1-U). \end{aligned}$$

If we consider the pair (X_1, X_3), it is clear that the correlation will be negative since

$$\text{cov}(X_1, X_3) = \text{cov}(Y_{12}, Y_{31}) + \text{cov}(Y_{12}, Y_{32}) + \text{cov}(Y_{13}, Y_{31}) + \text{cov}(Y_{13}, Y_{32}),$$

where each of the sets of pairs (Y_{12}, Y_{13}), (Y_{12}, Y_{32}), (Y_{13}, Y_{31}), (Y_{13}, Y_{32}) are counter-monotonic by construction. However, if we consider (X_1, X_2), the covariance is driven by both comonotonic and counter-monotonic pairs. In this case, the pairs (Y_{12}, Y_{21}), (Y_{13}, Y_{21}) are counter-monotonic while the remaining pairs (Y_{12}, Y_{23}) and (Y_{13}, Y_{23}) are comonotonic. Thus, the overall covariance may be either positive or negative depending on the magnitude of the parameters λ_{12}, λ_{13}, λ_{23}.

Moreover, in the multivariate Poisson case the proposed formulation does not yield margins that are univariate Poisson. In the trivariate setting, consider the first component: X_1 is the sum of an independent Poisson random variable, Y_{11}, along with two perfectly dependent Poisson random variables, (Y_{12}, Y_{13}), as the latter pair are functions of a common underlying uniform random variable U. It then follows that

$$\text{var}(X_1) = \text{var}(Y_{11}) + \text{var}(Y_{12}) + \text{var}(Y_{13}) + 2\text{cov}(Y_{12}, Y_{13}) \geqslant \lambda_1$$

while

$$\text{E}(X_1) = \text{E}(Y_{11}) + \text{E}(Y_{12}) + \text{E}(Y_{13}) = \lambda_1.$$

Clearly, X_1 does not follow a $\mathcal{P}(\lambda_1)$ distribution.

Note that a slight modification to the proposed construction rectifies this issue. Continuing with the trivariate case, consider the following stochastic representation in terms of mutually independent $\mathcal{U}(0,1)$ random variables U_1, U_2, U_3, U_{12}, U_{13}, U_{23}:

$$\begin{aligned}
X_1 &= Y_{11} + Y_{12} + Y_{13} = G^{-1}_{\lambda_{11}-\lambda_{12}-\lambda_{13}}(U_1) + G^{-1}_{\lambda_{12}}(U_{12}) + G^{-1}_{\lambda_{13}}(U_{13}),\\
X_2 &= Y_{22} + Y_{21} + Y_{23} = G^{-1}_{\lambda_{22}-\lambda_{12}-\lambda_{23}}(U_2) + G^{-1}_{\lambda_{12}}(1-U_{12}) + G^{-1}_{\lambda_{23}}(U_{23}),\\
X_3 &= Y_{33} + Y_{31} + Y_{32} = G^{-1}_{\lambda_{33}-\lambda_{13}-\lambda_{23}}(U_3) + G^{-1}_{\lambda_{13}}(1-U_{13}) + G^{-1}_{\lambda_{23}}(1-U_{23}).
\end{aligned}$$

Clearly any two components can be rewritten in terms of the bivariate model representation in that the pairwise margins are comprised of both an independent and counter-monotonic element. Thus, all pairwise covariances will be negative. Furthermore, each margin is the sum of independent Poisson random variables so that $X_i \sim \mathcal{P}(\lambda_i)$ for all $i \in \{1, 2, 3\}$. This model is revisited in Chapter 7.

2.3.4 Other approaches

Dependence between Poisson margins can also be generated via random effects models, as shown by Chib and Winkelmann (2001), among others. In allowing for an unrestricted covariance matrix for the random effects, the authors define a multivariate Poisson model with a flexible dependence structure, accommodating both positive and negative correlation.

A bivariate Poisson model can also be defined in terms of conditional distributions, as done in Papageorgiou (1983) and Berkhout and Plug (2004), for example. In particular, Berkhout and Plug (2004) construct a bivariate Poisson model by writing the conditional mean of one margin as a function of the other. Specifically, let (X_1, X_2) denote a random pair of Poisson random variables with respective means λ_1 and λ_2. Then, for a given choice of permutation π (where there are two possibilities for the bivariate case) and conditional on $X_2 = x_2$, write

$$\lambda_2 = \lambda_{20} \exp(\alpha x_2).$$

This approach allows for both positive and negative correlation, depending on the sign of α. Note that the authors allow the marginal and conditional means to depend on a set of covariates through the exponential link function.

There are numerous papers that explore multivariate Poisson regression models, in which the effects of covariates are incorporated into the model. For example, Kocherlakota and Kocherlakota (2001) write the marginal means in the classical bivariate Poisson model in terms of a GLM using a log link function. As an extension to this, Tsionas (2001) consider a set of marginal GLMs, again using the log link functions, for each marginal mean in the classical multivariate Poisson model. Karlis and Meligkotsidou (2005) allow for covariate effects via GLMs on the marginal means in the more complex version of a multivariate Poisson model while Nikoloulopoulos and Karlis (2010) incorporate marginal GLMs in a copula model framework. Bermúdez and Karlis (2012) propose a finite mixture of bivariate Poisson regression models and apply it to car insurance data. As yet another alternative, several authors define a multivariate Poisson distribution as the limiting case of a multivariate Binomial distribution; see, e.g., Krishnamoorthy (1951), Kawamura (1976), Kawamura (1979).

2.4 Other multivariate count models

There are numerous other models for correlated count data wherein the marginal distributions are not Poisson. Some examples include the Generalized Poisson, Zero-Inflated Poisson and Negative Binomial distributions. Famoye (2012) and Famoye (2015) provide detailed comparisons of some of these models in the bivariate and multivariate setting.

A bivariate Generalized Poisson distribution is developed in Famoye and Consul (1995), where the model construction uses the trivariate reduction method in an analogous manner to the classical bivariate Poisson model. As was the case in the Poisson model, this approach implies a positive correlation. Famoye (2010a) later considered an alternative derivation of a bivariate Generalized Poisson model with flexible correlation by working with the formulation proposed by Lakshminarayana et al. (1999). In the same way, Famoye (2010b) uses the method of Lakshminarayana et al. (1999) to define a bivariate Negative Binomial model, which also incorporates covariate effects. The bivariate Generalized Poisson model is then extended to higher dimensions in Famoye (2015). The latter model allows for a flexible covariance structure as well as both over and under dispersion, while incorporating covariate effects. Zamani et al. (2016) also consider a regression model in the context of the bivariate Generalized Poisson model. The authors explore various forms of the aforementioned model that have been proposed in the literature and apply these models to health care data.

Multivariate extensions to the Zero-Inflated Poisson model are discussed in various papers. Li et al. (1999) propose a multivariate Zero-Inflated Poisson distribution which involves a mass at 0 along with univariate and multivariate Poisson distributions. Walhin (2001) consider different techniques for constructing bivariate Zero-Inflated Poisson models via mixture models as well as a trivariate reduction approach. Gurmu and Elder (2008) also consider a mixture model wherein dependence is introduced via a pair of correlated latent variables acting on both marginal means while allowing for a flexible correlation structure. Bivariate Zero-Inflated Generalized Poisson models are considered by Zhang et al. (2015) and Faroughi and Ismail (2017). In particular, Zhang et al. (2015) define two models, one with a common inflation for both margins and a second with two separate inflation factors.

There are of course several other examples of multivariate models for count data. For example, Gurmu and Elder (2000) propose a bivariate Generalized Negative Binomial regression model and consider an application to health care data. Akpoue and Angers (2017) develop a multivariate Poisson-Skellam distribution and illustrate its use by modelling soccer data. Lindskog and McNeil (2003) discuss a general Poisson shock model applied to insurance data. In particular, the authors consider dependence in both the claim frequencies and severities for different types of losses.

3

Comonotonic Discrete Random Pairs

3.1 Introduction

As a building block towards defining an extension to the classical bivariate Poisson model, we begin by considering a special case: comonotonic discrete random variables. While the focus here is on the two-dimensional case, most of what is discussed can be extended to higher dimensions.

A pair of random variables (Z_1, Z_2) is said to be comonotonic if and only if it can be written as a non-decreasing function of a common underlying random variable; see, e.g., Dhaene et al. (2006). In particular, for $U \sim \mathcal{U}(0, 1)$, the comonotonic random pair can be written as

$$(Z_1, Z_2) =_d (F_1^{-1}(U), F_2^{-1}(U)), \tag{3.1}$$

where F_1 and F_2 denote the marginal cumulative distribution functions of Z_1 and Z_2, respectively, and the (generalized) inverse cumulative distribution function, or quantile function, is defined, for all $u \in [0, 1]$ and $x \in \bar{\mathbb{R}} = [-\infty, \infty]$, by

$$F_X^{-1}(u) = \inf\{x \in \mathbb{R} : F_X(x) \geqslant u\}. \tag{3.2}$$

An extensive review of the quantile function and its properties is given in Embrechts and Hofert (2013).

The joint distribution of a comonotonic pair (Z_1, Z_2) is given, for all $z_1, z_2 \in \mathbb{R}$, by

$$\begin{aligned}\Pr(Z_1 \leqslant z_1, Z_1 \leqslant z_2) &= \Pr\{F_1^{-1}(U) \leqslant z_1, F_2^{-1}(U) \leqslant z_2\} \\ &= \Pr\{U \leqslant F_1(z_1), U \leqslant F_2(z_2)\} \\ &= \Pr[U \leqslant \min\{F_1(z_1), F_2(z_2)\}] \\ &= \min\{F_1(z_1), F_2(z_2)\}. \end{aligned} \tag{3.3}$$

The above follows from the fact that for any $x \in \mathbb{R}$ and $u \in [0, 1]$,

$$F_X(x) \geqslant u \quad \Leftrightarrow \quad x \geqslant F_X^{-1}(u)$$

accordingly to the definition of F_X^{-1} and the fact that any CDF is right continuous.

The joint distribution function given in (3.3) is known as the Fréchet–Hoeffding upper bound and represents perfect positive dependence. It can be shown that any bivariate distribution function H with given margins F_1 and F_2 satisfies

$$\max\{0, F_1(x_1) + F_2(x_2) - 1\} \leqslant H(x_1, x_2) \leqslant \min\{F_1(x_1), F_2(x_2)\}, \tag{3.4}$$

for all $(x_1, x_2) \in \mathbb{R}^2$; see e.g., Eq. (2.5.1) on p. 30 of Nelsen (2006). In the above equation, $\max\{0, F_1(x_1) + F_2(x_2) - 1\}$ is referred to as the lower Fréchet–Hoeffding bound and is the joint distribution function of a counter-monotonic random pair (X_1, X_2) defined as

$$(X_1, X_2) =_d (F_1^{-1}(U), F_2^{-1}(1 - U)).$$

Analogously to the concept of comonotonicity, counter-monotonicity represents perfect negative dependence. In this thesis, $\mathcal{M}(F_1, F_2)$ and $\mathcal{W}(F_1, F_2)$ will denote the upper and lower Fréchet–Hoeffding boundary distributions, respectively, with margins F_1 and F_2.

Note that the Fréchet–Hoeffding bounds for bivariate distribution functions given in (3.4) extend to higher dimensions, as shown in, e.g., Theorem 3.1 of Joe (1997). Suppose the d-dimensional random vector $(X_1, \ldots, X_d)$ has marginal distribution functions given by $F_1, \ldots, F_d$. Then for any $(x_1, \ldots, x_d) \in \mathbb{R}^d$, the joint distribution H of $(X_1, \ldots, X_d)$ is bounded by

$$\max\{0, F_1(x_1) + \cdots + F_d(x_d) - (d - 1)\} \leqslant H(x_1, \ldots, x_d) \leqslant \min\{F_1(x_1), \ldots, F_d(x_d)\}.$$

While the upper Fréchet–Hoeffding bound $\min\{F_1(x_1), \ldots, F_d(x_d)\}$ is a valid distribution function for arbitrary dimension d, in general this will not hold for the lower bound $\max\{0, F_1(x_1) + \cdots + F_d(x_d) - (d - 1)\}$. Theorem 3.7 of Joe (1997) details the case where the lower bound is a proper CDF.

3.2 Continuous vs. discrete margins

Recall the definition of the quantile function given in Eq. (3.2). Since any cumulative distribution function is by definition non-decreasing and right continuous, F^{-1} itself is a non-decreasing and left continuous function with $F^{-1} : [0, 1] \mapsto \bar{\mathbb{R}}$. Moreover, it can be shown that

$$\forall_{x \in \mathbb{R}} \;\; F^{-1}\{F(x)\} \leqslant x \quad \text{and} \quad \forall_{u \in [0,1]} \;\; F\{F^{-1}(u)\} \geqslant u.$$

In the special case where the CDF F is strictly increasing, one has that $F^{-1}\{F(x)\} = x$; while when F is continuous, one has that $F\{F^{-1}(u)\} = u$.

For a continuous distribution function, the corresponding quantile function is a one-to-one mapping so that every x in the support of F can only be generated from a unique $u \in [0, 1]$, viz. $X = F^{-1}(U)$. However, when the cumulative distribution function F has jumps, the resulting F^{-1} will be a many-to-one mapping. To see this, suppose that F has a jump at x_0 so that $F(x_0^-) < F(x_0)$. Then any u in the jump interval $\big(F(x_0^-), F(x_0)\big]$ will be mapped to x_0 since $F^{-1}(u) = \inf\{x : F(x) \geqslant u\}$. For example, consider a Poisson random variable with mean λ and distribution function G_λ; the corresponding quantile function G_λ^{-1} can then be defined, for all $u \in [0, 1]$, by

$$G_\lambda^{-1}(u) = \sum_{k \in \mathbb{N}} k \times \mathbf{1}\left\{G_\lambda(k-1) < u \leqslant G_\lambda(k)\right\}, \tag{3.5}$$

where $\mathbb{N} = \{0, 1, 2, \ldots\}$ and $\mathbf{1}$ represents the indicator function.

Note that plateaus in the cumulative distribution function imply jumps in the corresponding inverse CDF. Suppose a distribution function F is flat for $x \in [a_1, a_2)$ so that $F(x) = F(a_1)$ for all $x \in [a_1, a_2)$, i.e., $\Pr(a_1 < X < a_2) = 0$. Accordingly, the quantile function will set $F^{-1}\{F(a_1)\} = a_1$ while $F^{-1}\{F(a_1) + \epsilon\} > a_2$ for any $\epsilon > 0$ such that $\epsilon \leqslant 1 - F(a_1)$. Thus, no $u \in [0, 1]$ will be mapped to an $x \in (a_1, a_2]$.

Recall the stochastic representation for comonotonic pairs, namely, for $U \sim \mathcal{U}(0, 1)$, the pair can be expressed as $(Z_1, Z_2) =_d (F_1^{-1}(U), F_2^{-1}(U))$. When the marginal distribution functions F_1 and F_2 are continuous, comonotonicity implies a functional dependence between the margins. In this case, every observed pair $(z_1, z_2) \in \mathbb{R}^2$ arises from a unique $u \in [0, 1]$ as the marginal quantile functions generating the pair are one-to-one. Accordingly, every observed pair (z_1, z_2) must be such that $F_1(z_1) = F_2(z_2) = u$. This translates into a deterministic relationship between the two random variables in terms of their marginal cumulative distribution functions, viz.

$$Z_2 =_d F_2^{-1}(U) = F_2^{-1}\{F_1(Z_1)\},$$

or, equivalently,

$$Z_1 =_d F_1^{-1}(U) = F_1^{-1}\{F_2(Z_2)\}.$$

Thus, for continuous random variables, comonotonicity reduces the dimensionality of the problem from the bivariate to the univariate setting since one margin is deterministically ascertained from the other. In other words, conditional on say $Z_1 = z_1$, Z_2 is a degenerate random variable which places all its mass at the point $F_2^{-1}\{F_1(z_1)\}$. To illustrate this, the following example considers the case of Exponential margins.

Example 3.1 Comonotonic Exponential random variables
Consider the case where the margins are Exponential. In particular, suppose that for $s \in \{1, 2\}$, $Z_s \sim \mathcal{E}(\beta_s)$ *with* $F_s(x) = 1 - \exp(-x/\beta_s)$ *for all* $x \geqslant 0$ *and corresponding inverse CDF given by* $F_s^{-1}(u) = -\beta_s \ln(1-u)$ *for all* $u \in (0, 1)$*. Then if* (Z_1, Z_2) *are comonotonic, it follows that*

$$1 - \exp(-Z_1/\beta_1) = 1 - \exp(-Z_2/\beta_2) \quad \Leftrightarrow \quad Z_1/\beta_1 = Z_2/\beta_2 \quad \Leftrightarrow \quad Z_1 = \beta_1 Z_2/\beta_2$$

almost surely. Thus, in the case of Exponential random variables, comonotonicity implies a linear relation between the margins, and conditional on $Z_1 = z_1$, $Z_2 = \beta_2 z_1/\beta_1$ *with probability* 1.

In fact, any comonotonic pair with strictly increasing marginal distribution functions belonging to a common location-scale family will exhibit a linear relation. To see this, suppose that Z_1 and Z_2 are members of the same location-scale family with baseline distribution function F, i.e., there exist $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1, \sigma_2 > 0$ such that, for all $z_1, z_2 \in \mathbb{R}$,

$$F_1(z_1) = F\left(\frac{z_1 - \mu_1}{\sigma_1}\right), \quad F_2(z_2) = F\left(\frac{z_2 - \mu_2}{\sigma_2}\right).$$

Suppose that the underlying standard location-scale distribution F is strictly increasing so that $F^{-1}\{F(x)\} = x$. Then, it follows that for each $s \in \{1, 2\}$,

$$\begin{aligned} Z_s = F_s^{-1}(U) \quad &\Leftrightarrow \quad U = F_s(Z_s) = F\left(\frac{Z_s - \mu_s}{\sigma_s}\right) \\ &\Leftrightarrow \quad F^{-1}(U) = (Z_s - \mu_s)/\sigma_s \\ &\Leftrightarrow \quad Z_s = \mu_s + \sigma_s F^{-1}(U). \end{aligned}$$

Comonotonicity then implies that

$$Z_2 = \mu_2 + \sigma_2 F^{-1}(U) = \mu_2 + \sigma_2 \left(\frac{Z_1 - \mu_1}{\sigma_1}\right) = \mu_2 - \mu_1 \left(\frac{\sigma_2}{\sigma_1}\right) + \left(\frac{\sigma_2}{\sigma_1}\right) Z_1$$

almost surely. Thus, Z_2 is a linear function of Z_1. Some examples of distributions belonging to the location-scale family include the Exponential, Normal and Uniform distributions, to name a few.

For discrete random variables, in contrast, the relation between the margins in the case of comonotonicity is no longer straightforward. For a discrete random variable, the cumulative distribution function has jump discontinuities at every $x \in \mathbb{R}$ for which the probability mass function is non-zero. This in turn causes the inverse CDF to also be a jump function, as shown in Eq. (3.5) for the case of a Poisson random variable. For a comonotonic pair with discrete margins, the joint distribution only places mass at values (z_1, z_2) such that the jump intervals overlap, i.e., such that

$$(F_1(z_1^-), F_1(z_1)] \cap (F_2(z_2^-), F_2(z_2)] \neq \varnothing.$$

This follows since comonotonicity implies that both margins are generated by the same underlying uniform random variable U such that $Z_1 = F_1^{-1}(u) = z_1 \Leftrightarrow u \in (F_1(z_1^-), F_1(z_1)]$ and $Z_2 = F_2^{-1}(u) = z_2 \Leftrightarrow u \in (F_2(z_2^-), F_2(z_2)]$, and so jointly

$$(Z_1, Z_2) = (z_1, z_2) \quad \Leftrightarrow \quad u \in (F_1(z_1^-), F_1(z_1)] \cap (F_2(z_2^-), F_2(z_2)].$$

Thus, for comonotonic random variables with discontinuous marginal distribution functions, the joint probability mass function (PMF) can be derived as

$$\begin{aligned}\Pr(Z_1 = z_1, Z_2 = z_2) &= \Pr\{U \in (F_1(z_1^-), F_1(z_1)] \cap (F_2(z_2^-), F_2(z_2)]\} \\ &= \Pr[\max\{F_1(z_1^-), F_2(z_2^-)\} < U \leqslant \min\{F_1(z_1), F_2(z_2)\}] \\ &= [\min\{F_1(z_1), F_2(z_2)\} - \max\{F_1(z_1^-), F_2(z_2^-)\}]_+ \end{aligned} \tag{3.6}$$

where $[x]_+ = x\mathbf{1}(x > 0)$. It is clear from the above formulation that the probability of observing (z_1, z_2) depends on how the two intervals, $(F_1(z_1^-), F_1(z_1)]$ and $(F_2(z_2^-), F_2(z_2)]$ overlap. Clearly, $F_1(z_1^-) < F_1(z_1)$ and $F_2(z_2^-) < F_2(z_2)$. Intersecting the two intervals could result in six possible orderings, namely

(i) $F_2(z_2^-) \leqslant F_1(z_1^-) \leqslant F_2(z_2) \leqslant F_1(z_1)$;
(ii) $F_1(z_1^-) \leqslant F_2(z_2^-) \leqslant F_2(z_2) \leqslant F_1(z_1)$;
(iii) $F_1(z_1^-) \leqslant F_2(z_2^-) \leqslant F_1(z_1) \leqslant F_2(z_2)$;
(iv) $F_2(z_2^-) \leqslant F_1(z_1^-) \leqslant F_1(z_1) \leqslant F_2(z_2)$;
(v) $F_1(z_1^-) < F_1(z_1) \leqslant F_2(z_2^-) < F_2(z_2)$;
(vi) $F_2(z_2^-) < F_2(z_2) \leqslant F_1(z_1^-) < F_1(z_1)$.

Each of the above ordering correspond to a different joint probability, specifically,

$$\Pr(Z_1 = z_1, Z_2 = z_2) = \begin{cases} F_2(z_2) - F_1(z_1^-), & \text{in case (i)}, \\ F_2(z_2) - F_2(z_2^-) = \Pr(Z_2 = z_2) & \text{in case (ii)}, \\ F_1(z_1) - F_2(z_2^-) & \text{in case (iii)}, \\ F_1(z_1) - F_1(z_1^-) = \Pr(Z_1 = z_1) & \text{in case (iv)}, \\ 0, & \text{in case (v)}, \\ 0, & \text{in case (vi)}. \end{cases} \tag{3.7}$$

Accordingly, conditional on Z_1, the distribution of Z_2 is no longer necessarily degenerate. Indeed, one has

$$\Pr(Z_2 = z_2 \mid Z_1 = z_1) = \frac{1}{\Pr(Z_1 = z_1)} \times [\min\{F_1(z_1), F_2(z_2)\} - \max\{F_1(z_1^-), F_2(z_2^-)\}]_+,$$

which is equal to 1 only when the ordering is as in case (iv), i.e., when $(F_1(z_1^-), F_1(z_1)] \subseteq (F_2(z_2^-), F_2(z_2)]$. Thus, in general, knowledge of one margin does not allow to identify its comonotonic counterpart in the case of discrete random variables, except in specific situations.

Example 3.2 Comonotonic Bernoulli random variables
Suppose that $Z_1 \sim$ Bernoulli(p), $Z_2 \sim$ Bernoulli(q), and $\Pr(Z_1 = 0) = p$ and $\Pr(Z_2 = 0) = q$. If $0 \leqslant p \leqslant q \leqslant 1$, then the joint probability mass function takes the following form:

$$\begin{aligned}
\Pr(0,0) &= [\min(p,q) - \max(0,0)]_+ = p,\\
\Pr(1,0) &= [\min(1,q) - \max(p,0)]_+ = q - p,\\
\Pr(1,1) &= [\min(1,1) - \max(p,q)]_+ = 1 - q,\\
\Pr(0,1) &= [\min(p,1) - \max(0,q)]_+ = 0.
\end{aligned}$$

The conditional probability mass function can then be derived as follows:

$$\Pr(Z_2 = z_2 \mid Z_1 = z_1) = \begin{cases} 1 & \text{if } (z_1, z_2) = (0,0),\\ (q-p)/(1-p) & \text{if } (z_1, z_2) = (1,0),\\ (1-q)/(1-p) & \text{if } (z_1, z_2) = (1,1),\\ 0 & \text{if } (z_1, z_2) = (0,1). \end{cases}$$

If instead $0 \leqslant q \leqslant p$, the joint mass function becomes

$$\begin{aligned}
\Pr(0,0) &= [\min(p,q) - \max(0,0)]_+ = q,\\
\Pr(1,0) &= [\min(1,q) - \max(p,0)]_+ = 0,\\
\Pr(1,1) &= [\min(1,1) - \max(p,q)]_+ = 1 - p,\\
\Pr(0,1) &= [\min(p,1) - \max(0,q)]_+ = p - q,
\end{aligned}$$

and the corresponding conditional probability mass function is

$$\Pr(Z_2 = z_2 | Z_1 = z_1) = \begin{cases} q/p & \text{if } (z_1, z_2) = (0,0),\\ 0 & \text{if } (z_1, z_2) = (1,0),\\ 1 & \text{if } (z_1, z_2) = (1,1),\\ (p-q)/p & \text{if } (z_1, z_2) = (0,1). \end{cases}$$

Clearly, the conditional probability mass function of Z_2 given Z_1 is only degenerate specific cases:

(i) when $0 \leqslant p \leqslant q$, conditional on $Z_1 = 0$, $Z_2 = 0$ with probability 1;

(ii) when $0 \leqslant q \leqslant p$, conditional on $Z_1 = 1$, $Z_2 = 1$ with probability 1.

Notice from the above example that for comonotonic pairs with Bernoulli margins, when $p < q$ (or equivalently when $\mathrm{E}(Z_1) = (1-p) > \mathrm{E}(Z_2) = (1-q)$) one has that $Z_1 \geqslant Z_2$ while when $p > q$ (or equivalently when $\mathrm{E}(Z_1) = (1-p) < \mathrm{E}(Z_2) = (1-q)$) the relation is reversed and $Z_1 \leqslant Z_2$. This ordering holds in general, i.e., perfect positive dependence implies an ordering to the random variables. In particular, depending on the magnitude of the marginal parameters, either $Z_1 \geqslant Z_2$ or $Z_2 \geqslant Z_1$, almost surely. This follows from the fact that the quantile function is non-decreasing and that both Z_1 and Z_2 are generated from the *same* uniform random variable U. The following example outlines this result for the case of Poisson margins.

Example 3.3 Comonotonic Poisson random variables
Suppose that $(Z_1, Z_2) \sim \mathcal{M}\{\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)\}$, where $\mathcal{P}(\lambda)$ denotes the Poisson distribution with mean λ. The Poisson distribution function G_λ is a decreasing function of λ because for all $k \in \mathbb{N}$,

$$\partial G_\lambda(k)/\partial\lambda = -e^{-\lambda}\lambda^k/k! = -g_\lambda(k).$$

Suppose that $\lambda_2 > \lambda_1$. Then, for any $z_2 < z_1$, we also have that $z_2 \leqslant z_1 - 1$ since $(z_1, z_2) \in \mathbb{N}^2$, and the ordering $G_{\lambda_2}(z_2 - 1) < G_{\lambda_2}(z_2) < G_{\lambda_1}(z_2) < G_{\lambda_1}(z_1 - 1) < G_{\lambda_1}(z_1)$ ensues. It then follows that

$$\begin{aligned}\Pr(Z_1 \leqslant Z_2) &= 1 - \Pr(Z_2 < Z_1) = 1 - \Pr(Z_2 \leqslant Z_1 - 1)\\ &= 1 - \sum_{z_1=0}^{\infty} \Pr(Z_2 \leqslant Z_1 - 1, Z_1 = z_1) = 1 - \sum_{z_1=0}^{\infty}\sum_{z_2=0}^{z_1-1} \Pr(Z_1 = z_1, Z_2 = z_2)\end{aligned}$$

which can be rewritten as

$$\begin{aligned}1 - \sum_{z_1=0}^{\infty}\sum_{z_2=0}^{z_1-1} \Big[\min\{G_{\lambda_1}(z_1), G_{\lambda_2}(z_2)\} - \max\{G_{\lambda_1}(z_1-1), G_{\lambda_2}(z_2-1)\}\Big]_+ \\ = 1 - \sum_{z_1=0}^{\infty}\sum_{z_2=0}^{z_1-1} \left[G_{\lambda_2}(z_2) - G_{\lambda_1}(z_1-1)\right]_+ .\end{aligned}$$

However, the right-hand side is $1 - 0 = 1$, because $G_{\lambda_2}(z_2) < G_{\lambda_1}(z_1 - 1)$ for $z_2 \leqslant z_1 - 1$. Thus, for $(Z_1, Z_2) \sim \mathcal{M}\{\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)\}$, we have $\lambda_1 < \lambda_2 \;\Rightarrow Z_1 \;\leqslant Z_2$.

3.3 Comonotonic pairs: Data generation and visualization

It is clear from the stochastic representation, as given in Eq. (3.1), that a sample of comonotonic pairs $(Z_{11}, Z_{12}), \ldots, (Z_{n1}, Z_{n2})$ can easily be generated from a random sample of standard uniform variables. In particular, generate independent $U_1, \ldots, U_n \sim \mathcal{U}(0, 1)$ and for each $i \in \{1, \ldots, n\}$, set $(Z_{i1}, Z_{i2}) = (F_1^{-1}(U_i), F_2^{-1}(U_i))$.

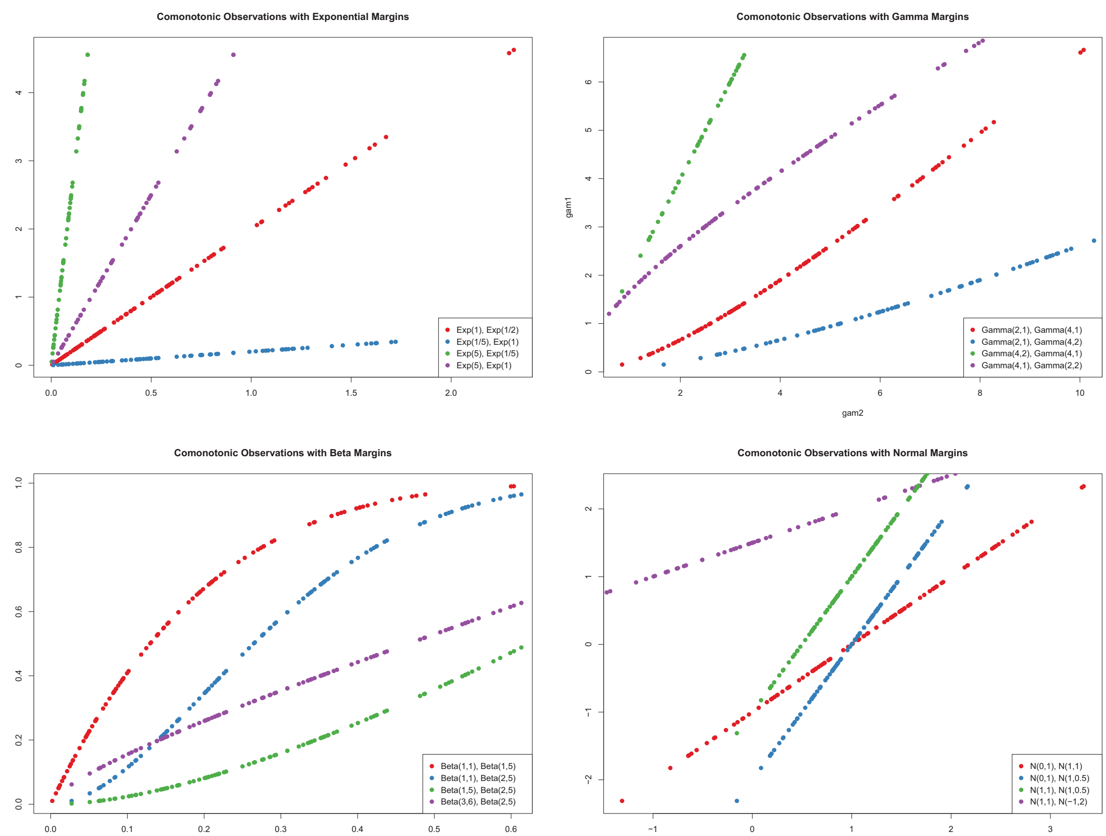

Figure 3.1: Comonotonic pairs for Exponential margins (top left), Gamma margins (top right), Beta margins (bottom left) and Normal margins (bottom right).

As already mentioned, when dealing with continuous random variables, this data generation procedure leads to a functional relation between the margins. Indeed, plotting the observations z_{i1} against z_{i2} for all $i \in \{1, \dots, n\}$ will exhibit a perfect relation that is induced by the form of the marginal quantile functions, namely F_1^{-1} and F_2^{-1}. The graphs shown in Figure 3.1 depict these relations for various choices of continuous marginal distributions. In contrast, plotting the ranks of the observations leads to a perfect linear relation as $F_1(Z_1) = F_2(Z_2)$ for comonotonic random variables with continuous margins.

When it comes to discrete random variables, comonotonicity does not necessarily imply a functional relationship between the margins, as was previously explored. This is caused by the form of the marginal cumulative distribution functions, which are step functions with jump discontinuities at each $x \in \mathbb{R}$ where the marginal probability mass function is non-zero. As a result, the inverse CDF is a many-to-one mapping and rather than having the relationship $F_1(z_{i1}) = F_2(z_{i2})$ for all $i \in \{1, \dots, n\}$, one has a relationship with regards to the marginal jumps where, for each

$i \in \{1, \ldots, n\}$,

$$(F_1(z_{i1}^-), F_1(z_{i1})] \cap (F_2(z_{i2}^-), F_2(z_{i2})] \neq \varnothing.$$

Thus, for each comonotonic observation (z_{i1}, z_{i2}), rather than knowing the value of the underlying uniform random variable u that generated the pair, one can only identify an interval in which u is contained, namely $u \in (F_1(z_{i1}^-), F_1(z_{i1})] \cap (F_2(z_{i2}^-), F_2(z_{i2})]$. In other words, comonotonic discrete random pairs imply a type of *interval censoring* on the underlying uniform random variable.

This interval relationship can be visualized as a type of "staircase" along the main diagonal: Suppose the unit square is partitioned into all possible rectangles of the form

$$R_{ij} = (F_1(i^-), F_1(i)] \times (F_2(j^-), F_2(j)].$$

Then, the x-axis grid lines represent the marginal cumulative distribution function jumps $F_1(x)$ for all possible $x \in \mathbb{R}$ such that $F_1(x) - F_1(x^-) > 0$, and likewise for the y-axis in terms of F_2. Then any rectangle R_{ij} that touches the main diagonal line corresponds to a value (i, j) that the comonotonic pair (Z_1, Z_2) can take on with non-zero probability, i.e., such that $\Pr(Z_1 = i, Z_2 = j) > 0$. This follows since the diagonal line $y = x$ represents the underlying uniform random variable u generating both Z_1 and Z_2 and so only rectangles that intersect with this diagonal line have the property that $(F_1(z_{i1}^-), F_1(z_{i1})] \cap (F_2(z_{i2}^-), F_2(z_{i2})] \neq \varnothing$. Figure 3.2 depict this relation for Bernoulli margins while Figure 3.3 shows the relation for Poisson margins. Graphing the observations (z_{i1}, z_{i2}) for all $i \in \{1, \ldots, n\}$ will also show a staircase pattern, reflecting the "interval dependence" between the margins, as shown in Figure 3.4 for various choices of discrete marginal distributions.

3.4 Parameter estimation for comonotonic random pairs

For comonotonic random pairs, the joint distribution function is the upper Fréchet–Hoeffding bound, i.e., for all $z_1, z_2 \in \mathbb{R}$, we have

$$\Pr(Z_1 \leqslant z_1, Z_2 \leqslant z_2) = \min\{F_1(z_1), F_2(z_2)\}. \tag{3.8}$$

It is clear from the above form that the joint CDF of (Z_1, Z_2) is parametrized only by the marginal parameters, say θ_1 and θ_2 for F_1 and F_2, respectively, where both θ_1 and θ_2 may be vectors. There is thus no notion of a "dependence parameter" and there are only marginal parameters to estimate. Accordingly, one could consider carrying out parameter estimation separately for both margins and certainly appropriate estimation techniques, e.g., the method of moments or maximum likelihood estimation, will lead to consistent estimates. As shown in, e.g. Result 2.1 of Wakefield (2013), as long as an estimator is based on an unbiased estimating equation, consistency of the resulting estimate ensues. More specifically, suppose that for observations $\mathbf{Y} = (Y_1, \ldots, Y_n)$, the estimator

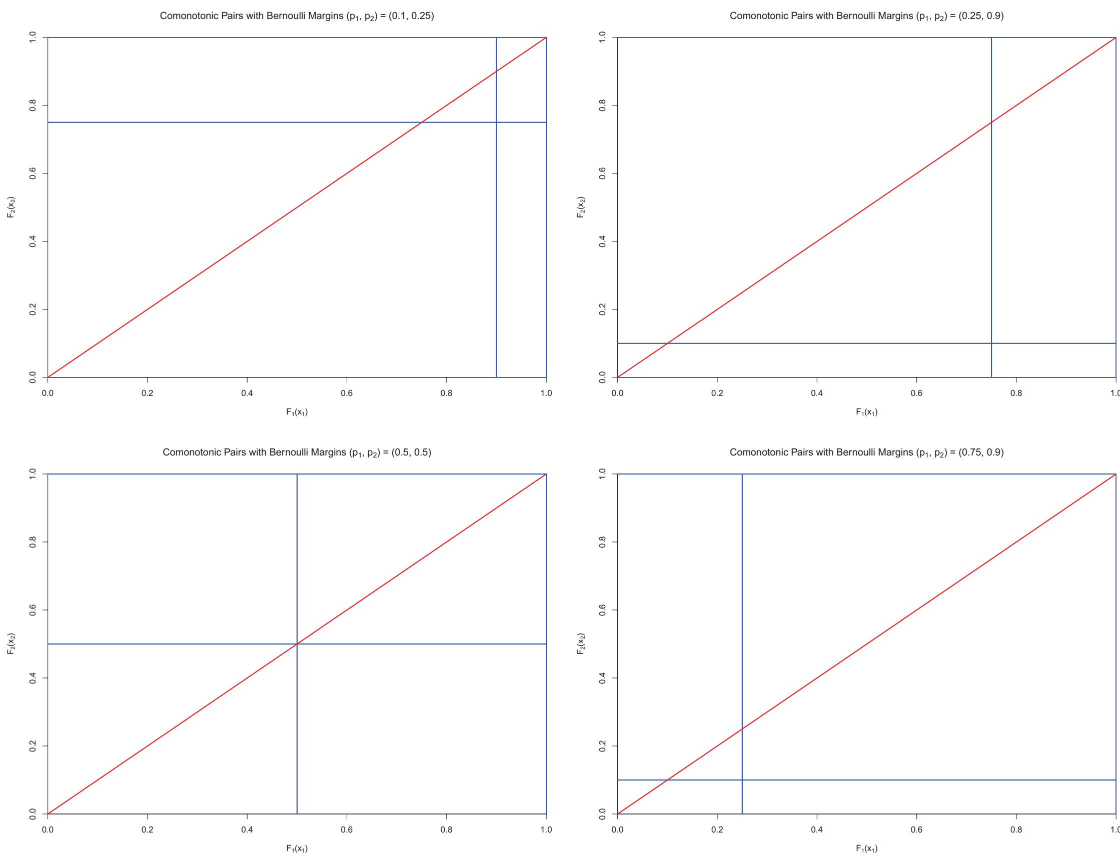

Figure 3.2: Comonotonic pairs cumulative distribution function gridlines for Bernoulli margins.

$\widehat{\mathbf{\Psi}}$ results from a general estimating equation of the form

$$\mathbf{G}_n(\mathbf{\Psi}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{G}(Y_i, \mathbf{\Psi}) = 0,$$

so that $\mathbf{G}_n(\widehat{\mathbf{\Psi}}) = 0$. Then it can be shown that $\widehat{\mathbf{\Psi}} \xrightarrow{p} \mathbf{\Psi}$.

The question then arises as to whether there is a loss in efficiency when estimation is done separately for θ_1 and θ_2 rather than considering a joint analysis. Note that separate analyses of marginal parameters via their respective marginal likelihoods is a common approach used for estimation in copula models. See, e.g., Section 10.1 of Joe (1997), who refers to this approach as the method of inference functions for margins (IFM). This estimation technique is usually preferable to a full maximum likelihood approach due to the complex form of the joint density, which renders optimization difficult. Certainly, in higher dimensions, the utility of this approach is even greater.

Focusing on the bivariate case, suppose that a copula C_ϕ characterizes the dependence between two random variables (X_1, X_2) with marginal distributions F_1 and F_2, respectively. According to Sklar's theorem, as stated in Theorem 2.3.3 of Nelsen (2006), there always exists a copula C_ϕ such

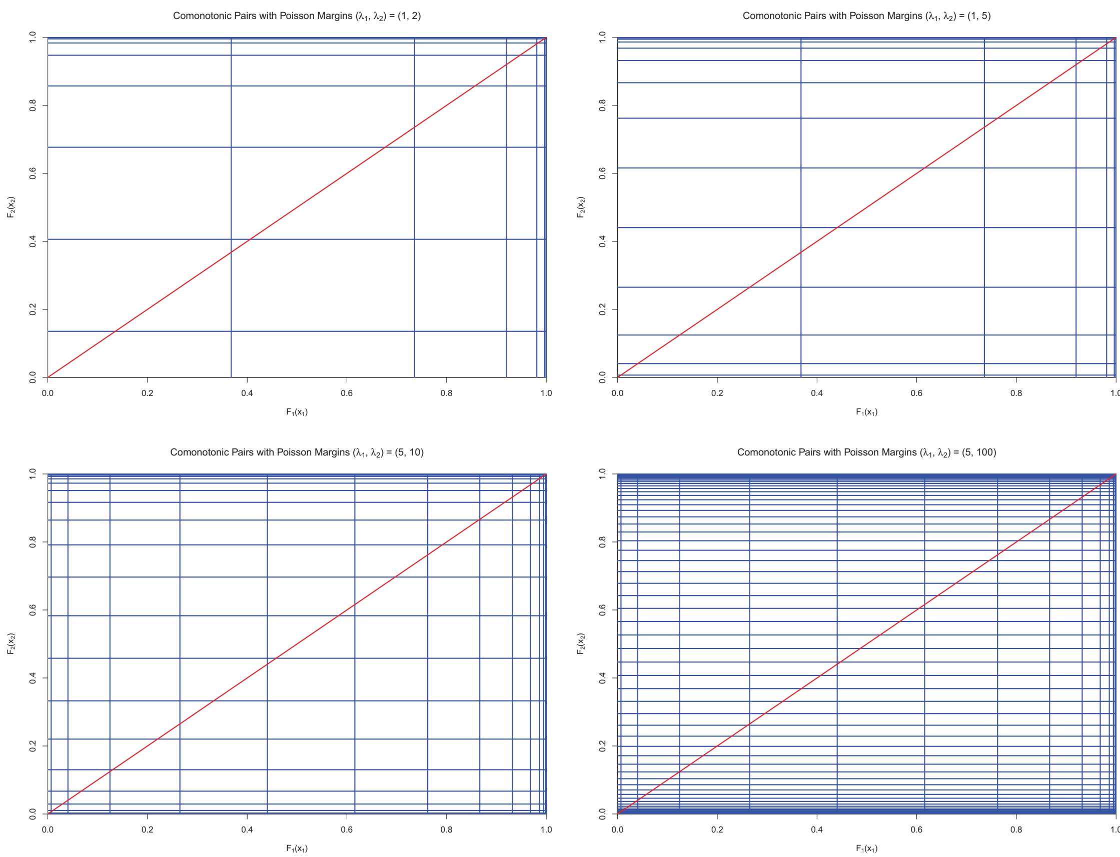

Figure 3.3: Comonotonic pairs cumulative distribution function gridlines for Poisson margins.

that the joint CDF of (X_1, X_2) has the form

$$H(x_1, x_2) = C_\phi\{F_1(x_1), F_2(x_2)\},$$

for all $(x_1, x_2) \in \bar{\mathbb{R}}^2$. In the case of continuous margins, the copula is unique and the joint density, when it exists, is obtained by differentiating $H(x_1, x_2)$ to obtain

$$h(x_1, x_2) = f_1(x_1)f_2(x_2)c_\phi\{F_1(x_1), F_2(x_2)\},$$

where f_1 and f_2 are the densities associated with F_1 and F_2, respectively, and for all $u_1, u_2 \in (0, 1)$,

$$c_\phi(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C_\phi(u_1, u_2).$$

Note that for discrete margins, a similar form for the joint probability function of (X_1, X_2) can be obtained by differencing the joint cumulative distribution function, although the underlying copula is unique only on supp$(F_1) \times$ supp(F_2). Given a set of n mutually independent observations

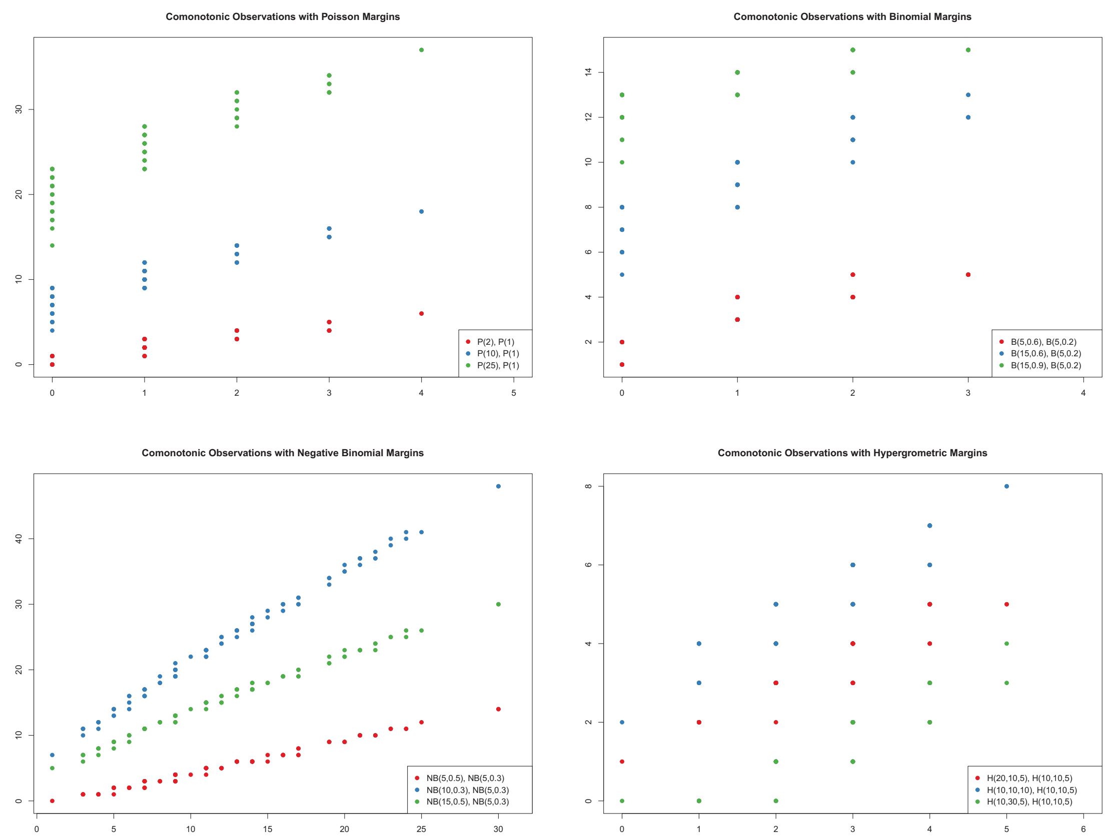

Figure 3.4: Comonotonic pairs for Poisson margins (top left), Binomial margins (top right), Negative Binomial margins (bottom left) and Hypergeometric margins (bottom right).

$(x_{11}, x_{12}), \ldots, (x_{n1}, x_{n2})$ from H, the corresponding log-likelihood has the form

$$\begin{aligned} \ell(\theta_1, \theta_2, \phi) &= \sum_{i=1}^{n} \left[\ln f_1(x_{i1}) + \ln f_2(x_{i2}) + \ln c_\phi \left\{ F_1(x_{i1}), F_2(x_{i2}) \right\} \right] \\ &= \ell_1(\theta_1) + \ell_2(\theta_2) + \ell_D(\theta_1, \theta_2, \phi), \end{aligned}$$

where $\ell_s(\theta_s)$ denotes the marginal log-likelihood for component $s \in \{1, 2\}$ and $\ell_D(\theta_1, \theta_2, \phi)$ is the remaining term encompassing the dependence.

For certain choices of copula, direct maximization of the full log-likelihood may be feasible but often, this is not the case. The IFM approach, as discussed in Joe (1997, 2005, 2015), suggests that estimation for the marginal parameters be carried out using their respective marginal log-likelihoods to obtain estimates $\check{\theta}_1$ and $\check{\theta}_2$. We will refer to the latter as the marginal maximum likelihood estimates or the IFM estimates. Estimation for the dependence parameter ϕ can then be accomplished by maximizing the portion of the log-likelihood embodying the dependence, while holding the marginal parameters fixed at their respective marginal maximum likelihood estimates, i.e.,

$$\check{\phi} = \arg\max_{\phi} \ \ell_D(\check{\theta}_1, \check{\theta}_2, \phi).$$

Note that, as pointed out by Joe (2005), this two-step procedure towards estimation is simply the computational implementation of the IFM approach. For theoretical purposes, under suitable regularity conditions, the IFM estimates are such that

$$\left(\frac{\partial}{\partial\theta_1}\ell_1(\theta_1), \frac{\partial}{\partial\theta_2}\ell_2(\theta_2), \frac{\partial}{\partial\phi}\ell(\theta_1,\theta_2,\phi)\right) = \mathbf{0}. \tag{3.9}$$

This approach is obviously not as computationally intensive as simultaneously estimating all parameters $(\theta_1, \theta_2, \phi)$ using the full log-likelihood and its benefits are likely to increase with the dimensionality of the problem. A full joint likelihood approach yields maximum likelihood estimates, denoted by $(\hat{\theta}_1, \hat{\theta}_2, \hat{\phi})$. Under regularity conditions, the MLEs are found as the solution to the set of score equations given by

$$\left(\frac{\partial}{\partial\theta_1}\ell(\theta_1,\theta_2,\phi), \frac{\partial}{\partial\theta_2}\ell(\theta_1,\theta_2,\phi), \frac{\partial}{\partial\phi}\ell(\theta_1,\theta_2,\phi)\right) = \mathbf{0}. \tag{3.10}$$

It will be assumed that the usual regularity conditions in the context of maximum likelihood theory hold for the joint model as well as all marginal specifications. Lending from the theory of estimating equations, one can derive the asymptotic variance of the IFM estimators. In particular, the IFM approach is simply a special case where the set of estimating equations takes on a certain form. Let g denote the general form of estimating equations, viz.

$$\sum_{i=1}^{n} \mathbf{g}(X_{i1}, X_{i2}, \Psi) = \mathbf{0},$$

where Ψ denotes the parameter vector $(\theta_1, \theta_2, \phi)$. Then, the IFM method is such that g is defined as in Eq. (3.9). That is, in the bivariate setting,

$$\mathbf{g}^\top = (\mathbf{g}_1^\top, \mathbf{g}_2^\top, \mathbf{g}_3^\top) = \left(\frac{\partial}{\partial\theta_1}\ell_1(\theta_1), \frac{\partial}{\partial\theta_2}\ell_2(\theta_2), \frac{\partial}{\partial\phi}\ell(\theta_1,\theta_2,\phi)\right).$$

As detailed in Joe (1997, 2005), it can be shown that the IFM estimators, denoted by $\check{\Psi} = (\check{\theta}_1, \check{\theta}_2, \check{\phi})$, are asymptotically Gaussian, viz.

$$\sqrt{n}\,(\check{\Psi} - \Psi) \rightsquigarrow \mathcal{N}(0, V), \tag{3.11}$$

where $\rightsquigarrow$ denotes weak convergence, also called convergence in distribution. Following the notation of Joe (2005), the asymptotic variance is given by

$$V = (-D_g^{-1}) M_g (-D_g^{-1})^\top,$$

for $M_g = \text{cov}\{\mathbf{g}(\mathbf{X};\Psi)\}$ and $D_g = \text{E}\{\partial \mathbf{g}(\mathbf{X};\Psi)/\partial\Psi^\top\}$, where $\mathbf{X}$ is used to denote the bivariate data $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$. Upon simplification, Joe (2005) shows that the asymptotic variance V is such that, for $j, k \in \{1, 2\}$,

$$\text{cov}(\check{\theta}_j, \check{\theta}_k) = \mathcal{J}_{jj}^{-1} \mathcal{J}_{jk} \mathcal{J}_{kk}^{-1}, \quad \text{cov}(\check{\theta}_j, \check{\phi}) = \mathcal{J}_{jj}^{-1} \sum_{k=1}^{2} \mathcal{J}_{jk} a_k^\top,$$

and

$$\text{cov}(\check{\phi}, \check{\phi}) = \mathcal{I}_{33}^{-1} + \sum_{j=1}^{2}\sum_{k=1}^{2} a_j \mathcal{J}_{jk} a_k^\top.$$

In the above, $\mathcal{J}_{jj}$ denotes the information matrix from the jth univariate margin while $\mathcal{I}$ denotes the joint model information matrix so that $\mathcal{I}_{33} = -\text{E}(\partial^2 \ell / \partial\phi\partial\phi^\top)$. Finally, $a_j = -\mathcal{I}_{33}^{-1}\mathcal{I}_{3j}\mathcal{J}_{jj}^{-1}$ for $j \in \{1, 2\}$. Note that these results readily extend to higher dimensions as outlined in Joe (1997, 2005). A detailed derivation of the resulting asymptotic variance of IFM estimators is given in Joe (1997) as well as Joe (2005).

From the above derivations, it is possible to quantify the loss in efficiency from using the IFM approach to estimate the marginal parameters rather than using the full likelihood. In particular, for $j \in \{1, 2\}$,

$$\mathcal{I}_{jj} = -\text{E}\left\{\frac{\partial^2}{\partial\theta_j^2}\,\ell(\theta_1, \theta_2, \phi)\right\} = -\text{E}\left\{\frac{\partial^2}{\partial\theta_j^2}\,\ell_j(\theta_j)\right\} - \text{E}\left\{\frac{\partial^2}{\partial\theta_j^2}\,\ell_D(\theta_1, \theta_2, \phi)\right\}.$$

Letting $\zeta_{jj} = -\text{E}\{\partial^2 \ell_D(\theta_1, \theta_2, \phi)/\partial\theta_j^2\}$, it follows that $\mathcal{I}_{jj} = \mathcal{J}_{jj} + \zeta_{jj}$. Accordingly, the loss in efficiency from using the IFM approach will be regulated by ζ_{jj}, which essentially quantifies the amount of information pertaining the marginal parameter θ_j that is contained in the dependence portion of the log-likelihood ℓ_D. Joe (2005) provides results from numerous simulation studies examining the asymptotic relative efficiency of IFM estimates in comparison to full maximum likelihood estimates.

The upper Fréchet–Hoeffding boundary distribution, which itself is a copula, can be regarded as a limiting case where there is no longer a dependence parameter, or when ϕ reaches the boundary of the parameter space for certain choices of copula C_ϕ. Joe (2005) considers several examples comparing a full maximum likelihood estimation approach to the IFM method in the case of the upper and lower Fréchet–Hoeffding boundary distributions and finds that for certain choices of marginally distributions, the two approaches coincide.

In the case where the marginal cumulative distribution functions are strictly increasing functions, there is a functional relationship between the margins induced by comonotonicity. Writing, say, Z_2 in terms of Z_1, i.e.,

$$Z_2 = F_2^{-1}\{F_1(Z_1)\},$$

one then has that the joint density of a comonotonic pair of continuous random variables (Z_1, Z_2) has the form

$$f(z_1, z_2) = f_1(z_1) f_{2|1}(z_2|z_1) = f_1(z_1)\mathbf{1}[z_2 = F_2^{-1}\{F_1(z_1)\}].$$

For observations $(z_{11}, z_{12}), \ldots, (z_{n1}, z_{n2})$ forming a random sample from $\mathcal{M}(F_1, F_2)$, the corresponding log-likelihood is then

$$\begin{aligned} \ell(\theta_1, \theta_2) &= \sum_{i=1}^{n} \ln \left[f_1(z_{i1})\mathbf{1}[z_{i2} = F_2^{-1}\{F_1(z_{i1})\}] \right] \\ &= \ell_1(\theta_1) + \sum_{i=1}^{n} \ln \mathbf{1}[z_{i2} = F_2^{-1}\{F_1(z_{i1})\}]. \end{aligned} \tag{3.12}$$

The first term of the above equation consists of the marginal log-likelihood corresponding to the first component Z_1. Clearly, $\ell_1(\theta_1)$ is maximized at the marginal MLE, i.e., the IFM estimate $\check{\theta}_1$.

The second term of (3.12) encompasses the dependence in the model. Since this quantity involves taking the log of indicator functions, it will either be equal to $-\infty$ or 0. The maximum value of 0 occurs when all observed values of z_{i2} follow the deterministic relation implied by comonotonicity, i.e., when $z_{i2} = F_2^{-1}\{F_1(z_{i1})\}$ for all $i \in \{1, \ldots, n\}$. Surely, there are several values of (θ_1, θ_2) that could yield

$$\sum_{i=1}^{n} \ln \mathbf{1}[z_{i2} = F_2^{-1}\{F_1(z_{i1})\}] = 0.$$

To illustrate this, recall Example 3.1 which considers a comonotonic pair of Exponential random variables. Suppose that the true marginal parameter values are β_{10} and β_{20}. Then, any value of β_1 and β_2 such that their ratio is equal to β_{10}/β_{20} would cause the second term in the log-likelihood to be equal to 0.

Accordingly, the form of the comonotonic log-likelihood in (3.12) results in maximum likelihood estimates $(\hat{\theta}_1, \hat{\theta}_2)$ where $\hat{\theta}_1 = \check{\theta}_1$ and $\hat{\theta}_2$ is deterministically ascertained as the value such that $z_{i2} = F_2^{-1}\{F_1(z_{i1})\}$ for all $i \in \{1, \ldots, n\}$. Thus, in the case of strictly increasing marginal distribution functions, IFM estimation will coincide with maximum likelihood estimation only when the form of the marginal estimates $\check{\theta}_1, \check{\theta}_2$ reflect the functional relationship between the margins determined by $Z_2 = F_2^{-1}\{F_1(Z_1)\}$.

Example 3.4 Maximum likelihood estimation for comonotonic pairs in the location-scale family

Suppose that Z_1 and Z_2 have strictly increasing cumulative distribution functions F_1 and F_2 respectively, that are members of the same location-scale family so that, for $s \in \{1, 2\}$,

$$F_s(z_s) = F\left(\frac{z_s - \mu_s}{\sigma_s}\right) \quad \Leftrightarrow \quad Z_s = \mu_s + \sigma_s F^{-1}(U),$$

where F is the standard distribution for the particular location-scale family in question. For continuous margins, the densities are then given, for $s \in \{1, 2\}$, by

$$f_s(z_1) = f\left(\frac{z_s - \mu_s}{\sigma_s}\right) \times \left(\frac{1}{\sigma_s}\right),$$

where $f(x) = \partial F(x)/\partial x$. Suppose that a random sample $(z_{11}, z_{12}), \ldots, (z_{n1}, z_{n2})$ is observed, where $(Z_1, Z_2) \sim \mathcal{M}(F_1, F_2)$. If a separate analysis of the marginal likelihoods is considered, then for each $s \in \{1, 2\}$, one obtains a log-likelihood function of the form

$$\ell_s(\mu_s, \sigma_s) = \sum_{i=1}^{n} \ln f\left(\frac{z_s - \mu_s}{\sigma_s}\right) - n \ln \sigma_s.$$

Differentiating the log-likelihood with respect to the parameters $\theta_s = (\mu_s, \sigma_s)$ yields the following score equations:

$$\begin{aligned}
\frac{\partial}{\partial \mu_s} \ell_s(\mu_s, \sigma_s) &= \sum_{i=1}^{n} -\frac{1}{\sigma_s} \left\{ \frac{f'\left(\frac{z_s - \mu_s}{\sigma_s}\right)}{f\left(\frac{z_s - \mu_s}{\sigma_s}\right)} \right\} = 0 \\
&\Leftrightarrow \quad \sum_{i=1}^{n} \left\{ \frac{f'\left(\frac{z_s - \mu_s}{\sigma_s}\right)}{f\left(\frac{z_s - \mu_s}{\sigma_s}\right)} \right\} = 0, \\
\frac{\partial}{\partial \sigma_s} \ell_s(\mu_s, \sigma_s) &= \sum_{i=1}^{n} \left[-\left(\frac{z_{is} - \mu_s}{\sigma_s^2}\right) \left\{ \frac{f'\left(\frac{z_s - \mu_s}{\sigma_s}\right)}{f\left(\frac{z_s - \mu_s}{\sigma_s}\right)} \right\} - \frac{1}{\sigma_s} \right] = 0 \\
&\Leftrightarrow \quad \sum_{i=1}^{n} \left[\left(\frac{z_{is} - \mu_s}{\sigma_s}\right) \left\{ \frac{f'\left(\frac{z_s - \mu_s}{\sigma_s}\right)}{f\left(\frac{z_s - \mu_s}{\sigma_s}\right)} \right\} \right] + n = 0,
\end{aligned}$$

where $f'(x) = \partial f(x)/\partial x$. For each $s \in \{1, 2\}$, the IFM estimates $\check{\mu}_s, \check{\sigma}_s$ are then determined as the roots to the above score equations. A full maximum likelihood approach would use the above score equations for one set of marginal parameters, say μ_1, σ_1, and then the remaining parameters μ_2, σ_2 would be computed according to the functional relation

$$Z_2 = \sigma_2 Z_1/\sigma_1 + \mu_2 - \sigma_2 \mu_1/\sigma_2.$$

However, the form of the score equations derived above are in fact identical for both margins (and hence redundant) since comonotonicity implies the deterministic relation

$$(Z_1 - \mu_1)/\sigma_1 = (Z_2 - \mu_2)/\sigma_2.$$

Thus, in the case of comonotonic pairs with strictly increasing distribution functions that are members of a common location-scale family, the IFM and ML estimation approaches will coincide. This result will thus hold, for example, in the case of Normal margins and Exponential margins.

For further examples and comparisons, please refer to Joe (2005).

When dealing with discrete margins, the problem becomes more difficult as perfect positive dependence does not necessarily imply a functional relationship between the margins. It is then no longer clear how the IFM approach will compare with full maximum likelihood estimation and rather the problem must be treated case by case. The following two examples illustrate both when the IFM and MLE agree and when the two methods do not coincide.

Example 3.5 Maximum likelihood estimation for comonotonic Bernoulli random variables

Suppose that $Z_1 \sim Bernoulli(p)$ *and* $Z_2 \sim Bernoulli(q)$ *such that* $\Pr(Z_1 = 0) = p$ *and* $\Pr(Z_2 = 0) = q$. *It was previously shown in Example 3.2 that if* $0 \leqslant p \leqslant q \leqslant 1$, *then* $Z_1 \geqslant Z_2$ *almost surely and the upper Fréchet–Hoeffding boundary joint probability mass function is given by:*

$$\begin{aligned}
\Pr(0,0) &= [\min(p,q) - \max(0,0)]_+ = p, \\
\Pr(1,0) &= [\min(1,q) - \max(p,0)]_+ = q - p, \\
\Pr(1,1) &= [\min(1,1) - \max(p,q)]_+ = 1 - q, \\
\Pr(0,1) &= [\min(p,1) - \max(0,q)]_+ = 0.
\end{aligned}$$

Suppose a random sample $(z_{11}, z_{12}), \ldots, (z_{n1}, z_{n2})$ *is obtained from the above joint distribution and that the observations are tabulated in a contingency table of the form*

m_{00}	m_{01}
m_{10}	m_{11}

where m_{ij} *denotes the number of observations such that* $(z_{i1}, z_{i2}) = (i, j)$. *Note that since* $Z_1 \geqslant Z_2$ *almost surely, the count* $m_{01} = 0$. *The joint likelihood can be written as*

$$L(p,q) = p^{m_{00}}(q-p)^{m_{10}}(1-q)^{m_{11}}$$

with corresponding log-likelihood given by

$$\ell(p,q) = m_{00}\ln p + m_{10}\ln(q-p) + m_{11}\ln(1-q).$$

The score equations are then

$$\frac{\partial}{\partial p}\,\ell(p,q) = \frac{m_{00}}{p} - \frac{m_{10}}{q-p} = 0 \quad \Leftrightarrow \quad p = \frac{m_{00}q}{m_{00}+m_{10}}\,, \tag{3.13}$$

$$\frac{\partial}{\partial q}\,\ell(p,q) = \frac{m_{10}}{q-p} - \frac{m_{11}}{1-q} = 0 \quad \Leftrightarrow \quad q = \frac{m_{10}+m_{11}p}{m_{10}+m_{11}}. \tag{3.14}$$

Substituting (3.13) *into* (3.14)*, one obtains*

$$q = \frac{m_{10}(m_{10}+m_{00}) + m_{11}m_{00}q}{(m_{10}+m_{11})(m_{10}+m_{00})}.$$

Simplifying the above yields the following:

$$\begin{aligned} q(m_{10}+m_{11})(m_{10}+m_{00}) &= m_{10}(m_{10}+m_{00}) + m_{11}m_{00}\,q \\ \Leftrightarrow q(m_{10}^2 + m_{10}m_{00} + m_{11}m_{10} + m_{11}m_{00} - m_{11}m_{00}) &= m_{10}(m_{10}+m_{00}) \\ \Leftrightarrow qm_{10}(m_{10}+m_{00}+m_{11}) &= m_{10}(m_{10}+m_{00}) \\ \Leftrightarrow q &= \frac{m_{10}+m_{00}}{m_{10}+m_{00}+m_{11}} \end{aligned}$$

and thus

$$p = \frac{m_{00}q}{m_{10}+m_{00}} = \frac{m_{00}}{m_{10}+m_{00}+m_{11}}.$$

Notice that in this example, the marginal parameter estimators are linearly dependent.

Since $m_{01} = 0$*, the total number of observations is the sum* $n = m_{00} + m_{10} + m_{11}$*. Let* m_{+0} *denote the number of observations such that* $Z_2 = 0$*, i.e.,* $m_{+0} = m_{00} + m_{10}$*. Similarly, let* m_{0+} *denote the number of observations with* $Z_1 = 0$*, so that* $m_{0+} = m_{00}$*. Then, the maximum likelihood estimates are*

$$\hat{p} = m_{0+}/n, \quad \hat{q} = m_{+0}/n.$$

The above estimates $\hat{p}$ *and* $\hat{q}$ *agree with the marginal maximum likelihood estimates, i.e.,* $\check{p} = \hat{p}$ *and* $\check{q} = \hat{q}$*. Indeed, the marginal likelihood for the two components are given by*

$$L_1(p) = p^{m_{0+}}(1-p)^{n-m_{0+}}, \quad L_2(q) = q^{m_{+0}}(1-q)^{n-m_{+0}},$$

with corresponding log-likelihoods

$$\ell_1(p) = m_{0+}\ln\left(\frac{p}{1-p}\right) + n\ln(1-p), \quad \ell_2(q) = m_{+0}\ln\left(\frac{q}{1-q}\right) + n\ln(1-q).$$

This yields the following set of marginal score equations:

$$\frac{\partial}{\partial p}\ell_1(p) = \frac{1}{1-p}\left(\frac{m_{0+}}{p} - n\right), \quad \frac{\partial}{\partial q}\ell_2(q) = \frac{1}{1-q}\left(\frac{m_{+0}}{q} - n\right).$$

Setting the score equations to zero yields the marginal maximum likelihood, or IFM, estimates, viz.

$$\check{p} = m_{0+}/n, \quad \check{q} = m_{+0}/n.$$

Thus, the estimates obtained from the IFM method and those from a full maximum likelihood approach coincide in the case of comonotonic Bernoulli random pairs. Seemingly, the manner in which the dependence is reflected in the data (in this specific example, by setting $m_{01} = 0$) adjusts the marginal log-likelihoods appropriately.

Standard maximum likelihood theory ensures that as $n \to \infty$, one has that marginally

$$\sqrt{n}\,(\hat{p} - p) \rightsquigarrow \mathcal{N}[0, p(1-p)], \quad \sqrt{n}\,(\hat{q} - q) \rightsquigarrow \mathcal{N}[0, q(1-q)].$$

Working with the full likelihood, one can obtain an expression for the expected Fisher Information matrix $\mathcal{I}_n$. For $\Psi = (p,q)^\top$, it can be shown that

$$\mathcal{I}_n = -\mathrm{E}\left\{\frac{\partial^2}{\partial\Psi\partial\Psi^\top}\,\ell(p,q)\right\} = \frac{n}{(q-p)}\begin{bmatrix} q/p & -1 \\ -1 & (1-p)/(1-q)\end{bmatrix}.$$

The joint asymptotic distribution of the MLEs $\hat{\Psi} = (\hat{p},\hat{q})^\top$ is then given by

$$\sqrt{n}\,(\hat{\Psi} - \Psi) \rightsquigarrow \mathcal{N}(0, V),$$

where

$$V = p(1-q)\begin{bmatrix} (1-p)/(1-q) & 1 \\ 1 & q/p\end{bmatrix}.$$

It is clear from this example that although the IFM approach will yield asymptotically efficient estimates for the marginal parameters, this method systematically ignores the correlation between $\hat{p}$ and $\hat{q}$.

Example 3.6 Maximum likelihood estimation for comonotonic Poisson random variables

Suppose that Z_1 and Z_2 are comonotonic Poisson random variables with marginal means λ_1 and λ_2, respectively. It is possible to find an example where the IFM and maximum likelihood estimates do not concur, as will be done using R. Specifying the seed as 1234, a random sample of size $n = 10$ comonotonic Poisson random pairs (z_{i1}, z_{i2}) may be generated from a common sample of uniform observations $u_1, \ldots, u_n$. Precisely, for $\lambda_1 = 5$ and $\lambda_2 = 2$, one obtains the following sample:

Z_1	2	6	5	6	7	6	1	3	6	5
Z_2	0	2	2	2	4	2	0	1	2	2

The marginal maximum likelihood estimates are simply the marginal sample means, and hence one obtains $\check{\lambda}_1 = \bar{z}_1 = 4.7$ and $\check{\lambda}_2 = \bar{z}_2 = 1.7$. However, if (Z_1, Z_2) are comonotonic, one has that

$$\Pr(Z_1 = i, Z_2 = j) = \big[\min\{F_1(i), F_2(j)\} - \max\{F_1(i-1), F_2(j-1)\}\big]_+.$$

Setting $\lambda_1 = \check{\lambda}_1$ yields the interval $(F_1(6), F_1(7)] = (0.8046051, 0.8960312]$ while for $\lambda_2 = \check{\lambda}_2$ one obtains $(F_2(3), F_2(4)] = (0.9068106, 0.9703852]$. Since the two intervals do not overlap, it follows that setting $(\lambda_1, \lambda_2) = (\check{\lambda}_1, \check{\lambda}_2)$ results in $\Pr(Z_1 = 7, Z_2 = 4) = 0$. Thus, in this example, the IFM estimates cannot possibly coincide with the estimates obtained from optimizing the full likelihood as the marginal sample means yield an observed likelihood of zero!

Note that although marginal maximum likelihood estimates may deviate from those obtained by optimizing the full likelihood, as long as the marginal specifications are correct, both methods will provide consistent estimates for the marginal parameters (θ_1, θ_2). It then follows that the difference between IFM and ML estimates should tend to zero as the sample size increases. It is also interesting to remark that the IFM method is equivalent to maximum likelihood estimation in the case of independence. This implies that in situations where the IFM and maximum likelihood methods coincide, estimation of the marginal parameters are not affected by the dependence structure.

3.5 Comonotonic Poisson random pairs

Comonotonic Poisson random pairs present an interesting example. As illustrated in Example 3.6, estimating the marginal parameters via their respective marginal likelihoods could lead to inappropriate estimates. This in turn implies that method of moments estimation is not suitable as this approach also estimates the marginal parameters by their marginal sample means $(\bar{z}_1, \bar{z}_2)$. Seemingly, a full maximum likelihood approach is preferable in this case.

As previously discussed, for $(Z_1, Z_2) \sim \mathcal{M}\{\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)\}$, the joint probability mass function has the form

$$\Pr(Z_1 = z_1, Z_2 = z_2) = \big[\min\{G_{\lambda_1}(i), G_{\lambda_2}(j)\} - \max\{G_{\lambda_1}(i-1), G_{\lambda_2}(j-1)\}\big]_+,$$

where G_λ is used to denote the cumulative distribution function of a Poisson random variable with mean λ. Then, for a random sample $(z_{11}, z_{12}), \ldots, (z_{n1}, z_{n2})$, the log-likelihood is given by

$$\ell(\lambda_1, \lambda_2) = \sum_{i=1}^{n} \ln\Big[\big[\min\{G_{\lambda_1}(i), G_{\lambda_2}(j)\} - \max\{G_{\lambda_1}(i-1), G_{\lambda_2}(j-1)\}\big]_+\Big].$$

Maximum likelihood estimates $(\hat{\lambda}_1, \hat{\lambda}_2)$ are then determined as

$$(\hat{\lambda}_1, \hat{\lambda}_2) = \underset{\lambda_1,\lambda_2}{\arg\max}\ \ell(\lambda_1, \lambda_2).$$

Certainly, numerical optimization techniques may be used to solve the above problem. In this case, the marginal MLEs or method of moments estimates could be used as starting values.

In the case of Poisson comonotonic pairs, the log-likelihood can be reformulated in terms of pairwise order statistics. Suppose that it is observed that $z_{i1} \leqslant z_{i2}$ for each $i \in \{1, \ldots, n\}$. As was previously shown, this ordering suggests that $\lambda_1 < \lambda_2$. In this situation, it is convenient to consider conditioning on the observed values of Z_1 to reformulate the joint probability mass function. Let N_1 denote the number of unique observed values of Z_1 and for $k \in \{1, \ldots, N_1\}$, define the following:

$z_{1(k)}$: the corresponding order statistics for Z_1;
n_k: the number of unique values of Z_2 for which $Z_1 = z_{1(k)}$;
$z_{2(j)}^{(k)}$: the corresponding order statistics for Z_2 conditional on $Z_1 = z_{1(k)}$;
m_{kj}: the number of observations such that $(Z_1, Z_2) = (z_{1(k)}, z_{2(j)}^{(k)})$;
m_{k0}: the number of observations in the special case where $n_k = 1$.

In terms of a contingency table, this formulation sets the ordered values of Z_1 as the rows while the Z_2 values appear in the columns. This set-up characterizes the values of Z_2 row-by-row, i.e., conditionally on the value of Z_1. One could equivalently consider reversing the row and column specifications. Note that in the case where $\lambda_1 < \lambda_2$, defining the contingency table in this way populates the cells in a convenient manner. In particular, when $\lambda_1 < \lambda_2$, one has that $Z_1 \leqslant Z_2$ and thus there will be fewer unique values of Z_1 observed as compared to Z_2. In this case, setting Z_1 as the rows leads to a staircase like pattern of non-zero cells diagonally down the table. Reversing the columns and rows would simply imply steeper steps. This set-up is convenient, as will be made more obvious in defining the probabilities in terms of the paired order statistics.

For example, the following contingency table summarizes a random sample of $n = 100$ comonotonic pairs with margins $Z_1 \sim \mathcal{P}(1)$ and $Z_2 \sim \mathcal{P}(2)$. The unique values of Z_1 are given in the rows while the columns display the unique values of Z_2. Since $\lambda_1 < \lambda_2$, it is observed that $z_{i1} \leqslant z_{i2}$ for each $i \in \{1, \ldots, 100\}$. Notice that the non-zero cells in the contingency fall along the diagonal forming a staircase like pattern.

	0	1	2	3	4	5
0	11	27	0	0	0	0
1	0	10	23	5	0	0
2	0	0	0	14	4	0
3	0	0	0	0	2	4

In what follows, it will be assumed that for each row, i.e., each observed value $z_{1(k)}$, there are no sparsity issues so that all cells corresponding to a non-zero probability in the contingency table accordingly have non-zero observations. Of course, in practical applications this is unrealistic, especially for very large values of λ_1 and λ_2, in which case the marginal Poisson distributions will become approximately Normal. By working with the order statistics of one component of the comonotonic pair, here Z_1, along with the conditional order statistics of the remaining component, i.e., $Z_2 \mid Z_1$, the joint probability mass function can be simplified case by case. Correspondingly, the log-likelihood can be rewritten in terms of the order and contidional order statistics of an observed set of comonotonic Poisson pairs.

Consider the kth row of the contingency table where $Z_1 = z_{1(k)}$ and the corresponding Z_2 values, ordered by $z_{2(1)}^{(k)} < \cdots < z_{2(n_k)}^{(k)}$. Comonotonicity implies a particular arrangement of the intersection of the marginal quantile intervals

$$(G_{\lambda_1}(z_{1(k)} - 1), G_{\lambda_1}(z_{1(k)})], (G_{\lambda_2}(z_{2(1)}^{(k)} - 1), G_{\lambda_2}(z_{2(1)}^{(k)})], \ldots, (G_{\lambda_2}(z_{2(n_k)}^{(k)} - 1), G_{\lambda_2}(z_{2(n_k)}^{(k)})].$$

First consider the special case where $n_k = 1$ so that there is only one possible value that Z_2 can take on when $Z_1 = z_{1(k)}$, say $z_{2(0)}^{(k)}$. In this case, the conditional probability mass function of Z_2 given that $Z_1 = z_{1(k)}$ must be a point mass at $z_{2(0)}^{(k)}$. It follows that the intersection of the corresponding quantile intervals must be given by

$$G_{\lambda_2}(z_{2(0)}^{(k)} - 1) \leqslant G_{\lambda_1}(z_{1(k)} - 1) < G_{\lambda_1}(z_{1(k)}) \leqslant G_{\lambda_2}(z_{2(0)}^{(k)}).$$

This then yields

$$\Pr\left\{Z_1 = z_{1(k)}, Z_2 = z_{2(0)}^{(k)}\right\} = G_{\lambda_1}(z_{1(k)}) - G_{\lambda_1}(z_{1(k)} - 1) = g_{\lambda_1}(z_{1(k)}),$$

and thus

$$\Pr\left\{Z_2 = z_{2(0)}^{(k)} \mid Z_1 = z_{1(k)}\right\} = 1.$$

Now consider the case where $n_k > 1$. The minimum value of Z_2 along the kth row, i.e., $z_{2(1)}^{(k)}$, is such that the intersection of the corresponding marginal quantile intervals is given by

$$G_{\lambda_2}(z_{2(1)}^{(k)} - 1) \leqslant G_{\lambda_1}(z_{1(k)} - 1) \leqslant G_{\lambda_2}(z_{2(1)}^{(k)}) \leqslant G_{\lambda_1}(z_{1(k)}). \tag{3.15}$$

This ordering must hold, for reasons given below.

(a) If either $G_{\lambda_2}(z_{2(1)}^{(k)}) < G_{\lambda_1}(z_{1(k)} - 1)$ or $G_{\lambda_1}(z_{1(k)}) < G_{\lambda_2}(z_{2(1)}^{(k)} - 1)$, then the intervals $(G_{\lambda_1}(z_{1(k)} - 1), G_{\lambda_1}(z_{1(k)})]$ and $(G_{\lambda_2}(z_{2(1)}^{(k)} - 1), G_{\lambda_2}(z_{2(1)}^{(k)})]$ would not overlap and the probability of observing the pair $(z_{1(k)}, z_{2(1)}^{(k)})$ would be zero.

(b) If $G_{\lambda_1}(z_{1(k)} - 1) < G_{\lambda_2}(z_{2(1)}^{(k)} - 1)$ then one could either have the ordering $G_{\lambda_2}(z_{2(1)}^{(k)} - 2) <$

$G_{\lambda_1}(z_{1(k)} - 1) < G_{\lambda_2}(z_{2(1)}^{(k)} - 1)$ or $G_{\lambda_2}(z_{2(1)}^{(k)} - m - 1) < G_{\lambda_1}(z_{1(k)} - 1) < G_{\lambda_2}(z_{2(1)}^{(k)} - m) < G_{\lambda_2}(z_{2(1)}^{(k)} - 1)$ for some $m \in \{2, 3, \ldots\}$. In either case, this would imply that the minimum possible value of Z_2 observable with $Z_1 = z_{1(k)}$ would not be equal to $z_{2(1)}^{(k)}$, which contradicts the definition of $z_{2(1)}^{(k)}$ in the first place.

(c) If $G_{\lambda_2}(z_{2(1)}^{(k)} - 1) \leqslant G_{\lambda_1}(z_{1(k)} - 1) \leqslant G_{\lambda_1}(z_{1(k)}) \leqslant G_{\lambda_2}(z_{2(1)}^{(k)})$, then one would have $n_k = 1$.

The ordering given in (3.15) implies that the cell corresponding to $(z_{1(k)}, z_{2(1)}^{(k)})$ has probability

$$\Pr\left\{Z_1 = z_{1(k)}, Z_2 = z_{2(1)}^{(k)}\right\} = G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)} - 1).$$

Using similar arguments, it can be shown that for the maximum value of Z_2 conditional on $Z_1 = z_{1(k)}$, intersecting the appropriate quantile intervals leads to the ordering

$$G_{\lambda_1}(z_{1(k)} - 1) \leqslant G_{\lambda_2}(z_{2(n_k)}^{(k)} - 1) \leqslant G_{\lambda_1}(z_{1(k)}) \leqslant G_{\lambda_2}(z_{2(n_k)}^{(k)}). \tag{3.16}$$

This implies that

$$\Pr\left\{Z_1 = z_{1(k)}, Z_2 = z_{2(n_k)}^{(k)}\right\} = G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)} - 1).$$

Finally, because distribution functions are non-decreasing, for all other values of Z_2 observable with $Z_1 = z_{1(k)}$, the following ordering must hold:

$$G_{\lambda_1}(z_{1(k)} - 1) \leqslant G_{\lambda_2}(z_{2(j)}^{(k)} - 1) \leqslant G_{\lambda_2}(z_{2(j)}^{(k)}) \leqslant G_{\lambda_1}(z_{1(k)})$$

for all $j \in \{2, \ldots, n_k - 1\}$. Therefore,

$$\Pr\left\{Z_1 = z_{1(k)}, Z_2 = z_{2(j)}^{(k)}\right\} = g_{\lambda_2}(z_{2(j)}).$$

Putting everything together, it follows that for each row corresponding to $Z_1 = z_{1(k)}$ with $n_k > 1$, the intersection of the quantile intervals is given by

$$\begin{aligned} G_{\lambda_2}(z_{2(1)}^{(k)} - 1) \leqslant G_{\lambda_1}(z_{1(k)} - 1) \leqslant G_{\lambda_2}(z_{2(1)}^{(k)}) \cdots < G_{\lambda_2}(z_{2(j)}^{(k)} - 1) < \\ G_{\lambda_2}(z_{2(j)}^{(k)}) < \cdots < G_{\lambda_2}(z_{2(n_k)}^{(k)} - 1) \leqslant G_{\lambda_1}(z_1(k)) \leqslant G_{\lambda_2}(z_{2(n_k)}^{(k)}). \end{aligned} \tag{3.17}$$

Notice that since the cumulative distribution function is non-decreasing, in the setting where the contingency table has non-zero counts wherever the corresponding probability is non-zero, one must also have that

$$z_{2(n_k)}^{(k)} = z_{2(1)}^{(k+1)}$$

for all $k \in \{1, \ldots, N_1 - 1\}$ and $z_{1(k)} + 1 = z_{1(k+1)}$.

Working with the reformulated joint probabilities involving the order statistics $(z_{1(k)}, z_{2(j)}^{(k)})$, one can then write the likelihood function as

$$\begin{aligned} L(\lambda_1, \lambda_2) &= \prod_{k=1}^{N_1} \prod_{j=1}^{n_k} \left[\Pr\left\{ Z_1 = z_{1(k)}, Z_2 = z_{2(j)}^{(k)} \right\} \right]^{m_{kj}} \\ &= \prod_{k=1}^{N_1} \left[\left\{ g_{\lambda_1}(z_{1(k)}) \right\}^{m_{k0}} \times \left\{ G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)} - 1) \right\}^{m_{k1}} \right. \\ &\qquad \left. \times \prod_{j=2}^{n_k - 1} \left\{ g_{\lambda_2}(z_{2(j)}^{(k)}) \right\}^{m_{kj}} \times \left\{ G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)} - 1) \right\}^{m_{kn_k}} \right]. \end{aligned}$$

The log-likelihood is then obtained as

$$\begin{aligned} \ell(\lambda_1, \lambda_2) = \sum_{k=1}^{N_1} \Bigg[& m_{k0} \left\{ -\lambda_1 + z_{1(k)} \ln(\lambda_1) - \ln(z_{1(k)}!) \right\} \\ & + m_{k1} \ln \left\{ G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)} - 1) \right\} \\ & + \sum_{j=2}^{n_k - 1} m_{kj} \left\{ -\lambda_2 + z_{2(j)}^{(k)} \ln(\lambda_2) - \ln(z_{2(j)}^{(k)}!) \right\} \\ & + m_{kn_k} \ln \left\{ G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)} - 1) \right\} \Bigg]. \end{aligned} \tag{3.18}$$

Note that in order for (3.18) to be correct, it is only necessary for $z_{2(1)}^{(k)}$ and $z_{2(n_k)}^{(k)}$ to respectively represent the minimum and maximum value that Z_2 can take on with non-zero probability given that $Z_1 = z_{1(k)}$. This follows since the conditional minimum and maximum values of Z_2 conditional on Z_1 dictate the cut-off points where the joint probability mass function switches amongst the plausible cases given in (3.7).

It is straightforward to show that the Poisson cumulative distribution function G_λ is decreasing in λ. Thus, in assigning Z_1 to the rows of the contingency table, where $\lambda_1 < \lambda_2$, one ensures that there are less incidences where $n_k = 1$. All cases where $n_k = 1$ leads to $m_{k0} > 0$ and accordingly $m_{k1} = \cdots = m_{kn_k} = 0$ so that only one component is contributing to the likelihood. Thus, when conditioning on the rows in the bivariate Poisson contingency table, it is more convenient to place the component with the smaller mean as the rows of the table.

The form of the log-likelihood in (3.18) allows to establish a set of score equations. In terms of the first parameter λ_1, the score equation is given by:

$$\frac{\partial}{\partial \lambda_1} \ell(\lambda_1,\lambda_2) = \sum_{k=1}^{N_1} \Bigg[m_{k0} \left\{ \frac{z_{1(k)}}{\lambda_1} - 1 \right\} + m_{k1} \left\{ \frac{g_{\lambda_1}(z_{1(k)}-1)}{G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)}-1)} \right\} + m_{kn_k} \left\{ \frac{-g_{\lambda_1}(z_{1(k)})}{G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)}-1)} \right\} \Bigg]. \tag{3.19}$$

For λ_2, one obtains

$$\frac{\partial}{\partial \lambda_2} \ell(\lambda_1,\lambda_2) = \sum_{k=1}^{N_1} \Bigg[\sum_{j=2}^{n_k-1} m_{kj} \left\{ \frac{z_{2(j)}^{(k)}}{\lambda_2} - 1 \right\} + m_{k1} \left\{ \frac{-g_{\lambda_2}(z_{2(1)}^{(k)})}{G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)}-1)} \right\} + m_{kn_k} \left\{ \frac{g_{\lambda_2}(z_{2(n_k)}^{(k)}-1)}{G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)}-1)} \right\} \Bigg]. \tag{3.20}$$

Using the recurrence relation $g_\lambda(k-1) = k/\lambda g_\lambda(k)$, Eqs. (3.19)–(3.20) can be rewritten as follows:

$$\frac{\partial}{\partial \lambda_1} \ell(\lambda_1,\lambda_2) = \sum_{k=1}^{N_1} \Bigg[\frac{z_{1(k)}}{\lambda_1} \left\{ m_{k0} + m_{k1} \left\{ \frac{g_{\lambda_1}(z_{1(k)})}{G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)}-1)} \right\} \right\} - m_{k0} - m_{kn_k} \left\{ \frac{g_{\lambda_1}(z_{1(k)})}{G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)}-1)} \right\} \Bigg],$$

$$\frac{\partial}{\partial \lambda_2} \ell(\lambda_1,\lambda_2) = \sum_{k=1}^{N_1} \Bigg[\frac{1}{\lambda_2} \left\{ \sum_{j=2}^{n_k-1} m_{kj} z_{2(j)}^{(k)} + m_{kn_k} z_{2(n_k)}^{(k)} \left\{ \frac{g_{\lambda_2}(z_{2(n_k)}^{(k)})}{G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)}-1)} \right\} \right\} - \sum_{j=2}^{n_k-1} m_{kj} - m_{k1} \left\{ \frac{g_{\lambda_2}(z_{2(1)}^{(k)})}{G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)}-1)} \right\} \Bigg].$$

The above score equations can be further simplified by reformulating the counts in terms of the conditional probabilities. To this end, define the following modified counts:

$$\omega_{k0} = m_{k0} \left\{ \Pr(Z_2 = z_{2(0)}^{(k)} \mid Z_1 = z_{1(k)}) \right\}^{-1} = m_{k0} \times 1,$$

$$\omega_{k1} = m_{k1} \left\{ \Pr(Z_2 = z_{2(1)}^{(k)} \mid Z_1 = z_{1(k)}) \right\}^{-1} = m_{k1} \left\{ \frac{g_{\lambda_1}(z_{1(k)})}{G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)}-1)} \right\},$$

$$\omega_{kn_k} = m_{kn_k} \left\{ \Pr(Z_2 = z_{2(n_k)}^{(k)} \mid Z_1 = z_{1(k)}) \right\}^{-1} = m_{kn_k} \left\{ \frac{g_{\lambda_1}(z_{1(k)})}{G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)}-1)} \right\},$$

$$\upsilon_{k1} = m_{k1} \left\{ \Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(1)}^{(k)}) \right\}^{-1} = m_{k1} \left\{ \frac{g_{\lambda_2}(z_{2(1)}^{(k)})}{G_{\lambda_2}(z_{2(1)}^{(k)}) - G_{\lambda_1}(z_{1(k)} - 1)} \right\},$$
$$\upsilon_{kj} = m_{kj} \left\{ \Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(j)}^{(k)}) \right\}^{-1} = m_{kj} \times 1,$$
$$\upsilon_{kn_k} = m_{kn_k} \left\{ \Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(n_k)}^{(k)}) \right\}^{-1} = m_{kn_k} \left\{ \frac{g_{\lambda_2}(z_{2(n_k)}^{(k)})}{G_{\lambda_1}(z_{1(k)}) - G_{\lambda_2}(z_{2(n_k)}^{(k)} - 1)} \right\}.$$

In terms of these modified counts, the score equations can be simplified to yield the following maximum likelihood estimators:

$$\hat{\lambda}_1 = \frac{\sum_{k=1}^{N_1} (\omega_{k0} + \omega_{k1})\, z_{1(k)}}{\sum_{k=1}^{N_1} (\omega_{k0} + \omega_{kn_k})}, \tag{3.21}$$

$$\hat{\lambda}_2 = \frac{\sum_{k=1}^{N_1} \sum_{j=2}^{n_k} \upsilon_{kj}\, z_{2(j)}^{(k)}}{\sum_{k=1}^{N_1} \sum_{j=1}^{n_k - 1} \upsilon_{kj}}. \tag{3.22}$$

Note that the modified counts are functions of the unknown parameters λ_1 and λ_2 and thus the maximum likelihood estimates are fount by simultaneously solving Eqs. (3.21)–(3.22).

Remark 3.1 In the trivial case where the marginal Poisson rates are identical, i.e., $\lambda_1 = \lambda_2$, it follows that $Z_1 = Z_2$ almost surely. Following the construction given in (3.18), the joint log-likelihood would reduce to the marginal log-likelihood for Z_1 as it would be observed that $m_{k0} > 0$ for each $k \in \{1, \ldots, N_1\}$. That is, the joint log-likelihood $\ell(\lambda_1, \lambda_2)$ reduces to $\ell_1(\lambda_1)$, where the latter denotes the marginal log-likelihood in the univariate Poisson setting. Thus, one obtains $\hat{\lambda}_1 = \bar{z}_1$ while there is no estimate for λ_2. In that sense, the set-up involving pairwise order statistics recognizes that the problem is really univariate in that $(Z_1, Z_2) = (Z_1, Z_1)$.

The Hessian matrix of the log-likelihood is established by differentiating the score equations (3.19) and (3.20). With some work, the elements of the Hessian matrix can be shown to have the following form:

$$\begin{aligned}
\frac{\partial^2}{\partial\lambda_1^2}\,\ell(\lambda_1,\lambda_2) = \sum_{k=1}^{N_1}\Bigg[& -m_{k0}\frac{z_{1(k)}}{\lambda_1^2} - m_{k1}\frac{z_{1(k)}}{\lambda_1}\left\{1-\frac{(z_{1(k)}-1)}{\lambda_1}\right\}\left\{\Pr(Z_2 = z_{2(1)}^{(k)} \mid Z_1 = z_{1(k)})\right\}^{-1} \\
& - m_{k1}\left(\frac{z_{1(k)}}{\lambda_1}\right)^2\left\{\Pr(Z_2 = z_{2(1)}^{(k)} \mid Z_1 = z_{1(k)})\right\}^{-2} \\
& - m_{kn_k}\left(\frac{z_{1(k)}}{\lambda_1}-1\right)\left\{\Pr(Z_2 = z_{2(n_k)}^{(k)} \mid Z_1 = z_{1(k)})\right\}^{-1} \\
& - m_{kn_k}\left\{\Pr(Z_2 = z_{2(n_k)}^{(k)} \mid Z_1 = z_{1(k)})\right\}^{-2}\Bigg] \\
= \sum_{k=1}^{N_1}\Bigg[& -\omega_{k0}\frac{z_{1(k)}}{\lambda_1^2} - \omega_{k1}\frac{z_{1(k)}}{\lambda_1}\left\{1-\frac{(z_{1(k)}-1)}{\lambda_1}\right\} - \frac{\omega_{k1}^2}{m_{k1}}\left(\frac{z_{1(k)}}{\lambda_1}\right)^2 \\
& -\omega_{kn_k}\left(\frac{z_{1(k)}}{\lambda_1}-1\right) - \frac{\omega_{kn_k}^2}{m_{kn_k}}\Bigg].
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial\lambda_2^2}\,\ell(\lambda_1,\lambda_2) = \sum_{k=1}^{N_1}\Bigg[& -\left(\sum_{j=2}^{n_k-1} m_{kj}\frac{z_{2(j)}^{(k)}}{\lambda_2^2}\right) - m_{k1}\left\{\frac{z_{2(1)}^{(k)}}{\lambda_2}-1\right\}\left\{\Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(1)}^{(k)})\right\}^{-1} \\
& - m_{k1}\left\{\Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(1)}^{(k)})\right\}^{-2} \\
& - m_{kn_k}\frac{z_{2(n_k)}^{(k)}}{\lambda_2}\left\{1-\frac{(z_{2(n_k)}^{(k)}-1)}{\lambda_2}\right\}\left\{\Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(n_k)}^{(k)})\right\}^{-1} \\
& - m_{kn_k}\left(\frac{z_{2(n_k)}^{(k)}}{\lambda_2}\right)^2\left\{\Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(n_k)}^{(k)})\right\}^{-2}\Bigg] \\
= \sum_{k=1}^{N_1}\Bigg[& -\left(\sum_{j=2}^{n_k-1} \upsilon_{kj}\frac{z_{2(j)}^{(k)}}{\lambda_2^2}\right) - \upsilon_{k1}\left\{\frac{z_{2(1)}^{(k)}}{\lambda_2}-1\right\} - \frac{\upsilon_{k1}^2}{m_{k1}} \\
& - \upsilon_{kn_k}\frac{z_{2(n_k)}^{(k)}}{\lambda_2}\left\{1-\frac{(z_{2(n_k)}^{(k)}-1)}{\lambda_2}\right\} - \frac{\upsilon_{kn_k}^2}{m_{kn_k}}\left(\frac{z_{2(n_k)}^{(k)}}{\lambda_2}\right)^2\Bigg],
\end{aligned}$$

and

$$\frac{\partial^2}{\partial\lambda_1\partial\lambda_2}\,\ell(\lambda_1,\lambda_2) = \sum_{k=1}^{N_1}\Bigg[m_{k1}\frac{z_{1(k)}}{\lambda_1}\left\{\Pr(Z_2 = z_{2(1)}^{(k)} \mid Z_1 = z_{1(k)})\Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2(1)}^{(k)})\right\}^{-1}$$
$$+m_{kn_k}\frac{z_{2n_k}^{(k)}}{\lambda_2}\left\{\Pr(Z_2 = z_{2(n_k)}^{(k)} \mid Z_1 = z_{1(k)})\Pr(Z_1 = z_{1(k)} \mid Z_2 = z_{2n_k}^{(k)})\right\}^{-1}\Bigg]$$
$$= \sum_{k=1}^{N_1}\left[\left(\frac{\omega_{k1}\upsilon_{k1}}{m_{k1}}\right)\left(\frac{z_{1(k)}}{\lambda_1}\right) + \left(\frac{\omega_{kn_k}\upsilon_{kn_k}}{m_{kn_k}}\right)\left(\frac{z_{2n_k}^{(k)}}{\lambda_2}\right)\right].$$

Maximum likelihood theory ensures that the MLE $\hat{\Lambda} = (\hat{\lambda}_1, \hat{\lambda}_2)^\top$ is consistent and asymptotically efficient. Specifically, as $n \to \infty$,

$$\sqrt{n}\,(\hat{\Lambda} - \Lambda) \rightsquigarrow \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1}),$$

where $\mathcal{I}$ denotes the Fisher information given by

$$\mathcal{I} = -\mathrm{E}\left\{\frac{\partial^2}{\partial\Lambda\partial\Lambda^\top}\ \ln\Pr(Z_1, Z_2)\right\}.$$

For practical purposes, this can be estimated by the observed Information matrix, viz.

$$-\frac{1}{n}\left\{\frac{\partial^2}{\partial\Lambda\partial\Lambda^\top}\,\ell(\lambda_1,\lambda_2)\right\}.$$

There are several variants in the statement of the usual regularity conditions under which such statements hold. For example, Serfling (1980) lists on p. 144 the regularity conditions as follows. Let Θ be an open interval in $\mathbb{R}$ and let $f(x;\theta)$ denote either a probability density or probability mass function with distribution function F_θ belonging to the family $\mathcal{F} = \{F_\theta : \theta \in \Theta\}$. The following regularity conditions on $\mathcal{F}$ ensure consistency, asymptotic Normality and efficiency of maximum likelihood estimators.

(1) For each $\theta \in \Theta$, the following derivatives exist for all $x \in \mathbb{R}$:

$$\frac{\partial}{\partial\theta}\ \ln f(x;\theta), \qquad \frac{\partial^2}{\partial\theta^2}\ \ln f(x;\theta), \qquad \frac{\partial^3}{\partial\theta^3}\ \ln f(x;\theta).$$

(2) For each $\theta_0 \in \Theta$, there exist functions $g(x)$, $h(x)$ and $H(x)$ (posibly depending on θ_0) such that for θ in a neighbourhood of $N(\theta_0)$ the following relations hold, for all $x \in \mathbb{R}$

$$\left|\frac{\partial}{\partial\theta}\,f(x;\theta)\right| \leqslant g(x), \qquad \left|\frac{\partial^2}{\partial\theta^2}\,f(x;\theta)\right| \leqslant h(x), \qquad \left|\frac{\partial^3}{\partial\theta^3}\,f(x;\theta)\right| \leqslant H(x),$$

and

$$\int g(x)dx < \infty, \quad \int h(x)dx < \infty, \quad \mathrm{E}_\theta\{H(X)\} < \infty.$$

(3) For each $\theta \in \Theta$,

$$0 < \mathrm{E}_\theta\left\{\left(\frac{\partial}{\partial\theta}\ln f(X;\theta)\right)^2\right\} < \infty.$$

Casella and Berger (2002) specify the regularity conditions in a slightly different manner (see p. 516):

(1) We observe $X_1, \ldots, X_n$, where $X_i \sim f(x;\theta)$ are iid.

(2) The parameter is identifiable; i.e., if $\theta \neq \theta'$, then $f(x;\theta) \neq f(x;\theta')$.

(3) The densities $f(x;\theta)$ have common support and $f(x;\theta)$ is differentiable in θ.

(4) The parameter space Ω contains an open set ω of which the true parameter value θ_0 is an interior point.

(5) For every $x \in \mathcal{X}$, $f(x;\theta)$ is three time differentiable with respect to θ, the third derivative is continuous in θ, and $\int f(x;\theta)dx$ can be differentiated three times under the integral sign.

(6) For any $\theta_0 \in \Omega$, there exists a positive number c and a function $M(x)$ (both of which may depend on θ_0) such that for all $x \in \mathcal{X}$ and $\theta_0 - c < \theta < \theta_0 + c$,

$$\left|\frac{\partial^3}{\partial\theta^3}\ln f(x;\theta)\right| \leqslant M(x),$$

with $\mathrm{E}_{\theta_0}\{M(X)\} < \infty$.

Note that the while the above statements are made in the univariate setting, the multivariate extension follows directly.

The log-likelihood of the comonotonic Poisson pair (Z_1, Z_2) involves evaluations of the univariate Poisson probability mass function and cumulative distribution function. Accordingly, the usual regularity conditions necessary for maximum likelihood theory will hold.

3.5.1 Simulation results

A set of simulations was carried out in order to further examine the effects of comonotonicity in estimating the marginal parameters in the case of Poisson distributed observations. Specifically, for varying values of the marginal parameters (λ_1, λ_2) and sample size n, maximum likelihood estimation was implemented both marginally and jointly. As previously mentioned, the marginal maximum likelihood, or IFM, estimates in this setting are simply the marginal sample means $(\bar{Z}_1, \bar{Z}_2)$. The joint maximum likelihood estimates $(\hat{\lambda}_1, \hat{\lambda}_2)$ were obtained using the `optim` function in R. In particular, the Nelder–Mead method was used, with starting values set equal to the true parameter values for simplicity. In order to avoid difficulties in the numerical estimation, a reparametrization of the log-likelihood was considered by optimizing over the parameters $(\alpha_1, \alpha_2) = (\ln(\lambda_1), \ln(\lambda_2))$.

In the simulation study, the value of λ_1 was varied in $\{1, 2, 5, 10\}$ while λ_2 was held fixed at 5 throughout. This was done as the problem is symmetric in the marginal parameters and so varying both marginal rates is redundant. The sample size n was varied from 5, 10, 100 to 1000. For each possible combination of λ_1 and n, 500 replications were performed. The results for estimating λ_1 are summarized in Figure 3.5. In each plot, the light blue boxplot depicts the marginal estimate $\bar{Z}_1$ while the pink boxplot represents the full maximum likelihood estimate $\hat{\lambda}_1$. The boxplots for the two analyses are shown side by side for each distinct value of n. The solid red line denotes the true value of λ_1. The figures suggest that both estimators behave similarly, as might be expected since both $\bar{Z}_1$ and $\hat{\lambda}_1$ are consistent estimators of λ_1.

As was illustrated in Example 3.6, there were several cases where evaluating the log-likelihood at the marginal parameter estimates $(\bar{Z}_1, \bar{Z}_2)$ returned a value of $-\infty$. Specifically, in the 16 distinct scenarios that arose by considering $\lambda_1 \in \{1, 2, 5, 10\}$ and $n \in \{1, 5, 100, 1000\}$, this particular phenomenon occurred in as few as 0% and at most 6.8% of the 500 replications. Note that the only instance where this did not occur in the simulation study was when the marginal parameters coincided so that $Z_1 = Z_2$ almost surely.

3.6 Counter-monotonic random pairs

The concept of counter-monotonicity parallels that of comonotonicity. As was previously discussed, a random pair (X_1, X_2) with marginal distribution functions F_1 and F_2, respectively, is said to be counter-monotonic if the following relation holds:

$$(X_1, X_2) =_d (F_1^{-1}(U), F_2^{-1}(1 - U)),$$

where U is a $\mathcal{U}(0, 1)$ random variable.

For continuous margins with strictly increasing cumulative distribution functions, counter-monotonicity then implies a perfect *negative* dependence wherein X_1 is a decreasing function of X_2. This follows immediately from the fact that while X_1 is a monotone increasing function of U, X_2 is a monotone increasing function of $1-U$. Consider the case where both margins are members of the same location-scale family with standard distribution function F. In particular, let X_1 have location and scale parameters μ_1 and σ_1, respectively, and let X_2 be analogously parametrized by μ_2 and σ_2. Then, one has that for any $(x_1, x_2) \in \mathbb{R}^2$,

$$F_1(x_1) = F\left(\frac{x_1 - \mu_1}{\sigma_1}\right), \quad F_2(x_2) = F\left(\frac{x_2 - \mu_2}{\sigma_2}\right).$$

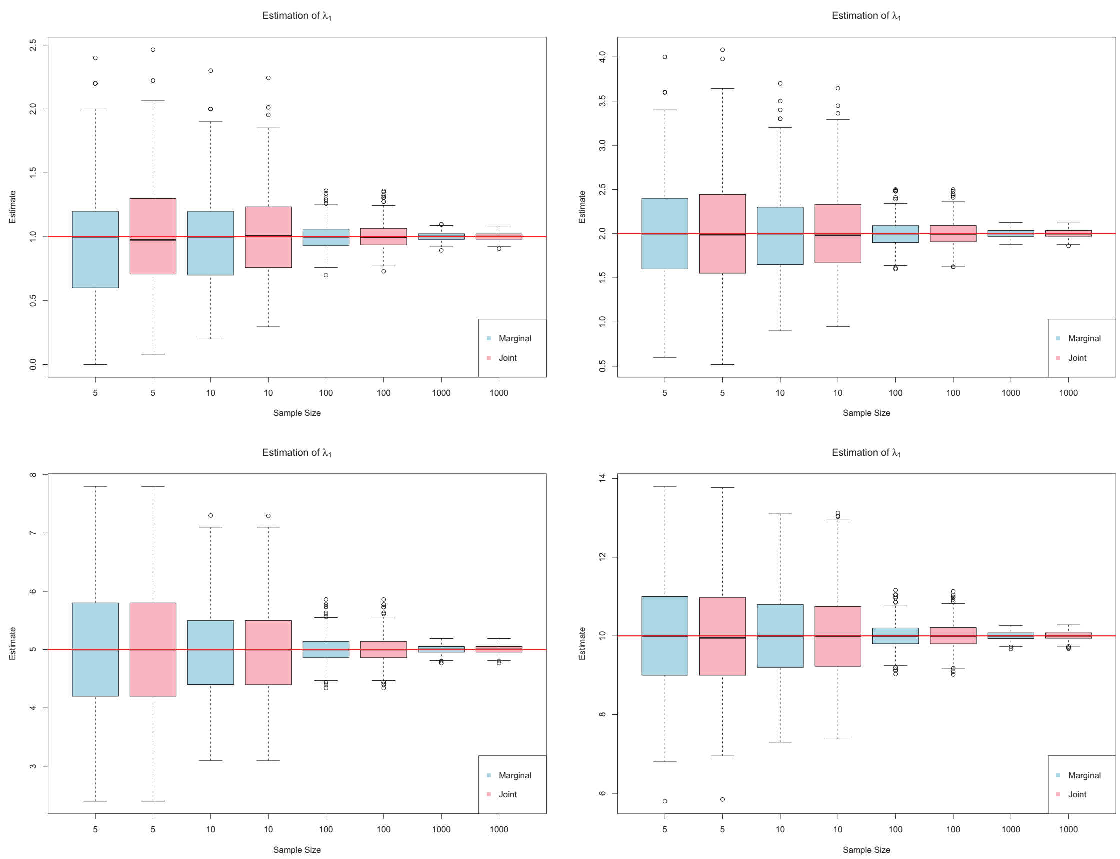

Figure 3.5: Simulation study results: The top left plot shows the results for $\lambda_1 = 1$, the top right for $\lambda_1 = 2$, bottom left for $\lambda_1 = 5$ and bottom right for $\lambda_1 = 10$. The x-axis shows the sample size n, which took on values 5, 10, 100 and 1000.

From this it follows that

$$X_1 = F_1^{-1}(U) \quad \Leftrightarrow \quad X_1 = \mu_1 + \sigma_1 F^{-1}(U)$$

and

$$X_2 = F_2^{-1}(1-U) \quad \Leftrightarrow \quad X_2 = \mu_2 + \sigma_2 F^{-1}(1-U).$$

For F strictly increasing, F^{-1} will also be strictly increasing so that X_1 is a decreasing function of X_2. The form of the cumulative distribution function will then dictate the structure of the functional dependence implied by counter-monotonicity in the case of continuous margins. Consider the case where X_1 and X_2 are both marginally Exponential with respective means β_1 and β_2.

Counter-monotonicity then yields the following:

$$X_1 = F_1^{-1}(U) = -\beta_1 \ln(1-U) \quad \Leftrightarrow \quad U = 1 - \exp\left(-X_1/\beta_1\right),$$
$$X_2 = F_2^{-1}(1-U) = -\beta_2 \ln(U) \quad \Leftrightarrow \quad U = \exp\left(-X_2/\beta_2\right).$$

Thus, the dependence between X_1 and X_2 has the functional form

$$X_1 = -\beta_1 \ln\left\{1 - \exp\left(-X_2/\beta_2\right)\right\}.$$

The treatment of counter-monotonic random pairs will be left brief as many of the properties and issues are analogous to that of comonotonic pairs. For more details on the concept of counter-monotonicity in the bivariate Poisson setting, see Pfeifer and Nešlehová (2004).

4

Proposed Bivariate Poisson Model

4.1 Introduction

As detailed in Chapter 2, the classical bivariate Poisson model introduces dependence via a common shock variable. By construction, this model can only characterize dependence in the restrictive range $[0, \min(\lambda_1, \lambda_2)/\sqrt{\lambda_1\lambda_2}]$, where λ_1 and λ_2 denote the marginal means. However, as mentioned by Griffiths et al. (1979), an arbitrary pair of Poisson random variables $X_1 \sim \mathcal{P}(\lambda_1)$ and $X_2 \sim \mathcal{P}(\lambda_2)$ can have correlation falling anywhere in the interval $[\rho_{\min}(\lambda_1, \lambda_2), \rho_{\max}(\lambda_1, \lambda_2)]$. In particular, the authors show that these minimum and maximum correlation values, $\rho_{\min}(\lambda_1, \lambda_2)$ and $\rho_{\max}(\lambda_1, \lambda_2)$, result respectively from the upper and lower Fréchet–Hoeffding boundary distributions and are given by

$$\rho_{\min}(\lambda_1, \lambda_2) = \frac{1}{\sqrt{\lambda_1\lambda_2}}\left[-\lambda_1\lambda_2 - \sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} \min\{0, G_{\lambda_1}(i) + G_{\lambda_2}(j) - 1\}\right], \tag{4.1}$$

$$\rho_{\max}(\lambda_1, \lambda_2) = \frac{1}{\sqrt{\lambda_1\lambda_2}}\left[-\lambda_1\lambda_2 + \sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}} \min\{\bar{G}_{\lambda_1}(i), \bar{G}_{\lambda_2}(j)\}\right]. \tag{4.2}$$

Recall that the classical bivariate Poisson model is constructed via the trivariate reduction technique wherein

$$X_1 = Y_1 + Z, \quad X_2 = Y_2 + Z \tag{4.3}$$

for independent random variables $Y_1 \sim \mathcal{P}(\lambda_1 - \xi)$, $Y_2 \sim \mathcal{P}(\lambda_2 - \xi)$ and $Z \sim \mathcal{P}(\xi)$. For a random pair (X_1, X_2) generated in this manner, it follows that $\mathrm{corr}(X_1, X_2) = \xi/\sqrt{\lambda_1\lambda_2}$. As was previously pointed out, only in the trivial case where $\lambda_1 = \lambda_2$ will the classical bivariate Poisson model span the full range of possible correlation. Indeed when $\lambda_1 = \lambda_2$, the correlation will range from 0 to 1. Figure 4.1 depicts the difference between the maximum possible correlation in the classical bivariate Poisson model and the upper bound derived by Griffiths et al. (1979), i.e., the difference $\rho_{\max}(\lambda_1, \lambda_2) - \min(\lambda_1, \lambda_2)/\sqrt{\lambda_1\lambda_2}$. The graph suggests that the difference tends to be

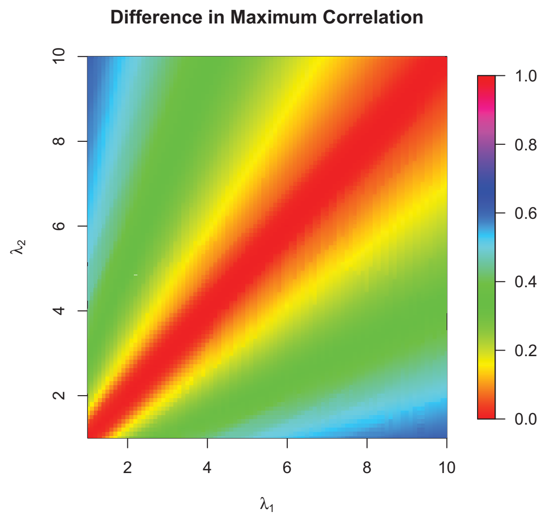

Figure 4.1: Difference between the maximum possible correlation implied by the classical bivariate Poisson model and the upper bound $\rho_{\max}$.

greatest when one of the marginal means is small while the other is large, i.e., when either $\lambda_1 \ll \lambda_2$ or $\lambda_1 \gg \lambda_2$. In contrast, the difference decreases when the difference between λ_1 and λ_2 is small.

In order to address the shortcomings of the classical model, an alternative bivariate Poisson model is proposed wherein dependence is induced through a comonotonic shock. The proposed model allows for greater flexibility in the implied correlation structure and spans the full range of possible positive correlation in $[0, \rho_{\max}(\lambda_1, \lambda_2)]$. The remainder of this chapter is dedicated to exploring the proposed model. The model construction will be introduced followed by its distributional properties. Various approaches to parameter estimation will also be examined along with simulations and a data illustration.

4.2 The proposed model

The random pair (X_1, X_2) is said to follow a *comonotonic shock* bivariate Poisson model if the margins can be written as

$$X_1 = Y_1 + Z_1, \quad X_2 = Y_2 + Z_2 \tag{4.4}$$

for

$$Y_1 \sim \mathcal{P}\{(1-\theta)\lambda_1\}, \quad Y_2 \sim \mathcal{P}\{(1-\theta)\lambda_2\}, \quad (Z_1, Z_2) \sim \mathcal{M}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\},$$

where Y_1 and Y_2 are mutually independent random variables which are also independent of the comonotonic pair (Z_1, Z_2). The formulation in (4.4) generates a pair of correlated Poisson random variables with marginal rates $\Lambda = (\lambda_1, \lambda_2) \in (0, \infty)^2$ and dependence parameter $\theta \in [0, 1]$. The notation $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$ will be used to denote the proposed model. In terms of mutually independent standard uniform random variables V_1, V_2 and U, a pair $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$ can be generated from

$$Y_1 = G^{-1}_{(1-\theta)\lambda_1}(V_1), \quad Y_2 = G^{-1}_{(1-\theta)\lambda_2}(V_2), \quad Z_1 = G^{-1}_{\theta\lambda_1}(U), \quad Z_2 = G^{-1}_{\theta\lambda_2}(U)$$

and setting $(X_1, X_2) = (Y_1 + Z_1, Y_2 + Z_2)$.

The comonotonic shock (Z_1, Z_2) induces the dependence while the strength of the association is regulated by θ. In fact, it will be shown that as θ increases, the strength of dependence increases accordingly. As such, the proposed comonotonic shock model spans the full range of possible dependence for a pair of positively associated Poisson random variables, i.e., $[0, \rho_{\max}]$. In particular, when $\theta = 0$ it follows that $(Z_1, Z_2) \equiv (0, 0)$ and thus the proposed model reduces to a pair of independent Poisson random variables with correlation $\rho = 0$. On the other hand, the limiting case $\theta = 1$ yields a comonotonic pair such that $\rho = \rho_{\max}$. This property will be formally presented in Lemma 4.1.

Note that in the case where $\lambda_1 = \lambda_2 = \lambda$, the $\mathcal{BP}(\Lambda, \theta)$ model coincides with the classical common shock model with $\theta = \xi/\lambda$. In this setting, the upper bound $\rho_{\max}$ is reached when $\theta = 1$ or equivalently when $\xi = \lambda$ so that $X_1 = X_2$ almost surely. However, when the marginal rates differ, the proposed model provides a more flexible framework for characterizing positively correlated Poisson random variables as the implied correlation can extend beyond $[0, \min(\lambda_1, \lambda_2)/\sqrt{\lambda_1 \lambda_2}]$.

Let $f_{\Lambda,\theta}$ and $F_{\Lambda,\theta}$ denote the corresponding probability mass function and cumulative distribution function in the proposed $\mathcal{BP}(\Lambda, \theta)$ family. Further write the probability mass function for the comonotonic pair $(Z_1, Z_2) \sim \mathcal{M}\left\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_1)\right\}$ as

$$\begin{aligned} c_{\Lambda,\theta}(z_1, z_2) &= \Pr(Z_1 = z_1, Z_2 = z_2) \\ &= \left[\min\{G_{\theta\lambda_1}(z_1), G_{\theta\lambda_2}(z_2)\} - \max\{G_{\theta\lambda_1}(z_1 - 1), G_{\theta\lambda_2}(z_2 - 1)\}\right]_+ , \end{aligned}$$

where $[x]_+ = x\mathbf{1}(x > 0)$. Then for $(x_1, x_2) \in \mathbb{N}^2$ it is straightforward to show that

$$f_{\Lambda,\theta}(x_1, x_2) = \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2)$$

and

$$\begin{aligned} F_{\Lambda,\theta}(x_1, x_2) &= \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) \min\left\{G_{\theta\lambda_1}(z_1), G_{\theta\lambda_2}(z_2)\right\} \\ &= \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} G_{(1-\theta)\lambda_1}(x_1 - z_1) G_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2). \end{aligned}$$

4.2.1 PQD ordering

The proposed $\mathcal{BP}(\Lambda, \theta)$ family of distributions is ordered in the Positive Quadrant Dependent (PQD) ordering, as regulated by the dependence parameter θ. Following Lehmann (1966), the notion of positive quadrant dependence in the two-dimensional case is defined as follows:

Definition 4.1 ***Positive quadrant dependence***
Suppose a random vector (X_1, X_2) has distribution function F, where X_1 and X_2 have respective marginal distributions given by F_1 and F_2. Then the random pair (X_1, X_2) is said to be PQD if and only if, for all x_1 and x_2,

$$F(x_1, x_2) \geqslant F_1(x_1)F_2(x_2).$$

Note that in the above equation, the right-hand side is the joint distribution under independence. Thus, PQD implies that the probability of the event $\{X_1 \leqslant x_1, X_2 \leqslant x_2\}$ is always greater under the joint distribution F as compared to the independence distribution. The notion of PQD ordering extends this definition to allow for comparisons between distribution functions with the same fixed margins. The following definition introduces the concept of PQD ordering; see, e.g., Eq. (9.A.3) in Shaked and Shanthikumar (2007).

Definition 4.2 ***PQD ordering***
Suppose the random pair (X_1, X_2) has distribution function F and corresponding survival function $\bar{F}$, where $\bar{F}(x_1, x_2) = \Pr(X_1 > x_1, X_2 > x_2)$ for all $x_1, x_2 \in \mathbb{R}$. Further suppose (Y_1, Y_2) has distribution function G and survival function $\bar{G}$, where both F and G have common marginal distributions given by F_1 and F_2. (That is, both X_1 and Y_1 have distribution function F_1 while both X_2 and Y_2 have distribution function F_2). If for all x_1 and x_2

$$F(x_1, x_2) \leqslant G(x_1, x_2), \tag{4.5}$$

then (X_1, X_2) is said to be smaller than (Y_1, Y_2) in the PQD order. This will be denoted as $(X_1, X_2) \prec_{PQD} (Y_1, Y_2)$, or analogously in terms of the distribution functions as $F \prec_{PQD} G$.

Note that in the bivariate setting, Eq. (4.5) is equivalent to $\bar{F}(x_1, x_2) \leqslant \bar{G}(x_1, x_2)$. This can be shown as follows:

$$\begin{aligned}
F(x_1, x_2) \leqslant G(x_1, x_2) &\Rightarrow\ 1 - F(x_1, x_2) \geqslant 1 - G(x_1, x_2) \\
&\Rightarrow\ 1 - F(x_1, x_2) - \bar{F}_1(x_1) - \bar{F}_2(x_2) \geqslant 1 - G(x_1, x_2) - \bar{F}_1(x_1) - \bar{F}_2(x_2) \\
&\Rightarrow\ -\bar{F}(x_1, x_2) \geqslant -\bar{G}(x_1, x_2) \\
&\Rightarrow\ \bar{F}(x_1, x_2) \leqslant \bar{G}(x_1, x_2),
\end{aligned}$$

since

$$\begin{aligned} 1 - F(x_1, x_2) - \bar{F}_1(x_1) - \bar{F}_2(x_2) &= \Pr(X_1 > x_1 \cup X_2 > x_2) - \Pr(X_1 > x_1) - \Pr(X_2 > x_2) \\ &= - \Pr(X_1 > x_1 \cap X_2 > x_2) \\ &= -\bar{F}(x_1, x_2). \end{aligned}$$

The notion of PQD ordering is also related to that of orthant orders. Following the definitions given in Section 6.G.1 of Shaked and Shanthikumar (2007), suppose (X_1, X_2) has distribution function F with corresponding survival function $\bar{F}$. Further suppose (Y_1, Y_2) has distribution function G and survival function $\bar{G}$. If for all (x_1, x_2) the inequality $\bar{F}(x_1, x_2) \leqslant \bar{G}(x_1, x_2)$ holds, then (X_1, X_2) is said to be smaller than (Y_1, Y_2) in the upper orthant order, which will be denoted as $(X_1, X_2) \prec_{UO} (Y_1, Y_2)$. Analogously, if $F(x_1, x_2) \leqslant G(x_1, x_2)$ for all (x_1, x_2), then (X_1, X_2) is said to be smaller than (Y_1, Y_2) in the lower orthant order, denote as $(X_1, X_2) \prec_{LO} (Y_1, Y_2)$. Clearly, for two distribution functions F and G with the same marginal distributions F_1 and F_2, the following equivalences ensue:

$$\begin{aligned} (X_1, X_2) \prec_{PQD} (Y_1, Y_2) &\Leftrightarrow (X_1, X_2) \prec_{UO} (Y_1, Y_2), \\ (X_1, X_2) \prec_{PQD} (Y_1, Y_2) &\Leftrightarrow (X_1, X_2) \prec_{LO} (Y_1, Y_2). \end{aligned}$$

Note that while the concept of PQD ordering requires the distribution functions F and G to have the same fixed marginals, the notions of UO and LO do not have this restriction.

The concept of PQD ordering has several convenient properties. For example, it follows immediately from Definition 4.2 that any bivariate distribution function F satisfies

$$F_L \prec_{PQD} F \prec_{PQD} F_U \tag{4.6}$$

where F_L and F_U denote the Fréchet–Hoeffding boundary distributions, defined for all $x_1, x_2 \in \mathbb{R}$, by

$$F_L(x_1, x_2) = \max\{0, F_1(x_1) + F_2(x_2) - 1\}, \quad F_U(x_1, x_2) = \min\{F_1(x_1), F_2(x_2)\}.$$

See Eq. (9.A.6) of Shaked and Shanthikumar (2007). In addition, the notion of PQD order is closed under convolutions and monotone increasing transformations. This property is detailed in Theorem (9.A.1) of Shaked and Shanthikumar (2007) and will be provided (without proof) in the following theorem for completeness.

Theorem 4.1 *Suppose that the random pairs (X_1, X_2), (Y_1, Y_2), (U_1, U_2) and (V_1, V_2) satisfy*

$$(X_1, X_2) \prec_{PQD} (Y_1, Y_2), \quad (U_1, U_2) \prec_{PQD} (V_1, V_2).$$

Further suppose that (X_1, X_2) and (U_1, U_2) are independent and that (Y_1, Y_2) and (V_1, V_2) are also independent. Then for all increasing functions ϕ and ψ, the following holds

$$\{\phi(X_1, U_1), \psi(X_2, U_2)\} \prec_{PQD} \{\phi(Y_1, V_1), \psi(Y_2, V_2)\}.$$

In the case where ϕ and ψ are taken to be sums, Theorem 4.1 implies that

$$(X_1 + U_1, X_2 + U_2) \prec_{PQD} (Y_1 + V_1, Y_2 + V_2). \tag{4.7}$$

There are several other closure properties that hold for PQD ordering; see Theorem (9.A.2) of Shaked and Shanthikumar (2007) for more details.

In terms of the comonotonic shock bivariate Poisson model, θ dictates the strength of the dependence as reflected by the PQD ordering. In particular, holding the marginal parameters λ_1 and λ_2 fixed, it can be shown that for any $x_1, x_2 \in \mathbb{N}$, if $\theta < \theta'$ then $F_{\Lambda,\theta}(x_1, x_2) \leqslant F_{\Lambda,\theta'}(x_1, x_2)$. This will be formally stated in the following lemma.

Lemma 4.1 ***PQD ordering in the $\mathcal{BP}(\Lambda, \theta)$ family***
Let $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$ and $(X_1', X_2') \sim \mathcal{BP}(\Lambda, \theta')$. Then $\theta < \theta' \Rightarrow (X_1, X_2) \prec_{PQD} (X_1', X_2')$.

Proof. Fix λ_1, λ_2 and $\theta < \theta'$. By definition of the proposed $\mathcal{BP}$ family, (X_1, X_2) can be expressed as

$$X_1 = Y_1 + Z_1, \quad X_2 = Y_2 + Z_2,$$

for $Y_1 \sim \mathcal{P}\{(1-\theta)\lambda_1\}$ independent of $Y_2 \sim \mathcal{P}\{(1-\theta)\lambda_2\}$ and (Y_1, Y_2) independent of the comonotonic pair $(Z_1, Z_2) \sim \mathcal{M}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\}$. In the same way, (X_1', X_2') has the representation

$$X_1' = Y_1' + Z_1', \quad X_2' = Y_2' + Z_2'$$

for $Y_1' \sim \mathcal{P}\{(1-\theta')\lambda_1\}$ independent of $Y_2' \sim \mathcal{P}\{(1-\theta')\lambda_2\}$ and (Y_1', Y_2') independent of $(Z_1', Z_2') \sim \mathcal{M}\{\mathcal{P}(\theta'\lambda_1), \mathcal{P}(\theta'\lambda_2)\}$. Since the univariate Poisson distribution is infinitely divisible, the independent components (Y_1, Y_2) can be rewritten as

$$Y_1 = T_1 + S_1, \quad Y_2 = T_2 + S_2,$$

where $T_1 \sim \mathcal{P}\{(1-\theta')\lambda_1\}$, $T_2 \sim \mathcal{P}\{(1-\theta')\lambda_2\}$, $S_1 \sim \mathcal{P}\{(\theta'-\theta)\lambda_1\}$, $S_2 \sim \mathcal{P}\{(\theta'-\theta)\lambda_2\}$ are mutually independent. It is clear that $(T_1, T_2) \prec_{PQD} (Y_1', Y_2')$ as both are independent pairs with the same marginal distributions. Both the pairs $(S_1 + Z_1, S_2 + Z_2)$ and (Z_1', Z_2') have marginal distributions that are $\mathcal{P}(\theta'\lambda_1)$ and $\mathcal{P}(\theta'\lambda_2)$ for the first and second components, respectively. The pair (Z_1', Z_2') is comonotonic and thus by (4.6) it follows that

$$(S_1 + Z_1, S_2 + Z_2) \prec_{PQD} (Z_1', Z_2').$$

Finally, note that the random vectors $(T_1 + S_1 + Z_1, T_2 + S_2 + Z_2)$ and $(Y_1' + Z_1', Y_2' + Z_2')$ have the same margins, specifically $\mathcal{P}(\lambda_1)$ and $\mathcal{P}(\lambda_2)$. Then, by (4.7), it follows that

$$(T_1 + S_1 + Z_1, T_2 + S_2 + Z_2) \prec_{PQD} (Y_1' + Z_1', Y_2' + Z_2'),$$

which is the desired result since by construction $(T_1 + S_1 + Z_1, T_2 + S_2 + Z_2) \sim \mathcal{BP}(\Lambda, \theta)$ and $(Y_1' + Z_1', Y_2' + Z_2') \sim \mathcal{BP}(\Lambda, \theta')$. □

4.2.2 Moments and measures of dependence

As shown in Shaked and Shanthikumar (2007), the PQD ordering correspondingly implies an ordering in measures of dependence. In particular, if $(X_1, X_2) \prec_{PQD} (Y_1, Y_2)$, then both the covariance and correlation inherit this ordering:

$$\text{cov}(X_1, X_2) \leqslant \text{cov}(Y_1, Y_2), \quad \text{corr}(X_1, X_2) \leqslant \text{corr}(Y_1, Y_2).$$

More generally, if $(X_1, X_2) \prec_{PQD} (Y_1, Y_2)$, then $\rho(X_1, X_2) \leqslant \rho(Y_1, Y_2)$ for any concordance measure ρ in the sense of Scarsini (1984). As noted in Shaked and Shanthikumar (2007), this includes Spearman's rho and Kendall's tau.

Applying these concepts to the proposed model, it follows that for $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$, the implied correlation $\rho_\theta = \text{corr}(X_1, X_2)$ will be an increasing function of θ. Since the independence model results when $\theta = 0$, it follows that $\rho_\theta \geqslant 0$. As will be shown subsequently, as θ increases to 1, the model reaches the upper Fréchet–Hoeffding bound and the resulting correlation coincides with the upper bound given in (4.2). As a first step towards establishing this result, the probability generating function for the $\mathcal{BP}$ family will be derived.

Proposition 4.2 *Suppose $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$. Then the pair (X_1, X_2) has probability generating function given by*

$$\text{E}(s_1^{X_1} s_2^{X_2}) = \exp\left\{(1-\theta)\lambda_1(s_1 - 1) + (1-\theta)\lambda_2(s_2 - 1)\right\} \varrho_{\Lambda,\theta}(s_1, s_2),$$

where

$$\varrho_{\Lambda,\theta}(s_1, s_2) = 1 + (s_1 - 1)\sum_{i=0}^{\infty} s_1^i \bar{G}_{\theta\lambda_1}(i) + (s_2 - 1)\sum_{j=0}^{\infty} s_2^j \bar{G}_{\theta\lambda_2}(j)$$
$$+ (s_1 - 1)(s_2 - 1)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \min\{\bar{G}_{\theta\lambda_1}(i), \bar{G}_{\theta\lambda_2}(j)\}.$$

Proof. The stochastic representation for (X_1, X_2) given in (4.4) allows to break down the probability generating function into three components stemming from the mutually independent random variables Y_1, Y_2 and the comonotonic pair (Z_1, Z_2), viz.

$$\mathrm{E}(s_1^{X_1} s_2^{X_2}) = \mathrm{E}(s_1^{Y_1})\mathrm{E}(s_2^{Y_2})\mathrm{E}(s_1^{Z_1} s_2^{Z_2}).$$

Since Y_1 and Y_2 are univariate Poisson random variables with respective means $(1-\theta)\lambda_1$ and $(1-\theta)\lambda_2$, it follows that

$$\mathrm{E}(s_1^{Y_1}) = \exp\{(1-\theta)\lambda_1 s_1\}, \quad \mathrm{E}(s_2^{Y_2}) = \exp\{(1-\theta)\lambda_2 s_2\}.$$

Thus, it only remains to establish the probability generating function for the comonotonic pair $(Z_1, Z_2) \sim \mathcal{M}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\}$. This function is denoted by $\varrho_{\Lambda,\theta}$. An alternative representation of the probability mass function for the comonotonic pair (Z_1, Z_2) can be derived in terms of the comonotonic survival function. In particular, for any $i, j \in \mathbb{N}$,

$$\Pr(Z_1 = i, Z_2 = j) = \Pr(Z_1 \geqslant i, Z_2 \geqslant j) + \Pr(Z_1 \geqslant i+1, Z_2 \geqslant j+1)$$
$$- \Pr(Z_1 \geqslant i+1, Z_2 \geqslant j) - \Pr(Z_1 \geqslant i, Z_2 \geqslant j+1).$$

Since comonotonicity implies that $(Z_1, Z_2) = \left\{G_{\theta\lambda_1}^{-1}(U), G_{\theta\lambda_2}^{-1}(U)\right\}$ for a standard uniform random variable $U \sim \mathcal{U}(0,1)$, it follows that

$$\begin{aligned}\Pr(Z_1 \geqslant i, Z_2 \geqslant j) &= \Pr\{U > G_{\theta\lambda_1}(i-1), U > G_{\theta\lambda_2}(j-1)\} \\ &= 1 - \max\{G_{\theta\lambda_1}(i-1), G_{\theta\lambda_2}(j-1)\} \\ &= \min\{\bar{G}_{\theta\lambda_1}(i-1), \bar{G}_{\theta\lambda_2}(j-1)\}.\end{aligned}$$

Accordingly, for any $i, j \in \mathbb{N}$ one has

$$\Pr(Z_1 = i, Z_2 = j) = \min\{\bar{G}_{\theta\lambda_1}(i-1), \bar{G}_{\theta\lambda_2}(j-1)\} + \min\{\bar{G}_{\theta\lambda_1}(i), \bar{G}_{\theta\lambda_2}(j)\}$$
$$- \min\{\bar{G}_{\theta\lambda_1}(i), \bar{G}_{\theta\lambda_2}(j-1)\} - \min\{\bar{G}_{\theta\lambda_1}(i-1), \bar{G}_{\theta\lambda_2}(j)\}.$$

It then follows that

$$
\begin{aligned}
\varrho_{\Lambda,\theta}(s_1,s_2) &= \mathrm{E}(s_1^{Z_1}s_2^{Z_2}) \\
&= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \big[\min\{\bar{G}_{\theta\lambda_1}(i-1),\bar{G}_{\theta\lambda_2}(j-1)\} - \min\{\bar{G}_{\theta\lambda_1}(i),\bar{G}_{\theta\lambda_2}(j-1)\} \\
&\qquad\qquad - \min\{\bar{G}_{\theta\lambda_1}(i-1),\bar{G}_{\theta\lambda_2}(j)\} + \min\{\bar{G}_{\theta\lambda_1}(i),\bar{G}_{\theta\lambda_2}(j)\}\big] \\
&= 1 + \sum_{j=0}^{\infty} s_2^j \bar{G}_{\theta\lambda_2}(j-1) + \sum_{i=1}^{\infty} s_1^i \bar{G}_{\theta\lambda_1}(i-1) \\
&\qquad + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} s_1^i s_2^j \min\{\bar{G}_{\theta\lambda_1}(i-1),\bar{G}_{\theta\lambda_2}(j-1)\} \\
&\qquad - \sum_{i=0}^{\infty} s_1^i \bar{G}_{\theta\lambda_1}(i) - \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} s_1^i s_2^j \min\{\bar{G}_{\theta\lambda_1}(i),\bar{G}_{\theta\lambda_2}(j-1)\} \\
&\qquad - \sum_{j=0}^{\infty} s_2^j \bar{G}_{\theta\lambda_2}(j) - \sum_{i=1}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \min\{\bar{G}_{\theta\lambda_1}(i-1),\bar{G}_{\theta\lambda_2}(j)\} \\
&\qquad + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \min\{\bar{G}_{\theta\lambda_1}(i),\bar{G}_{\theta\lambda_2}(j)\} \\
&= 1 + \sum_{i=0}^{\infty}(s_1^{i+1}-s_1^i)\bar{G}_{\theta\lambda_1}(i) + \sum_{j=0}^{\infty}(s_2^{j+1}-s_2^j)\bar{G}_{\theta\lambda_2}(j) \\
&\qquad + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(s_1^i s_2^j - s_1^{i+1}s_2^j - s_1^i s_2^{j+1} + s_1{}^j + 1s_2^{j+1}) \min\{\bar{G}_{\theta\lambda_1}(i),\bar{G}_{\theta\lambda_2}(j)\} \\
&= 1 + (s_1-1)\sum_{i=0}^{\infty} s_1^i \bar{G}_{\theta\lambda_1}(i) + (s_2-1)\sum_{j=0}^{\infty} s_2^j \bar{G}_{\theta\lambda_2}(j) \\
&\qquad + (s_1-1)(s_2-1)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \min\{\bar{G}_{\theta\lambda_1}(i),\bar{G}_{\theta\lambda_2}(j)\}.
\end{aligned}
$$

This completes the proof. □

The probability generating function can then be used to derive the moments of (X_1, X_2). The intrinsic dependence structure of the stochastic representation in (4.4) implies that

$$\mathrm{cov}(X_1, X_2) = \mathrm{cov}(Z_1, Z_2).$$

Working with $\varrho_{\Lambda,\theta}$, it is straightforward to verify that

$$\mathrm{E}(Z_1 Z_2) = \frac{\partial^2}{\partial s_1 \partial s_2}\varrho_{\Lambda,\theta}(s_1,s_2)\Big|_{s_1=s_2=1} = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\min\{\bar{G}_{\theta\lambda_1}(i),\bar{G}_{\theta\lambda_2}(j)\}.$$

Thus, the random pair $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$ has correlation

$$\rho_\theta = \text{corr}(X_1, X_2) = \frac{1}{\sqrt{\lambda_1 \lambda_2}} \left[-\theta^2 \lambda_1 \lambda_2 + \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \min\{\bar{G}_{\theta\lambda_1}(i), \bar{G}_{\theta\lambda_2}(j)\} \right]. \qquad (4.8)$$

It is clear that setting $\theta = 0$ yields $\rho_\theta = 0$. At the other end of the spectrum, $\theta = 1$ results in $\rho_\theta = \rho_{\max}$ so that the implied correlation in the $\mathcal{BP}(\Lambda, \theta)$ model reaches the upper bound given in (4.2). Moreover, Lemma 4.1 ensures that for fixed margins (λ_1, λ_2), the correlation ρ_θ is an increasing function of θ. Accordingly, the $\mathcal{BP}(\Lambda, \theta)$ family can accommodate the full range of possible positive dependence, with correlation varying from 0 to $\rho_{\max}$. Therefore, the proposed comonotonic shock bivariate Poisson model provides a fully flexible stochastic representation for positively correlated Poisson random variables.

4.2.3 Recurrence relations

In the classical bivariate Poisson model, the construction is based on a single common shock variable. Accordingly, establishing a set of recurrence relations is simplified by the conditional independence that ensues when conditioning on the common shock Z. In the proposed comonotonic shock model, however, the dependence structure is more complex and thus renders the computation of recurrence relations less straightforward.

Recall that the univariate Poisson distribution has the simple recurrence relation:

$$g_\lambda(x) = \frac{\lambda}{x} g_\lambda(x-1), \quad x \in \{1, 2, \ldots\}.$$

Since the classical common shock bivariate Poisson model formulation relies on convolutions of independent Poisson random variables, this recurrence relation remains relevant. As shown in, e.g., Holgate (1964) and Kawamura (1985) and reviewed in Chapter 2, for the classic bivariate Poisson distribution with probability mass function given, for all $x_1, x_2 \in \mathbb{N}$, by

$$h_{\Lambda,\xi}(x_1, x_2) = e^{-(\lambda_1 + \lambda_2 - \xi)} \sum_{z=0}^{\min\{x_1, x_2\}} \frac{(\lambda_1 - \xi)^{x_1 - z} (\lambda_2 - \xi)^{x_2 - z} \xi^z}{(x_1 - z)!(x_2 - z)! z!}$$

the following recurrence relations hold: for all $x_1, x_2 \in \{1, 2, \ldots\}$

$$x_1 h_{\Lambda,\xi}(x_1, x_2) = (\lambda_1 - \xi) h_{\Lambda,\xi}(x_1 - 1, x_2) + \xi h_{\Lambda,\xi}(x_1 - 1, x_2 - 1),$$
$$x_2 h_{\Lambda,\xi}(x_1, x_2) = (\lambda_2 - \xi) h_{\Lambda,\xi}(x_1, x_2 - 1) + \xi h_{\Lambda,\xi}(x_1 - 1, x_2 - 1).$$

Recall that in the proposed $\mathcal{BP}(\Lambda, \theta)$ model, the joint probability mass function has the form

$$f_{\Lambda,\theta}(x_1, x_2) = \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2)$$

for all $x_1, x_2 \in \mathbb{N}$. Working with the recurrence relations for the univariate Poisson PMF, the following relations hold:

$$\begin{aligned}
f_{\Lambda,\theta}(x_1 - 1, x_2) &= \sum_{z_1=0}^{x_1-1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - 1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2) \\
&= \sum_{z_1=0}^{x_1-1} \sum_{z_2=0}^{x_2} \frac{x_1 - z_1}{(1-\theta)\lambda_1} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2) \\
&= \{(1-\theta)\lambda_1\}^{-1} \sum_{z_1=0}^{x_1-1} \sum_{z_2=0}^{x_2} (x_1 - z_1) g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2) \\
&= \{(1-\theta)\lambda_1\}^{-1} \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} (x_1 - z_1) g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2) \\
&= \{(1-\theta)\lambda_1\}^{-1} f_{\Lambda,\theta}(x_1, x_2) \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} (x_1 - z_1) p_{\Lambda,\theta}(z_1, z_1, x_1, x_1) / f_{\Lambda,\theta}(x_1, x_2) \\
&= \{(1-\theta)\lambda_1\}^{-1} f_{\Lambda,\theta}(x_1, x_2) \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} (x_1 - z_1) p_{\Lambda,\theta}(z_1, z_1 \mid x_1, x_1) \\
&= \{(1-\theta)\lambda_1\}^{-1} f_{\Lambda,\theta}(x_1, x_2) \mathrm{E}(X_1 - Z_1 \mid x_1, x_2) \\
&= \{(1-\theta)\lambda_1\}^{-1} \{x_1 - \mathrm{E}(Z_1 \mid x_1, x_2)\} f_{\Lambda,\theta}(x_1, x_2).
\end{aligned}$$

In the above equations, $p_{\Lambda,\theta}(z_1, z_2, x_1, x_2)$ is used to denote the joint probability

$$\Pr(Z_1 = z_1, Z_2 = z_2, X_1 = x_1, X_2 = x_2)$$

and $p_{\Lambda,\theta}(z_1, z_2 \mid x_1, x_2)$ denotes the conditional probability

$$\Pr(Z_1 = z_1, Z_2 = z_2 \mid X_1 = x_1, X_2 = x_2).$$

In the same way, it follows that

$$f_{\Lambda,\theta}(x_1, x_2 - 1) = \{(1-\theta)\lambda_2\}^{-1} \{x_2 - \mathrm{E}(Z_2 \mid x_1, x_2)\} f_{\Lambda,\theta}(x_1, x_2).$$

Similarly, the following recursion can be derived:

$$\begin{aligned}
f_{\Lambda,\theta}(x_1-1,x_2-1) &= \sum_{z_1=0}^{x_1-1}\sum_{z_2=0}^{x_2-1} g_{(1-\theta)\lambda_1}(x_1-1-z_1)g_{(1-\theta)\lambda_2}(x_2-1-z_2)c_{\Lambda,\theta}(z_1,z_2)\\
&= \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}\sum_{z_1=0}^{x_1}\sum_{z_2=0}^{x_2}(x_1-z_1)(x_2-z_2)\\
&\qquad g_{(1-\theta)\lambda_1}(x_1-z_1)g_{(1-\theta)\lambda_2)}(x_2-z_2)c_{\Lambda,\theta}(z_1,z_2)\\
&= \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}f_{\Lambda,\theta}(x_1,x_2)\\
&\qquad \sum_{z_1=0}^{x_1}\sum_{z_2=0}^{x_2}(x_1-z_1)(x_2-z_2)p_{\Lambda,\theta}(z_1,z_2,x_1,x_1)/f_{\Lambda,\theta}(x_1,x_2)
\end{aligned}$$

and hence

$$\begin{aligned}
f_{\Lambda,\theta}(x_1-1,x_2-1) &= \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}f_{\Lambda,\theta}(x_1,x_2)\\
&\qquad \sum_{z_1=0}^{x_1}\sum_{z_2=0}^{x_2}(x_1-z_1)(x_2-z_2)p_{\Lambda,\theta}(z_1,z_2\mid x_1,x_1)\\
&= \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}f_{\Lambda,\theta}(x_1,x_2)\mathrm{E}\{(X_1-Z_1)(X_2-Z_2)\mid x_1,x_2\}\\
&= \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}f_{\Lambda,\theta}(x_1,x_2)\times\\
&\qquad \Big[x_1x_2-x_2\mathrm{E}(Z_1\mid x_1,x_2)-x_1\mathrm{E}(Z_2\mid x_1,x_2)+\mathrm{E}(Z_1Z_2\mid x_1,x_2)\Big]\\
&= \left\{(1-\theta)^2\lambda_1,\lambda_2\right\}^{-1}\Big[x_2\left\{x_1-\mathrm{E}(Z_1\mid x_1,x_2)\right\}+\\
&\qquad x_1\left\{x_2-\mathrm{E}(Z_2\mid x_1,x_2)\right\}-\left\{x_1x_2-\mathrm{E}(Z_1Z_2\mid x_1,x_2)\right\}\Big]f_{\Lambda,\theta}(x_1,x_2).
\end{aligned}$$

To recap, the following three recurrence relations hold in the proposed $\mathcal{BP}(\Lambda,\theta)$ family

$$f_{\Lambda,\theta}(x_1-1,x_2) = f_{\Lambda,\theta}(x_1,x_2)\{(1-\theta)\lambda_1\}^{-1}\left\{x_1-\mathrm{E}(Z_1\mid x_1,x_2)\right\}, \tag{4.9}$$

$$f_{\Lambda,\theta}(x_1,x_2-1) = f_{\Lambda,\theta}(x_1,x_2)\{(1-\theta)\lambda_2\}^{-1}\left\{x_2-\mathrm{E}(Z_2\mid x_1,x_2)\right\}, \tag{4.10}$$

$$\begin{aligned}
f_{\Lambda,\theta}(x_1-1,x_2-1) &= f_{\Lambda,\theta}(x_1,x_2)\{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}\big[x_2\left\{x_1-\mathrm{E}(Z_1\mid x_1,x_2)\right\}\\
&\qquad +x_1\left\{x_2-\mathrm{E}(Z_2\mid x_1,x_2)\right\}-\left\{x_1x_2-\mathrm{E}(Z_1Z_2\mid x_1,x_2)\right\}\big].
\end{aligned} \tag{4.11}$$

Note that these recursions also lead to several other relations. For example,

$$\begin{aligned}
&f_{\Lambda,\theta}(x_1-1,x_2)\times f_{\Lambda,\theta}(x_1,x_2-1)\\
&\quad= \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}\big\{x_1x_2-x_2\mathrm{E}(Z_1\mid x_1,x_2)-x_1\mathrm{E}(Z_2\mid x_1,x_2)\\
&\qquad\qquad +\mathrm{E}(Z_1\mid x_1,x_2)\mathrm{E}(Z_2\mid x_1,x_2)\big\}\left\{f_{\Lambda,\theta}(x_1,x_2)\right\}^2,
\end{aligned}$$

so

$$\begin{aligned} f_{\Lambda,\theta}(x_1-1,x_2)\times f_{\Lambda,\theta}(x_1,x_2-1) \\ = \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}\{f_{\Lambda,\theta}(x_1,x_2)\}^2\{x_1x_2 - x_2\mathrm{E}(Z_1\mid x_1,x_2) \\ - x_1\mathrm{E}(Z_2\mid x_1,x_2) + \mathrm{E}(Z_1Z_2\mid x_1,x_2) - \mathrm{cov}(Z_1,Z_2\mid x_1,x_2)\}, \end{aligned}$$

whence

$$\begin{aligned} f_{\Lambda,\theta}(x_1-1,x_2)\times f_{\Lambda,\theta}(x_1,x_2-1) \\ = f_{\Lambda,\theta}(x_1,x_2)\big[f_{\Lambda,\theta}(x_1-1,x_2-1) - \{(1-\theta)^2\lambda_1,\lambda_2\}^{-1}f_{\Lambda,\theta}(x_1,x_2)\mathrm{cov}(Z_1,Z_2\mid x_1,x_2)\big] \end{aligned}$$

and also

$$\begin{aligned} f_{\Lambda,\theta}(x_1-1,x_2-1) = x_1\{(1-\theta)\lambda_1\}^{-1}f_{\Lambda,\theta}(x_1,x_2-1) + x_2\{(1-\theta)\lambda_2\}^{-1}f_{\Lambda,\theta}(x_1-1,x_2) \\ - \{x_1x_2 - \mathrm{E}(Z_1,Z_2\mid x_1,x_2)\}\, f_{\Lambda,\theta}(x_1,x_2). \end{aligned}$$

4.2.4 Convolutions in the $\mathcal{BP}$ family

In the proposed bivariate Poisson model, the dependence structure is dictated by the comonotonic shock. As will be shown, comonotonicity is not retained under convolutions and thus the $\mathcal{BP}(\Lambda,\theta)$ family is not closed under convolution.

Suppose that (X_{11},X_{12}) and (X_{21},X_{22}) are independent random vectors from the proposed $\mathcal{BP}$ model with respective marginal rates given by $\Lambda_1=(\lambda_{11},\lambda_{12})$ and $\Lambda_2=(\lambda_{21},\lambda_{22})$ and a common dependence parameter θ. In terms of independent $\mathcal{U}(0,1)$ random variables $V_{11},V_{12},U_1,V_{21},V_{22},U_2$, these pairs can be expressed as

$$\begin{aligned} X_{11} &= Y_{11}+Z_{11} = G^{-1}_{(1-\theta)\lambda_{11}}(V_{11}) + G^{-1}_{\theta\lambda_{11}}(U_1),\\ X_{12} &= Y_{12}+Z_{12} = G^{-1}_{(1-\theta)\lambda_{12}}(V_{12}) + G^{-1}_{\theta\lambda_{12}}(U_1),\\ X_{21} &= Y_{21}+Z_{21} = G^{-1}_{(1-\theta)\lambda_{21}}(V_{21}) + G^{-1}_{\theta\lambda_{21}}(U_2),\\ X_{22} &= Y_{22}+Z_{22} = G^{-1}_{(1-\theta)\lambda_{22}}(V_{22}) + G^{-1}_{\theta\lambda_{22}}(U_2). \end{aligned}$$

Define the random sum (X_1,X_2) by

$$(X_1,X_2) = (X_{11}+X_{21},X_{12}+X_{22}).$$

It is then of interest whether the distribution of (X_1,X_2) falls within the $\mathcal{BP}$ family, i.e., whether the proposed bivariate Poisson distribution is closed under convolution.

Clearly, both X_1 and X_2 have marginal univariate Poisson distributions with respective means $\lambda_1 = \lambda_{11} + \lambda_{21}$ and $\lambda_2 = \lambda_{12} + \lambda_{22}$ as both components are the sum of independent Poisson random variables. Setting $Y_1 = Y_{11} + Y_{21}$ and $Y_2 = Y_{12} + Y_{22}$, it is also clear that $Y_1 \sim \mathcal{P}\{(1-\theta)\lambda_1\}$ is independent of $Y_2 \sim \mathcal{P}\{(1-\theta)\lambda_2\}$. Let (Z_1, Z_2) denote the the sum $(Z_{11} + Z_{21}, Z_{12} + Z_{22})$ so that marginally both Z_1 and Z_2 have Poisson distribution with rates $\theta\lambda_1$ and $\theta\lambda_2$, respectively. By design, Y_1 and Y_2 are independent of (Z_1, Z_2). Then, if (Z_1, Z_2) are comonotonic the sum $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$, where $\Lambda = (\lambda_1, \lambda_2)$. Thus, the proposed class of distributions $\mathcal{BP}(\Lambda, \theta)$ is closed under convolution if comonotonic Poisson random variables are closed under convolution.

Recall that (Z_1, Z_2) are comonotonic if and only if the random pair can be written in terms of a common underlying uniform random variable. In this set-up, comonotonicity then implies that $(Z_1, Z_2) = \{G^{-1}_{\theta\lambda_1}(U), G^{-1}_{\theta\lambda_2}(U)\}$ for some $U \sim \mathcal{U}(0,1)$. Consider a specific example: set $\Lambda_1 = (1,2)$, $\Lambda_2 = (3,8)$, $\theta = 0.5$. In this scenario $Z_1 \sim \mathcal{P}(2)$ and $Z_2 \sim \mathcal{P}(5)$. Take two independent realizations from a standard uniform distributions as $u_1 = 0.5$, $u_2 = 0.8$. It then follows that

$$\begin{aligned} z_{11} &= G^{-1}_{\theta\lambda_{11}}(u_1) = G^{-1}_{0.5}(0.5) = 0, \quad & z_{12} &= G^{-1}_{\theta\lambda_{12}}(u_1) = G^{-1}_{1}(0.5) = 1, \\ z_{21} &= G^{-1}_{\theta\lambda_{21}}(u_2) = G^{-1}_{1.5}(0.8) = 2, \quad & z_{22} &= G^{-1}_{\theta\lambda_{22}}(u_2) = G^{-1}_{4}(0.8) = 6, \end{aligned}$$

and so $(z_1, z_2) = (0,1) + (2,6) = (2,7)$. A value $z_1 = G^{-1}_2(u) = 2$ will result for any underlying value of $u \in (G_2(1), G_2(2)] = (0.41, 0.68]$. However, $z_2 = G^{-1}_5(u) = 7$ is generated by any $u \in (G_5(6), G_5(7)] = (0.76, 0.87]$. Since these two intervals do not overlap, there is no underlying $U \sim \mathcal{U}(0,1)$ which can generate a pair of comonotonic Poisson random variable equal to $(2,7)$ when the marginal means are respectively given by 2 and 5. Consequently, (X_1, X_2) cannot be expressed in terms of the $\mathcal{BP}$ model and the family is thus not closed under convolution.

Note that even sums of independent and identically distributed comonotonic pairs do not retain the property of comonotonicity. Consider the case where both (Z_{11}, Z_{21}) and (Z_{12}, Z_{22}) are i.i.d. $\mathcal{M}\{\mathcal{P}(1), \mathcal{P}(2)\}$. Suppose $u_1 = 0.7$ and $u_2 = 0.2$, thus generating

$$\begin{aligned} Z_{11} &= G^{-1}_1(0.7) = 1, \quad & Z_{21} &= G^{-1}_2(0.7) = 3, \\ Z_{12} &= G^{-1}_1(0.2) = 0, \quad & Z_{22} &= G^{-1}_2(0.2) = 1. \end{aligned}$$

Marginally, $Z_1 = Z_{11} + Z_{12} \sim \mathcal{P}(2)$ and $Z_2 = Z_{21} + Z_{22} \sim \mathcal{P}(4)$. The values of u_1 and u_2 yield $Z_1 = 1$ and $Z_2 = 4$. Writing $Z_1 = G^{-1}_2(U)$ implies that an observation of 1 can be generated from an underlying uniform variable in the interval $(G_2(0), G_2(1)] = (0.14, 0.41]$. In the same way, for $Z_2 = G^{-1}_4(U)$, observing a value of 4 corresponds to a Uniform variable in the interval $(G_4(3), G_4(4)] = (0.43, 0.63]$. Again, these intervals do not overlap and thus the pair (Z_1, Z_2) are not comonotonic.

4.3 Estimation

Let $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ denote a random sample from the proposed $\mathcal{BP}(\Lambda, \theta)$ model. It is then of interest to obtain estimates of the marginal parameters $\Lambda = (\lambda_1, \lambda_2) \in (0, \infty)^2$ and the dependence parameter $\theta \in [0, 1]$. This section will give an overview of various estimation methods, namely the method of moments (MM), maximum likelihood (ML) and inference functions for margins (IFM).

4.3.1 Method of moments

The method of moments approach to estimation relies on matching the theoretical moments implied by the assumed model to the empirical moments observed in the sample. In the case of the $\mathcal{BP}(\Lambda, \theta)$ family, method of moments estimation involves matching three moments as there are three parameters to estimate. Since the marginal parameters are in fact equal to the marginal means, the method of moments estimates for λ_1 and λ_2, denoted by $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ respectively, are given by

$$\tilde{\lambda}_1 = \bar{X}_1, \quad \tilde{\lambda}_2 = \bar{X}_2,$$

for$\bar{X}_k = (X_{1k} + \cdots + X_{nk})/n$ for each $k \in \{1, 2\}$. Note that these estimates coincide with maximum likelihood estimation when $\theta = 0$, i.e., in the case where the components X_1 and X_2 are independent. The Central Limit Theorem then ensures that, as $n \to \infty$,

$$\sqrt{n}\,(\tilde{\lambda}_1 - \lambda_1) \rightsquigarrow \mathcal{N}(0, \lambda_1), \quad \sqrt{n}\,(\tilde{\lambda}_2 - \lambda_2) \rightsquigarrow \mathcal{N}(0, \lambda_2),$$

where $\rightsquigarrow$ denotes convergence in law. Again, these asymptotic results are equivalent to those from maximum likelihood theory in the case of independence.

Estimation of the dependence parameter requires mixed moments. In particular, an estimate of θ can be obtained by matching the model covariance to the sample covariance. Let S_{12} denote the sample covariance given by

$$S_{12} = \frac{1}{n-1} \sum_{\ell=1}^{n} (X_{\ell 1} - \bar{X}_1)(X_{\ell 2} - \bar{X}_2).$$

As shown in Section 4.2.2, the covariance of $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$ is given by

$$\mathrm{cov}(X_1, X_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \min\{\bar{G}_{\theta\lambda_1}(i), \bar{G}_{\theta\lambda_2}(j)\} - \theta^2 \lambda_1 \lambda_2. \tag{4.12}$$

Let $m_{\lambda_1,\lambda_2}(\theta)$ denote the above covariance. The PQD ordering of the $\mathcal{BP}(\Lambda, \theta)$ family ensures that for fixed Λ, the covariance $m_{\lambda_1,\lambda_2}(\theta)$ is an increasing function of θ. Thus, there will be a unique value of θ, say $\tilde{\theta}$, for which $m_{\lambda_1,\lambda_2}(\theta) = S_{12}$, provided that $S_{12} \in [0, m_{\lambda_1,\lambda_2}(1)]$. Note that the latter condition ensures that the sample covariance falls within the range of permissible values for $m_{\lambda_1,\lambda_2}(\theta)$ in the $\mathcal{BP}(\Lambda,\theta)$ family. Indeed, $\theta = 0$ implies independence and thus $m_{\lambda_1,\lambda_2}(0) = 0$ while $\theta = 1$ results in perfect positive dependence such that $m_{\lambda_1,\lambda_2}(1) = \rho_{\max}(\lambda_1,\lambda_2)\sqrt{\lambda_1\lambda_2}$.

The MM estimator $\tilde{\theta}$ is then the unique value of θ such that

$$m_{\bar{X}_1,\bar{X}_2}(\theta) = S_{12},$$

for $S_{12} \in [0, \rho_{\max}(\bar{X}_1,\bar{X}_2)\sqrt{\bar{X}_1\bar{X}_2}]$. By convention, if $S_{12} > \rho_{\max}(\bar{X}_1,\bar{X}_2)\sqrt{\bar{X}_1\bar{X}_2}$, set $\tilde{\theta} = 1$. However if it is observed that $S_{12} < 0$, a model for negative dependence should be used instead; see Chapter 5.

The sample covariance S_{12} is a consistent estimator of $m_{\lambda_1,\lambda_2}(\theta)$. Moreover, as shown in, e.g., Theorem 8 of Ferguson (1996), as $n \to \infty$,

$$\sqrt{n}\,\{S_{12} - m_{\lambda_1,\lambda_2}(\theta)\} \rightsquigarrow \mathcal{N}[0, \sigma^2(\theta,\lambda_1,\lambda_2)],$$

where

$$\sigma^2(\theta,\lambda_1,\lambda_2) = \text{var}\,\{(X_1-\lambda_1)(X_2-\lambda_2)\}\,.$$

A straightforward application of the Delta Method then ensures that the MM estimator $\tilde{\theta}$ is also asymptotically Gaussian, as is shown in the following proposition.

Proposition 4.3 ***Asymptotic normality of the method of moments estimator***
Let $(X_{11},X_{12}),\ldots,(X_{n1},X_{n2})$ denote a random sample from the $\mathcal{BP}(\Lambda,\theta)$ family and let S_{12} denote the sample covariance given by $\left\{\sum_{\ell=1}^{n}(X_{\ell 1}-\bar{X}_1)(X_{\ell 2}-\bar{X}_2)\right\}/(n-1)$. The method of moments estimator is the unique solution of the equation

$$m_{\bar{X}_1,\bar{X}_2}(\theta) = S_{12}.$$

Moreover, for $S_{12} \in [0, \rho_{\max}(\bar{X}_1,\bar{X}_2)\sqrt{\bar{X}_1\bar{X}_2}]$, as $n \to \infty$,

$$\sqrt{n}\,(\tilde{\theta}-\theta) \rightsquigarrow \mathcal{N}[0, \{\gamma'(\sigma_{12})\}^2\,\sigma^2(\theta,\lambda_1,\lambda_2)],$$

where σ_{12} is used to denote the covariance $m_{\lambda_1,\lambda_2}(\theta)$ and γ denotes the inverse of the function $m_{\lambda_1,\lambda_2}(\theta)$, i.e., $\gamma : \theta \mapsto m^{-1}_{\lambda_1,\lambda_2}(\theta)$, with corresponding derivative γ'.

In general, computation of the asymptotic variance is tedious, although some simplifications can be made. First note that by definition,

$$\text{var}\{(X_1-\lambda_1)(X_2-\lambda_2)\} = \text{E}\{(X_1-\lambda_1)^2(X_2-\lambda_2)^2\} - \{m_{\lambda_1,\lambda_2}(\theta)\}^2\,.$$

By construction of the $\mathcal{BP}(\Lambda,\theta)$ model, for $k \in \{1,2\}$, $X_k = Y_k + Z_k$, where $Y_k \sim \mathcal{P}\{(1-\theta)\lambda_k\}$

is independent of $Z_k \sim \mathcal{P}(\theta\lambda_k)$. Then for $k \in \{1, 2\}$, write

$$X_k - \lambda_k = \mathring{Y}_k + \mathring{Z}_k,$$

where $\mathring{Y}_k$ and $\mathring{Z}_k$ represent the centred random variables, i.e.,

$$\mathring{Y}_k = Y_k - (1-\theta)\lambda_k, \quad \mathring{Z}_k = Z_k - \theta\lambda_k.$$

It can then be shown that

$$\begin{aligned}\mathrm{E}\{(X_1-\lambda_1)^2(X_2-\lambda_2)^2\} = \mathrm{E}(&\mathring{Y}_1^2\mathring{Y}_2^2 + 2\mathring{Y}_1^2\mathring{Y}_2\mathring{Z}_2 + \mathring{Y}_1^2\mathring{Z}_2^2 + 2\mathring{Y}_1\mathring{Z}_1\mathring{Y}_2^2 \\ &+ 4\mathring{Y}_1\mathring{Z}_1\mathring{Y}_2\mathring{Z}_2 + 2\mathring{Y}_1\mathring{Z}_1\mathring{Z}_2^2 + \mathring{Z}_1^2\mathring{Y}_2^2 + 2\mathring{Z}_1^2\mathring{Y}_2\mathring{Z}_2 + \mathring{Z}_1^2\mathring{Z}_2^2).\end{aligned}$$

Since the pairs $(\mathring{Y}_1, \mathring{Y}_2)$ and $(\mathring{Z}_1, \mathring{Z}_2)$ are independent, the right-hand side reduces to

$$\begin{aligned}&\mathrm{E}(\mathring{Y}_1^2)\mathrm{E}(\mathring{Y}_2^2) + 2\mathrm{E}(\mathring{Y}_1^2)\mathrm{E}(\mathring{Y}_2)\mathrm{E}(\mathring{Z}_2) + \mathrm{E}(\mathring{Y}_1^2)\mathrm{E}(\mathring{Z}_2^2) + 2\mathrm{E}(\mathring{Y}_1)\mathrm{E}(\mathring{Z}_1)\mathrm{E}(\mathring{Y}_2^2) \\ &\quad + 4\mathrm{E}(\mathring{Y}_1)\mathrm{E}(\mathring{Y}_2)\mathrm{E}(\mathring{Z}_1\mathring{Z}_2) + 2\mathrm{E}(\mathring{Y}_1)\mathrm{E}(\mathring{Z}_1\mathring{Z}_2^2) + \mathrm{E}(\mathring{Z}_1^2)\mathrm{E}(\mathring{Y}_2^2) + 2\mathrm{E}(\mathring{Y}_2)\mathrm{E}(\mathring{Z}_1^2\mathring{Z}_2) + \mathrm{E}(\mathring{Z}_1^2\mathring{Z}_2^2).\end{aligned}$$

Now by construction, the components of $(\mathring{Y}_1, \mathring{Y}_2)$ are themselves independent and $\mathrm{E}(\mathring{Y}_k) = \mathrm{E}(\mathring{Z}_k) = 0$ for $k \in \{1, 2\}$. The above expression thus becomes

$$\begin{aligned}&\mathrm{E}(\mathring{Y}_1^2)\mathrm{E}(\mathring{Y}_2^2) + \mathrm{E}(\mathring{Y}_1^2)\mathrm{E}(\mathring{Z}_2^2) + \mathrm{E}(\mathring{Z}_1^2)\mathrm{E}(\mathring{Y}_2^2) + \mathrm{E}(\mathring{Z}_1^2\mathring{Z}_2^2) \\ &\qquad = \mathrm{var}(Y_1)\mathrm{var}(Y_2) + \mathrm{var}(Y_1)\mathrm{var}(Z_2) + \mathrm{var}(Z_1)\mathrm{var}(Y_2) + \mathrm{E}\{(Z_1-\theta\lambda_1)^2(Z_2-\theta\lambda_2)^2\}\end{aligned}$$

in terms of the original variables Y_1, Y_2, Z_1, Z_2. Moreover, the last summand satisfies

$$\begin{aligned}\mathrm{E}\left\{(Z_1-\theta\lambda_1)^2(Z_2-\theta\lambda_2)^2\right\} &= \mathrm{var}\left\{(Z_1-\theta\lambda_1)(Z_2-\theta\lambda_2)\right\} + [\mathrm{E}\left\{(Z_1-\theta\lambda_1)(Z_2-\theta\lambda_2)\right\}^2] \\ &= \mathrm{var}\left\{(Z_1-\theta\lambda_1)(Z_2-\theta\lambda_2)\right\} + \left\{m_{\lambda_1,\lambda_2}(\theta)\right\}^2.\end{aligned}$$

Putting this all together, it follows that

$$\begin{aligned}\mathrm{var}\{(X_1-\lambda_1)(X_2-\lambda_2)\} = \mathrm{var}(Y_1)\mathrm{var}(Y_2) &+ \mathrm{var}(Y_1)\mathrm{var}(Z_2) + \mathrm{var}(Z_1)\mathrm{var}(Y_2) \\ &+ \mathrm{var}\left\{(Z_1-\theta\lambda_1)(Z_2-\theta\lambda_2)\right\},\end{aligned}$$

which, upon substitution, further reduces to

$$\mathrm{var}\{(X_1-\lambda_1)(X_2-\lambda_2)\} = (1-\theta)^2\lambda_1\lambda_2 + 2(1-\theta)\theta\lambda_1\lambda_2 + \mathrm{var}\left\{(Z_1-\theta\lambda_1)(Z_2-\theta\lambda_2)\right\}. \quad (4.13)$$

When $\theta = 0$, X_1 and X_2 are independent and $\sigma^2(0, \lambda_1, \lambda_2)$ reduces to $\lambda_1\lambda_2$. In the special case where $\lambda_1 = \lambda_2 = \lambda$, the comonotonic shock reduces to a common shock variable, i.e., $Z_1 = Z_2$

almost surely. Letting Z denote this common variable, one has that

$$\text{var}\{(Z_1 - \theta\lambda_1)(Z_2 - \theta\lambda_2)\} = \text{var}\{(Z - \theta\lambda)^2\} = \theta\lambda(1 + 3\theta\lambda) - \theta^2\lambda^2 = \theta\lambda(1 + 2\theta\lambda).$$

It then follows from (4.13) that

$$\text{var}\{(X_1 - \lambda_1)(X_2 - \lambda_2)\} = (1-\theta)^2\lambda^2 + 2(1-\theta)\theta\lambda^2 + \theta\lambda(1 + 2\theta\lambda) = \lambda(\lambda + \theta + \theta^2\lambda).$$

Moreover, in this case $m_{\lambda,\lambda}(\theta) = \theta\lambda$ and hence $\gamma'(\sigma_{12}) = 1/\lambda$. The asymptotic variance of the moment-based estimator $\tilde{\theta}$ thus reduces to $\lambda(\lambda + \theta + \theta^2\lambda)/\lambda^2 = (\lambda + \theta + \theta^2\lambda)/\lambda$.

4.3.2 Maximum likelihood estimation

Recall that the joint probability mass function in the proposed $\mathcal{BP}(\Lambda, \theta)$ family is given, for any $(x_1, x_2) \in \mathbb{N}^2$, by

$$f_{\Lambda,\theta}(x_1, x_2) = \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2).$$

It then follows that for observations $\mathbf{x} = \{(x_{11}, x_{12}), \ldots, (x_{n1}, x_{n2})\}$, the likelihood is given by

$$L(\Lambda, \theta; \mathbf{x}) = \prod_{i=1}^{n} f_{\Lambda,\theta}(x_{i1}, x_{i2})$$

with corresponding log-likelihood

$$\begin{aligned} \ell(\Lambda, \theta; \mathbf{x}) &= \sum_{i=1}^{n} \ln f_{\Lambda,\theta}(x_{i1}, x_{i2}) \\ &= \sum_{i=1}^{n} \ln \left\{ \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c_{\Lambda,\theta}(z_1, z_2) \right\}, \end{aligned} \tag{4.14}$$

where, as previously defined,

$$c_{\Lambda,\theta}(z_1, z_2) = \left[\min\{G_{\theta\lambda_1}(z_1), G_{\theta\lambda_2}(z_2)\} - \max\{G_{\theta\lambda_1}(z_1 - 1), G_{\theta\lambda_2}(z_2 - 1)\}\right]_+ .$$

Maximum likelihood estimates (MLEs) are then determined as

$$\underset{\lambda_1,\lambda_2,\theta}{\arg\max}\ \ell(\Lambda, \theta; \mathbf{x}),$$

subject to the constraints $\Lambda \in (0,\infty)^2$ and $\theta \in [0,1]$. If they exist, the MLEs are found by setting the score equations equal to zero, i.e., solving for the values $(\lambda_1, \lambda_2, \theta)$ such that

$$\left(\frac{\partial}{\partial \lambda_1}\,\ell(\Lambda,\theta), \frac{\partial}{\partial \lambda_2}\,\ell(\Lambda,\theta), \frac{\partial}{\partial \theta}\,\ell(\Lambda,\theta)\right) = \mathbf{0}^\top.$$

The score equations in the $\mathcal{BP}$ model are obtained by differentiating the log-likelihood as given in (4.14) with respect to $\Psi = (\lambda_1, \lambda_2, \theta)$. As a first step, it is straightforward to obtain the following derivatives involving the univariate Poisson probability mass functions:

$$\begin{aligned}
\frac{\partial}{\partial \lambda_1}\, g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) &= -(1-\theta) g_{(1-\theta)\lambda_2}(x_2 - z_2) \\
&\qquad \times \{g_{(1-\theta)\lambda_1}(x_1 - z_1) - g_{(1-\theta)\lambda_1}(x_1 - 1 - z_1)\}, \\
\frac{\partial}{\partial \lambda_2}\, g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) &= -(1-\theta) g_{(1-\theta)\lambda_1}(x_1 - z_1) \\
&\qquad \times \{g_{(1-\theta)\lambda_2}(x_2 - z_2) - g_{(1-\theta)\lambda_2}(x_2 - 1 - z_2)\}, \\
\frac{\partial}{\partial \theta}\, g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) &= (\lambda_1 + \lambda_2) g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) \\
&\qquad - \lambda_1 g_{(1-\theta)\lambda_1}(x_1 - 1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2), \\
&\qquad - \lambda_2 g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - 1 - z_2).
\end{aligned}$$

From the above equations, one can proceed to differentiate the log-likelihood. Beginning with the marginal parameter λ_1, one finds that $\partial \ell(\Lambda,\theta) \partial \lambda_1$ equals

$$\begin{aligned}
&\sum_{i=1}^{n} \{f_{\Lambda,\theta}(x_{i1}, x_{i2})\}^{-1} \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \left[\left\{\frac{\partial}{\partial \lambda_1}\, g_{(1-\theta)\lambda_1}(x_{i1} - z_1)\right\} g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c_{\Lambda,\theta}(z_1, z_2)\right. \\
&\qquad\qquad \left. + \; g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) \left\{\frac{\partial}{\partial \lambda_1}\, c_{\Lambda,\theta}(z_1, z_2)\right\}\right] \\
&= \sum_{i=1}^{n} \{f_{\Lambda,\theta}(x_{i1}, x_{i2})\}^{-1} \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \Big[-(1-\theta) g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c_{\Lambda,\theta}(z_1, z_2) \\
&\qquad\qquad + (1-\theta) g_{(1-\theta)\lambda_1}(x_{i1} - 1 - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c_{\Lambda,\theta}(z_1, z_2) \\
&\qquad\qquad + g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c_{\Lambda,\theta}(z_1, z_2) \left\{\frac{\partial}{\partial \lambda_1}\, \ln c_{\Lambda,\theta}(z_1, z_2)\right\} \Big].
\end{aligned}$$

and hence $\partial \ell(\Lambda, \theta) \partial \lambda_1$ equals

$$\sum_{i=1}^{n} \{f_{\Lambda,\theta}(x_{i1}, x_{i2})\}^{-1} \left[-(1-\theta) f_{\Lambda,\theta}(x_{i1}, x_{i2}) + (1-\theta) f_{\Lambda,\theta}(x_{i1}-1, x_{i2})\right]$$

$$+ \sum_{i=1}^{n} \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \left\{ \frac{\partial}{\partial \lambda_1} \ln c_{\Lambda,\theta}(z_1, z_2) \right\} p_{\Lambda,\theta}(z_1, z_2, x_{i1}, x_{i2}) / f_{\Lambda,\theta}(x_{i1}, x_{i2})$$

$$= -n(1-\theta) + \frac{1}{\lambda_1} \sum_{i=1}^{n} \{x_{i1} - \mathrm{E}(Z_1 \mid x_{i1}, x_{i2})\} + \sum_{i=1}^{n} \mathrm{E} \left\{ \frac{\partial}{\partial \lambda_1} \ln c_{\Lambda,\theta}(Z_1, Z_1) \middle| x_{i1}, x_{i2} \right\},$$

where the last step follows from the recurrence relations, given in Eqs. (4.9)–(4.11), since

$$f_{\Lambda,\theta}(x_1 - 1, x_2) / f_{\Lambda,\theta}(x_1, x_2) = \{(1-\theta)\lambda_1\}^{-1} \{x_1 - \mathrm{E}(Z_1 \mid x_1, x_2)\}.$$

With further manipulations, the score equation can be expressed as

$$\frac{\partial}{\partial \lambda_1} \ell(\Lambda, \theta) = \frac{n}{\lambda_1}(\bar{x}_1 - \lambda_1) - \frac{n}{\lambda_1} \{\bar{q}_1(\Lambda, \theta) - \theta \lambda_1\} + \sum_{i=1}^{n} \mathrm{E} \left\{ \frac{\partial}{\partial \lambda_1} \ln c_{\Lambda,\theta}(Z_1, Z_1) \middle| x_{i1}, x_{i2} \right\}, \quad (4.15)$$

where

$$\bar{q}_1(\Lambda, \theta) = \frac{1}{n} \sum_{i=1}^{n} q_{i1}(\Lambda, \theta) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}(Z_1 \mid x_{i1}, x_{i2}).$$

Analogously, let $\bar{q}_2(\Lambda, \theta) = \sum_{i=1}^{n} \mathrm{E}(Z_2 \mid x_{i1}, x_{i2})/n$. In the same way it can be shown that

$$\frac{\partial}{\partial \lambda_2} \ell(\Lambda, \theta) = \frac{n}{\lambda_2}(\bar{x}_2 - \lambda_2) - \frac{n}{\lambda_2} \{\bar{q}_2(\Lambda, \theta) - \theta \lambda_2\} + \sum_{i=1}^{n} \mathrm{E} \left\{ \frac{\partial}{\partial \lambda_2} \ln c_{\Lambda,\theta}(Z_1, Z_1) \middle| x_{i1}, x_{i2} \right\}. \quad (4.16)$$

The score equation involving the dependence parameter θ can be derived as follows. First, we see that $\partial \ell(\Lambda, \theta) \partial \theta$ equals

$$\sum_{i=1}^{n} \{f_{\Lambda,\theta}(x_{i1}, x_{i2})\}^{-1} \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \left[\left\{ \frac{\partial}{\partial \theta} g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) \right\} c_{\Lambda,\theta}(z_1, z_2) \right.$$

$$\left. + \; g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) \left\{ \frac{\partial}{\partial \theta} c_{\Lambda,\theta}(z_1, z_2) \right\} \right]$$

$$= \sum_{i=1}^{n} \{f_{\Lambda,\theta}(x_{i1}, x_{i2})\}^{-1} \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \Big[(\lambda_1 + \lambda_2) g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c_{\Lambda,\theta}(z_1, z_2)$$

$$- \lambda_1 g_{(1-\theta)\lambda_1}(x_{i1} - 1 - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c_{\Lambda,\theta}(z_1, z_2)$$

$$- \lambda_2 g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - 1 - z_2) c_{\Lambda,\theta}(z_1, z_2)$$

$$+ g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c_{\Lambda,\theta}(z_1, z_2) \left\{ \frac{\partial}{\partial \theta} \ln c_{\Lambda,\theta}(z_1, z_2) \right\} \Big],$$

which reduces to

$$n(\lambda_1+\lambda_2)-\sum_{i=1}^n \{f_{\Lambda,\theta}(x_{i1},x_{i2})\}^{-1}\left[-\lambda_1 f_{\Lambda,\theta}(x_{i1}-1,x_{i2})-\lambda_2 f_{\Lambda,\theta}(x_{i1},x_{i2}-1)\right]$$
$$+\sum_{i=1}^n\sum_{z_1=0}^{x_{i1}}\sum_{z_2=0}^{x_{i2}}\left\{\frac{\partial}{\partial\theta}\ \ln c_{\Lambda,\theta}(z_1,z_2)\right\} p_{\Lambda,\theta}(z_1,z_2,x_{i1},x_{i2})/f_{\Lambda,\theta}(x_{i1},x_{i2}),$$

where, again, the recurrence relations in Eqs. (4.9) through (4.11) were used to establish

$$f_{\Lambda,\theta}(x_1-1,x_2)/f_{\Lambda,\theta}(x_1,x_2)=\{(1-\theta)\lambda_1\}^{-1}\{x_1-\mathrm{E}(Z_1\mid x_1,x_2)\},$$
$$f_{\Lambda,\theta}(x_1,x_2-1)/f_{\Lambda,\theta}(x_1,x_2)=\{(1-\theta)\lambda_2\}^{-1}\{x_1-\mathrm{E}(Z_2\mid x_1,x_2)\}.$$

Further simplifications then yield

$$\frac{\partial}{\partial\theta}\ \ell(\Lambda,\theta)=n(\lambda_1+\lambda_2)-\frac{n}{(1-\theta)}\{\bar{x}_1+\bar{x}_2-\bar{q}_1(\Lambda,\theta)-\bar{q}_2(\Lambda,\theta)\}$$
$$+\sum_{i=1}^n \mathrm{E}\left\{\frac{\partial}{\partial\theta}\ \ln c_{\Lambda,\theta}(z_1,z_2)\middle| x_{i1},x_{i2}\right\}. \quad (4.17)$$

Equivalently, (4.17) can be written as

$$\frac{\partial}{\partial\theta}\ \ell(\Lambda,\theta)=-\frac{n}{(1-\theta)}(\bar{x}_1-\lambda_1)-\frac{n}{(1-\theta)}(\bar{x}_2-\lambda_2)+\frac{n}{(1-\theta)}\ \{\bar{q}_1(\Lambda,\theta)-\theta\lambda_1\}$$
$$+\frac{n}{(1-\theta)}\{\bar{q}_2(\Lambda,\theta)-\theta\lambda_2\}+\sum_{i=1}^n \mathrm{E}\left\{\frac{\partial}{\partial\theta}\ \ln c_{\Lambda,\theta}(z_1,z_2)\middle| x_{i1},x_{i2}\right\}. \quad (4.18)$$

Standard results ensure that

$$\mathrm{E}\left\{\frac{\partial}{\partial\lambda_1}\ \ell(\Lambda,\theta)\right\}=\mathrm{E}\left\{\frac{\partial}{\partial\lambda_2}\ \ell(\Lambda,\theta)\right\}=\mathrm{E}\left\{\frac{\partial}{\partial\theta}\ \ell(\Lambda,\theta)\right\}=0.$$

Moreover, both $\bar{X}_1-\lambda_1$ and $\bar{X}_2-\lambda_2$ have expectation equal to zero and for $k\in\{1,2\}$

$$\mathrm{E}\{\bar{q}_k(\Lambda,\theta)-\theta\lambda_k\}=\mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}(Z_k\mid X_{i1},X_{i2})\right\}-\theta\lambda_k=\frac{1}{n}\sum_{i=1}^n \mathrm{E}\{\mathrm{E}(Z_k\mid X_{i1},X_{i2})\}-\theta\lambda_k$$
$$=\frac{1}{n}\sum_{i=1}^n \mathrm{E}(Z_k)-\theta\lambda_k=\theta\lambda_k-\theta\lambda_k=0.$$

From Eqs. (4.15), (4.16) and (4.18), it follows that

$$\mathrm{E}\left[\sum_{i=1}^{n}\mathrm{E}\left\{\frac{\partial}{\partial\lambda_1}\ \ln c_{\Lambda,\theta}(Z_1,Z_2)\Big|X_{i1},X_{i2}\right\}\right]=0,$$
$$\mathrm{E}\left[\sum_{i=1}^{n}\mathrm{E}\left\{\frac{\partial}{\partial\lambda_2}\ \ln c_{\Lambda,\theta}(Z_1,Z_2)\Big|X_{i1},X_{i2}\right\}\right]=0,$$
$$\mathrm{E}\left[\sum_{i=1}^{n}\mathrm{E}\left\{\frac{\partial}{\partial\theta}\ \ln c_{\Lambda,\theta}(Z_1,Z_2)\Big|X_{i1},X_{i2}\right\}\right]=0.$$

It is clear from the expressions for the score equations given in Eqs. (4.15) through (4.17) that there is no explicit expression for the MLEs $(\hat{\lambda}_1,\hat{\lambda}_2,\hat{\theta})$ and that numerical techniques must be used. In such procedures, the method of moments estimates $(\tilde{\lambda}_1,\tilde{\lambda}_2,\tilde{\theta})$ could be used as starting values. In order to avoid difficulties in the optimization, one could consider a reparametrization that removes parameter constraints. In particular, let

$$\zeta_1=\ln(\lambda_1),\quad \zeta_2=\ln(\lambda_2),\quad \eta=\ln\{\theta/(1-\theta)\}.$$

so that $(\zeta_1,\zeta_2,\eta)\in\mathbb{R}^3$. The invariance property of maximum likelihood estimates then ensures that the MLEs $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\theta}$ are equal to $\exp(\hat{\zeta}_1)$, $\exp(\hat{\zeta}_2)$, $\exp(\hat{\eta})/\{1+\exp(\hat{\eta})\}$, respectively.

If $\theta\in(0,1)$, simple but tedious calculations show that the usual regularity conditions are satisfied, and hence the MLEs are consistent and asymptotically Gaussian. In particular, let Ψ denote the parameter vector and accordingly $\hat{\Psi}=(\hat{\lambda}_1,\hat{\lambda}_2,\hat{\theta})$ denote the MLE. Then, as $n\to\infty$,

$$\sqrt{n}\,(\hat{\Psi}-\Psi)\rightsquigarrow\mathcal{N}(0,\mathcal{I}^{-1})$$

where $\mathcal{I}$ denotes the Fisher information. As computation of the latter quantity is difficult, in practical implementations the asymptotic variance could be estimated using the bootstrap method.

Remark 4.1 There could be issues in optimizing the log-likelihood due to the form of the contribution of the comonotonic shock. In particular, as was demonstrated in Chapter 3 the method of moments estimates may yield non-finite values for the log-likelihood, in which case it becomes difficult to find reasonable starting values. However, this issue did not arise in our simulations.

The EM algorithm

As reviewed in Chapter 2, the EM algorithm is a useful tool for finding the MLEs in the classical bivariate Poisson model. Similarly to the classical set-up, the construction in the proposed bivariate Poisson model relies on latent variables, specifically the independent components (Y_1,Y_2) and the comonotonic shock (Z_1,Z_2). Conceivably, the EM algorithm could prove to be helpful in the $\mathcal{BP}(\Lambda,\theta)$ family as well.

The EM algorithm is an iterative procedure for determining maximum likelihood estimates in missing data problems. The procedure involves two main steps, namely the expectation step, or E-step, and the maximization step, called the M-step. To start, the data must be *augmented* so that an expression for the *complete data* log-likelihood can be written in terms of the observed and unobserved data. The E-step entails taking the conditional expectation of the complete data log-likelihood, given the observed data and current parameter estimates. The M-step then maximizes the resulting expectation from the E-step to yield the next iteration of parameter estimates, say $\Psi^{(k+1)}$ for $k \in \{0, 1, \ldots\}$. The algorithm then continues iteratively between the E- and M-step until convergence, where the convergence criterion can be in terms of the difference $L(\Psi^{(k+1)}; \mathbf{x}) - L(\Psi^{(k)}; \mathbf{x})$ or $\Psi^{(k+1)} - \Psi^{(k)}$.

It can be shown that at each iteration, the EM algorithm will yield updated parameter estimates $\Psi^{(k)}$ that increase the likelihood. Accordingly, if the underlying observed likelihood is bounded, the sequence $\Psi^{(0)}, \Psi^{(1)}, \ldots$ of EM estimates will converge to either a local or global maximum of the likelihood. There are other conditions which ensure convergence of the EM algorithm to the MLEs; see, e.g., McLachlan and Krishnan (2008) for more details.

In the context of the $\mathcal{BP}(\Lambda, \theta)$ family, one could consider using the comonotonic shock (Z_1, Z_2) as the missing data. In this case, the complete data consist of $(\mathbf{X}, \mathbf{Z})$, where

$$\mathbf{X} = \{(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})\}, \quad \mathbf{Z} = \{(Z_{11}, Z_{12}), \ldots, (Z_{n1}, Z_{n2})\}.$$

Accordingly, the complete data log-likelihood is given by

$$\ell_C(\Lambda, \theta; \mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \ln p_{\Lambda,\theta}(z_{i1}, z_{i2}, z_{i1}, z_{i2})$$

which expands as follows:

$$\begin{aligned}
&\sum_{i=1}^{n} \ln\{g_{(1-\theta)\lambda_1}(x_{i1} - z_{i1}) g_{(1-\theta)\lambda_2}(x_{i2} - z_{i2}) c_{\Lambda,\theta}(z_{i1}, z_{i2})\} \\
&\quad = \sum_{i=1}^{n} [-(1-\theta)\lambda_1 + (x_{i1} - z_{i1}) \ln\{(1-\theta)\lambda_1\} - \ln\{(x_{i1} - z_{i1})!\} \\
&\qquad\qquad - (1-\theta)\lambda_2 + (x_{i2} - z_{i2}) \ln\{(1-\theta)\lambda_2\} - \ln\{(x_{i2} - z_{i2})!\} + \ln c_{\Lambda,\theta}(z_{i1}, z_{i2})] \\
&\quad = -n(1-\theta)(\lambda_1 + \lambda_2) + n \ln(1-\theta)(\bar{x}_1 + \bar{x}_2 - \bar{z}_1 - \bar{z}_2) + n \ln(\lambda_1)(\bar{x}_1 - \bar{z}_1) \\
&\qquad + n \ln(\lambda_2)(\bar{x}_2 - \bar{z}_2) + \sum_{i=1}^{n} \ln c_{\Lambda,\theta}(z_{i1}, z_{i2}) - \sum_{i=1}^{n} \ln\{(x_{i1} - z_{i1})!(x_{i2} - z_{i2})!\}.
\end{aligned}$$

Let $\Psi^{(k)} = (\lambda_1^{(k)}, \lambda_2^{(k)}, \theta^{(k)})$ denote the parameter estimates at stage $k \in \{0, 1, \ldots\}$ and let

$$Q(\Psi \mid \Psi^{(k)}) = \mathrm{E}\{\ell_C(\Lambda, \theta; \mathbf{X}, \mathbf{Z}) \mid \mathbf{x}, \Psi^{(k)}\}$$

represent the conditional expectation of the complete data log-likelihood, given the observed data $\mathbf{x}$ and current parameter estimates $\Psi^{(k)} = (\lambda_1^{(k)}, \lambda_2^{(k)}, \theta^{(k)})$. Following the notation previously introduced, let

$$\bar{q}_1(\Lambda^{(k)}, \theta^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(Z_1 \mid x_{i1}, x_{i2}, \Psi^{(k)}), \quad \bar{q}_2(\Lambda^{(k)}, \theta^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(Z_2 \mid x_{i1}, x_{i2}, \Psi^{(k)}).$$

It is then straightforward to obtain

$$\begin{aligned} Q(\Psi \mid \Psi^{(k)}) = -n(1-\theta)(\lambda_1+\lambda_2) + n\ln(1-\theta)\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)}) - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\} \\ + n\ln(\lambda_1)\{\bar{x}_1 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)})\} + n\ln(\lambda_2)\{\bar{x}_2 - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\} \\ + \sum_{i=1}^{n} \mathrm{E}\{\ln c_{\Lambda,\theta}(Z_{i1}, Z_{i2}) \mid \mathbf{x}, \Psi^{(k)}\} - R(\mathbf{x}, \Psi^{(k)}), \end{aligned}$$

where

$$R(\mathbf{x}, \Psi^{(k)}) = \sum_{i=1}^{n} \mathrm{E}\left[\ln\{(x_{i1} - z_{i1})!(x_{i2} - z_{i2})!\}\right]$$

can be regarded as a remainder term that depends only on the observed data and current parameter estimates. In the maximization step, the parameter estimates are updated according to

$$\Psi^{(k+1)} = \arg\max_{\Psi} Q(\Psi \mid \Psi^{(k)}).$$

The updates are then found by solving

$$\left(\frac{\partial}{\partial \lambda_1} Q(\Psi \mid \Psi^{(k)}), \frac{\partial}{\partial \lambda_2} Q(\Psi \mid \Psi^{(k)}), \frac{\partial}{\partial \theta} Q(\Psi \mid \Psi^{(k)})\right) = \mathbf{0}^\top.$$

Since the left-hand side involves the terms

$$\mathrm{E}\left\{\frac{\partial}{\partial \lambda_1} \ln c_{\Lambda,\theta}(Z_1, Z_2)\middle|\mathbf{x}\right\}, \quad \mathrm{E}\left\{\frac{\partial}{\partial \lambda_2} \ln c_{\Lambda,\theta}(Z_1, Z_2)\middle|\mathbf{x}\right\}, \quad \mathrm{E}\left\{\frac{\partial}{\partial \theta} \ln c_{\Lambda,\theta}(Z_1, Z_2)\middle|\mathbf{x}\right\},$$

this optimization is just as difficult as directly optimizing the observed data log-likelihood to begin with. Seemingly, the EM algorithm is unhelpful when the comonotonic shock is used to augment the data.

As an alternative, one could consider using the underlying uniform random variables generating the comonotonic shock as the missing data. That is, the complete data consist of $\{(X_{11}, X_{12}, U_1),$

..., $(X_{n1}, X_{n2}, U_n)\}$, where for each $i \in \{1, \ldots, n\}$ the comonotonic pair is expressed as

$$(Z_{i1}, Z_{i2}) = \left(G^{-1}_{\theta\lambda_1}(U_i), G^{-1}_{\theta\lambda_2}(U_i)\right).$$

Let $p_{\Lambda,\theta}(x_1, x_2, u)$ denote the joint probability of (X_1, X_2, U). It is straightforward to show that

$$p_{\Lambda,\theta}(x_1, x_2, u) = g_{(1-\theta)\lambda_1}\{x_1 - G^{-1}_{\theta\lambda_1}(u)\} g_{(1-\theta)\lambda_2}\{x_2 - G^{-1}_{\theta\lambda_2}(u)\}.$$

It then follows that the complete data log-likelihood has the form

$$\begin{aligned}
\ell_C(\Lambda, \theta; \mathbf{x}, \mathbf{u}) &= \sum_{i=1}^{n} \ln p_{\Lambda,\theta}(u_i, x_{i1}, x_{i2}) \\
&= \sum_{i=1}^{n} \ln \left[g_{(1-\theta)\lambda_1}\{x_{i1} - G^{-1}_{\theta\lambda_1}(u_i)\} g_{(1-\theta)\lambda_2}\{x_{i2} - G^{-1}_{\theta\lambda_2}(u_i)\}\right] \\
&= \sum_{i=1}^{n} \Big[-(1-\theta)(\lambda_1+\lambda_2) + \ln(1-\theta)\{x_{i1} + x_{i2} - G^{-1}_{\theta\lambda_1}(u_i) - G^{-1}_{\theta\lambda_2}(u_i)\} \\
&\qquad + \ln(\lambda_1)\{x_{i1} - G^{-1}_{\theta\lambda_1}(u_i)\} + \ln(\lambda_2)\{x_{i2} - G^{-1}_{\theta\lambda_2}(u_i)\} \\
&\qquad - \ln\left[\{x_{i1} - G^{-1}_{\theta\lambda_1}(u_i)\}!\{x_{i2} - G^{-1}_{\theta\lambda_2}(u_i)\}!\,\right] \Big]. \qquad (4.19)
\end{aligned}$$

The E-step involves the expectations $\mathrm{E}\{G^{-1}_{\theta\lambda_1}(U_i) \mid \mathbf{x}, \Psi^{(k)}\}$ and $\mathrm{E}\{G^{-1}_{\theta\lambda_2}(U_i) \mid \mathbf{x}, \Psi^{(k)}\}$. Using the same notation previously introduced, these values simplify as follows:

$$\begin{aligned}
\mathrm{E}\{G^{-1}_{\theta\lambda_1}(U_i) \mid \mathbf{x}, \Psi^{(k)}\} &= \mathrm{E}\{Z_1 \mid x_{i1}, x_{i2}, \Psi^{(k)}\} = q_{i1}(\Lambda^{(k)}, \theta^{(k)}), \\
\mathrm{E}\{G^{-1}_{\theta\lambda_2}(U_i) \mid \mathbf{x}, \Psi^{(k)}\} &= \mathrm{E}\{Z_2 \mid x_{i1}, x_{i2}, \Psi^{(k)}\} = q_{i2}(\Lambda^{(k)}, \theta^{(k)}).
\end{aligned}$$

Thus, the E-step $Q(\Psi \mid \Psi^{(k)}) = \mathrm{E}\{\ell_C(\Lambda, \theta; \mathbf{x}, \mathbf{U}) \mid \mathbf{x}, \Psi^{(k)}\}$ yields

$$\begin{aligned}
Q(\Psi \mid \Psi^{(k)}) = &-n(1-\theta)(\lambda_1+\lambda_2) + n\ln(1-\theta)\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)}) - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)}\} \\
&+ n\ln(\lambda_1)\{\bar{x}_1 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)})\} + n\ln(\lambda_2)\{\bar{x}_2 - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\} - R(\mathbf{x}, \Psi^{(k)})
\end{aligned}$$

In the above, $R(\mathbf{x}, \Psi^{(k)})$ denotes the remainder term

$$\sum_{i=1}^{n} \mathrm{E}\left[\ln\left[\{x_{i1} - G^{-1}_{\theta\lambda_1}(u_i)\}!\,\{x_{i2} - G^{-1}_{\theta\lambda_2}(u_i)\}!\,\right] \mid \mathbf{x}, \Psi^{(k)}\right],$$

which depends only on the observed data and parameter estimates $\Psi^{(k)}$. Rearranging the terms in the E-step, $Q(\Psi \mid \Psi^{(k)})$ can be expressed as

$$Q(\Psi \mid \Psi^{(k)}) = -n(1-\theta)(\lambda_1+\lambda_2)+n\left\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)}) - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)}\right\} \ln\left\{(1-\theta)(\lambda_1+\lambda_2)\right\}$$
$$+ n\left\{\bar{x}_1 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)})\right\} \ln\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$
$$+ n\left\{\bar{x}_2 - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\right\} \ln\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right) - R(\mathbf{x}, \Psi^{(k)}). \quad (4.20)$$

Upon further inspection of the form of $Q(\Psi \mid \Psi^{(k)})$ given in (4.20), it is apparent that there are identifiability issues. Letting $\alpha = \lambda_1/(\lambda_1+\lambda_2)$ and $\beta = (1-\theta)(\lambda_1+\lambda_2)$, the expression in (4.20) can be rewritten as

$$Q(\Psi \mid \Psi^{(k)}) = -n\beta + n\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)}) - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\} \ln(\beta)$$
$$+ n\{\bar{x}_1 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)})\} \ln(\alpha) + n\{\bar{x}_2 - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\} \ln(1-\alpha)$$
$$- R(\mathbf{x}, \Psi^{(k)}). \quad (4.21)$$

In the above, both $\mathbf{x}$ and $\Psi^{(k)} = (\Lambda^{(k)}, \theta^{(k)})$ are fixed. Thus, any set of values $(\lambda_1, \lambda_2, \theta)$ that yield the same (α, β) will accordingly result in identical values for $Q(\Psi \mid \Psi^{(k)})$. For example, setting $(\lambda_1, \lambda_2, \theta) = (4, 6, 0.5)$ versus setting $(\lambda_1, \lambda_2, \theta) = (40, 60, 0.05)$ both give $(\alpha, \beta) = (0.4, 5)$ and will therefore both result in the same value of $Q(\Psi \mid \Psi^{(k)})$. Clearly, identifiability issues ensue if the EM algorithm is built using the underlying uniform random variable U generating the comonotonic shock as the latent variable. Conceivably, non-identifiability in the E-step is due to the singularity induced by the comonotonic shock since the underlying Uniform random variable is treated as the latent variable.

Note that this issue is not inherent in the underlying observed likelihood. For example, consider the two cases given by (i) $(\lambda_1, \lambda_2, \theta) = (4, 6, 0.5)$ and (ii) $(\lambda_1, \lambda_2, \theta) = (40, 60, 0.05)$. Suppose a pair $(X_1, X_2) = (0, 0)$ is observed. In the proposed $\mathcal{BP}$ family, the probability of observing $(0, 0)$ is calculated according to

$$f_{\Lambda,\theta}(0,0) = g_{(1-\theta)\lambda_1}(0) g_{(1-\theta)\lambda_2}(0) \min\{G_{\theta\lambda_1}(0), G_{\theta,\lambda_2}(0)\}$$

In case (i), this reduces to

$$0.1353353 \times 0.04978707 \times \min(0.1353353, 0.04978707) = 0.0003354627$$

whereas case (ii) yields

$$3.139133 \times 10^{-17} \times 1.758792 \times 10^{-25} \times \min(0.1353353, 0.04978707) = 2.748785 \times 10^{-43}.$$

4.3.3 Inference functions for margins estimation

The joint probability mass function $f_{\Lambda,\theta}$ can be rewritten in a slightly different manner, thus leading to an alternative formulation for the log-likelihood. Starting from $p_{\Lambda,\theta}(z_1, z_2, x_1, x_2)$, consider the following manipulations. First,

$$\begin{aligned}p_{\Lambda,\theta}(z_1, z_2, x_1, x_2) &= \Pr(X_1 = x_1, X_2 = x_2 \mid Z_1 = z_1, Z_2 = z_2)c_{\Lambda,\theta}(z_1, z_2)\\ &= \Pr(Y_1 = x_1 - z_1)\Pr(Y_2 = x_2 - z_2)c_{\Lambda,\theta}(z_1, z_2)\\ &= \Pr(X_1 = x_1 \mid Z_1 = z_1)\Pr(X_2 = x_1 \mid Z_2 = z_2)c_{\Lambda,\theta}(z_1, z_2).\end{aligned}$$

Now the right-hand side can be rewritten as

$$\Pr(X_1 = x_1, Z_1 = z_1)\Pr(X_2 = x_1, Z_2 = z_2) \times c_{\Lambda,\theta}(z_1, z_2)/\left\{\Pr(Z_1 = z_1)\Pr(Z_2 = z_2)\right\}.$$

This can be further decomposed as

$$\begin{aligned}&\Pr(X_1 = x_1)\Pr(X_2 = x_2) \times \{\Pr(X_1 = x_1, Z_1 = z_1)/\Pr(X_1 = x_1)\}\\ &\quad \times \{\Pr(X_2 = x_1, Z_2 = z_2)/\Pr(X_2 = x_2)\} \times \left[c_{\Lambda,\theta}(z_1, z_2)/\left\{\Pr(Z_1 = z_1)\Pr(Z_2 = z_2)\right\}\right],\end{aligned}$$

which is equivalent to

$$\begin{aligned}&g_{\lambda_1}(x_1)g_{\lambda_2}(x_2) \times \{\Pr(Z_1 = z_1 \mid X_1 = x_1)\Pr(Z_2 = z_2 \mid X_2 = x_2)\}\\ &\qquad\qquad\qquad\qquad \times \left[c_{\Lambda,\theta}(z_1, z_2)/\left\{g_{\theta\lambda_1}(z_1)g_{\theta\lambda_2}(z_2)\right\}\right].\end{aligned}$$

Therefore,

$$p_{\Lambda,\theta}(z_1, z_2, x_1, x_2) = g_{\lambda_1}(x_1)g_{\lambda_2}(x_2)b_{x_1,\theta}(z_1)b_{x_2,\theta}(z_2)c_{\Lambda,\theta}(z_1, z_2)/\left\{g_{\theta\lambda_1}(z_1)g_{\theta\lambda_2}(z_2)\right\},$$

where $b_{x_k,\theta}(z_k)$ denotes the probability mass function of a Binomial random variable with size x_k and probability θ. The above derivation follows from properties of the Poisson distribution, namely, if $W_1 \sim \mathcal{P}(\mu_1)$ is independent of $W_2 \sim \mathcal{P}(\mu_2)$, then $W_1 \mid W_1 + W_2 = w$ follows a Binomial distribution with size w and probability $\mu_1/(\mu_1 + \mu_2)$. Let

$$\omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2) = \frac{b_{x_1,\theta}(z_1)b_{x_2,\theta}(z_2)}{g_{\theta\lambda_1}(z_1)g_{\theta\lambda_2}(z_2)},$$

which can be regarded as a weighting term. Then, $f_{\Lambda,\theta}(x_1, x_2)$ can be rewritten as

$$f_{\Lambda,\theta}(x_1, x_2) = g_{\lambda_1}(x_1)g_{\lambda_2}(x_2) \times \left\{\sum_{z_1=0}^{x_1}\sum_{z_2=0}^{x_2} \omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2)c_{\Lambda,\theta}(z_1, z_2)\right\}, \tag{4.22}$$

which is simply the product of the marginal probability mass functions of X_1 and X_2 times a multiplicative factor accounting for the dependence induced by the comonotonic shocks (Z_1, Z_2). Note that this form resembles the approach considered in Lakshminarayana et al. (1999).

Working with the form (4.22), the likelihood becomes

$$\begin{aligned}
L(\Lambda, \theta; \mathbf{x}_1, \mathbf{x}_2) &= \prod_{i=1}^{n} f_{\Lambda,\theta}(x_{i1}, x_{i2}) \\
&= \prod_{i=1}^{n} \left\{ g_{\lambda_1}(x_{i1}) g_{\lambda_2}(x_{i2}) \times \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2) c_{\Lambda,\theta}(z_1, z_2) \right\} \\
&= \left\{ \prod_{i=1}^{n} g_{\lambda_1}(x_{i1}) \right\} \times \left\{ \prod_{i=1}^{n} g_{\lambda_2}(x_{i2}) \right\} \\
&\quad \times \prod_{i=1}^{n} \left\{ \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2) c_{\Lambda,\theta}(z_1, z_2) \right\} \\
&= L_1(\lambda_1; \mathbf{x}_1) \times L_2(\lambda_2; \mathbf{x}_2) \times L_D(\Lambda, \theta; \mathbf{x}_1, \mathbf{x}_2)
\end{aligned}$$

where for $k \in \{1, 2\}$, $L_k(\lambda_k; \mathbf{x}_k)$ denotes the marginal likelihood for component k with observations $\mathbf{x}_k$, and $L_D(\Lambda, \theta; \mathbf{x}_1, \mathbf{x}_2)$ encompasses the dependence. In the same way, the log-likelihood can be broken down into two marginal components and a dependence term, viz.

$$\ell(\Lambda, \theta; \mathbf{x}_1, \mathbf{x}_2) = \ell_1(\lambda_1; \mathbf{x}_1) + \ell_2(\lambda_2; \mathbf{x}_2) + \ell_D(\Lambda, \theta; \mathbf{x}_1, \mathbf{x}_2). \tag{4.23}$$

The decomposition of the log-likelihood given in (4.23) has a similar form as that in copula models, as shown in Chapter 3. As such, the inference function for margins (IFM) framework provides yet another approach for estimation in the proposed $\mathcal{BP}$ family. As was outlined in Chapter 3, under suitable regularity conditions, the IFM estimates, denoted by $(\check{\lambda}_1, \check{\lambda}_2, \check{\theta})$, are the solutions to the system

$$\left(\frac{\partial}{\partial \lambda_1} \ell_1(\lambda_1), \frac{\partial}{\partial \lambda_2} \ell_2(\lambda_2), \frac{\partial}{\partial \theta} \ell(\lambda_1, \lambda_2, \theta) \right) = \mathbf{0}^\top.$$

In the $\mathcal{BP}(\Lambda, \theta)$ family, the margins are univariate Poisson and thus, for $j \in \{1, 2\}$, the marginal log-likelihoods have the form

$$\ell_j(\lambda_j) = -n\lambda_j + n\bar{x}_j \ln(\lambda_j) - \sum_{i=1}^{n} \ln(x_{ij}!).$$

IFM estimation then yields the marginal parameter estimates $\check{\lambda}_1 = \bar{x}_1$ and $\check{\lambda}_2 = \bar{x}_2$. Note that the latter two estimates coincide with those obtained via the method of moments. The IFM estimate for the dependence parameter is determined by

$$\check{\theta} = \arg\max_{\theta} \ \ell(\check{\Lambda}, \theta),$$

which can be found by solving $\partial\ell(\check{\Lambda}, \theta)/\partial\theta = 0$. Based on the representation from (4.22), the dependence component of the log-likelihood is given by

$$\ell_D(\Lambda, \theta) = \sum_{i=1}^{n} \ln \left\{ \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} \omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2) c_{\Lambda,\theta}(z_1, z_2) \right\}.$$

In particular, since $\partial\ell(\Lambda, \theta)/\partial\theta = \partial\ell_D(\Lambda, \theta)/\partial\theta$, the estimating equation involving θ in the IFM approach is similar to that given in (4.17), namely

$$\frac{\partial}{\partial\theta} \ell_D(\Lambda, \theta) = n(\lambda_1 + \lambda_2) - \frac{n}{(1-\theta)} \{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda, \theta) - \bar{q}_2(\Lambda, \theta)\} + \sum_{i=1}^{n} \mathrm{E} \left\{ \frac{\partial}{\partial\theta} \ln c_{\Lambda,\theta}(z_1, z_2) \middle| x_{i1}, x_{i2} \right\}. \quad (4.24)$$

In the IFM setting, however, the marginal parameters are held fixed at $\check{\Lambda} = (\bar{x}_1, \bar{x}_2)$.

As was the case for maximum likelihood estimation, numerical procedures can be used to obtain the IFM estimate $\check{\theta}$. As an alternative, the EM algorithm can be applied to the pseudo likelihood $L(\check{\Lambda}, \theta)$, i.e., the observed likelihood wherein the marginal parameters are held fixed at their respective marginal MLEs, to obtain the IFM estimate $\check{\theta}$. In this setting, each EM iteration will update the dependence parameter only to yield a sequence $\theta^{(0)}, \theta^{(1)}, \ldots$

Suppose that the underlying uniform random variables generating the comonotonic shocks are used as the missing data in the EM set-up. Then, the complete data pseudo log-likelihood is similar to that given in (4.20), only with $(\lambda_1, \lambda_2) = (\bar{x}_1, \bar{x}_2)$ so that

$$\begin{aligned} \ell_C(\check{\Lambda}, \theta; \mathbf{x}, \mathbf{u}) = \sum_{i=1}^{n} \Big[&- (1-\theta)(\bar{x}_1 + \bar{x}_2) + \ln(1-\theta)\{x_{i1} + x_{i2} - G^{-1}_{\theta\bar{x}_1}(u_i) - G^{-1}_{\theta\bar{x}_2}(u_i)\} \\ &+ \ln(\bar{x}_1)\{x_{i1} - G^{-1}_{\theta\bar{x}_1}(u_i)\} + \ln(\bar{x}_2)\{x_{i2} - G^{-1}_{\theta\bar{x}_2}(u_i)\} \\ &- \ln\big[\{x_{i1} - G^{-1}_{\theta\bar{x}_1}(u_i)\}!\{x_{i2} - G^{-1}_{\theta\bar{x}_2}(u_i)\}!\big] \Big]. \end{aligned} \quad (4.25)$$

The E-step involves the expectations

$$\mathrm{E}\{G^{-1}_{\theta\bar{x}_1}(U_i) \mid x_{i1}, x_{i2}, \theta^{(k)}\} = \mathrm{E}\{Z_1 \mid x_{i1}, x_{i2}, \theta^{(k)}\} = q_{i1}(\check{\Lambda}, \theta^{(k)}),$$
$$\mathrm{E}\{G^{-1}_{\theta\bar{x}_2}(U_i) \mid x_{i1}, x_{i2}, \theta^{(k)}\} = \mathrm{E}\{Z_2 \mid x_{i1}, x_{i2}, \theta^{(k)}\} = q_{i2}(\check{\Lambda}, \theta^{(k)}).$$

Using similar notation as was previously introduced, let

$$\bar{q}_1(\check{\Lambda}, \theta^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} q_{i1}(\check{\Lambda}, \theta^{(k)}), \quad \bar{q}_2(\check{\Lambda}, \theta^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} q_{i2}(\check{\Lambda}, \theta^{(k)}),$$

with $\check{\Lambda} = (\bar{x}_1, \bar{x}_2)$. The E-step then yields

$$\begin{aligned} Q(\theta \mid \theta^{(k)}) &= -n(1-\theta)(\bar{x}_1 + \bar{x}_2) + n\ln(1-\theta)\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\check{\Lambda}, \theta^{(k)}) - \bar{q}_2(\check{\Lambda}, \theta^{(k)})\} \\ &\quad + n\ln(\bar{x}_1)\{\bar{x}_1 - \bar{q}_1(\check{\Lambda}, \theta^{(k)})\} + n\ln(\bar{x}_2)\{\bar{x}_2 - \bar{q}_2(\check{\Lambda}, \theta^{(k)})\} - R(\mathbf{x}, \theta^{(k)}), \end{aligned} \tag{4.26}$$

where the remainder term

$$R(\mathbf{x}, \theta^{(k)}) = \sum_{i=1}^{n} \mathrm{E}\Big[\ln[\{x_{i1} - G^{-1}_{\theta\bar{x}_1}(u_i)\}!\{x_{i2} - G^{-1}_{\theta\bar{x}_2}(u_i)\}!] \mid \mathbf{x}, \theta^{(k)} \Big]$$

is a function of the observed data $\mathbf{x}$ and parameter estimate $\theta^{(k)}$ only. This leads to the M-step, viz.

$$\begin{aligned} \frac{\partial}{\partial\theta} Q(\theta \mid \theta^{(k)}) &= n(\bar{x}_1 + \bar{x}_2) - \frac{n}{(1-\theta)}\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\check{\Lambda}, \theta^{(k)}) - \bar{q}_2(\check{\Lambda}, \theta^{(k)})\} = 0 \\ \Leftrightarrow \theta &= 1 - \left\{ \frac{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\check{\Lambda}, \theta^{(k)}) - \bar{q}_2(\check{\Lambda}, \theta^{(k)})}{\bar{x}_1 + \bar{x}_2} \right\}. \end{aligned}$$

Thus, the sequence of EM updates are given by

$$\theta^{(k+1)} = \frac{\bar{q}_1(\check{\Lambda}, \theta^{(k)}) + \bar{q}_2(\check{\Lambda}, \theta^{(k)})}{\bar{x}_1 + \bar{x}_2}. \tag{4.27}$$

The form of the EM update given in (4.27) is intuitive and can be interpreted as the proportion of the overall total average of X_1 and X_2 due to the comonotonic shock (Z_1, Z_2). The pooling of the information in terms of the summation is also intuitive in that θ is common to both components in (Z_1, Z_2). Moreover, the form of the updates ensures that the iterates $\theta^{(k)}$ will always fall within the parameter boundaries $[0, 1]$. Indeed, by construction $0 \leqslant Z_1 \leqslant X_1$ and $0 \leqslant Z_2 \leqslant X_2$, and thus for an observed pair (x_{i1}, x_{i2}) it follows that $0 \leqslant \bar{q}_1(\check{\Lambda}, \theta^{(k)}) \leqslant \bar{x}_1$ and $0 \leqslant \bar{q}_2(\check{\Lambda}, \theta^{(k)}) \leqslant \bar{x}_2$. Consequently, at each iteration $k \in \{1, 2, \ldots\}$, the EM update $\theta^{(k)} \in [0, 1]$.

Under certain conditions, the EM sequence $\theta^{(k)}$ will converge to the MLE corresponding to the pseudo likelihood $L(\check{\Lambda}, \theta)$, i.e., the IFM estimate. In particular, as stated in Theorem 7.5.2 of Casella and Berger (2002) and Theorem 3.2 of McLachlan and Krishnan (2008), if the E-step $Q(\Psi \mid \Psi')$ is continuous in both Ψ and Ψ', then all limit points of the sequence $\theta^{(k)}$ of EM updates are stationary points of the likelihood $L(\check{\Lambda}, \theta)$ and $L(\check{\Lambda}, \theta^{(k)})$ converges monotonically to $L(\check{\Lambda}, \theta^*)$ for some stationary point θ^*. As the form of $Q(\theta \mid \theta')$ given in (4.26) meets this regularity criterion, the sequence $\theta^{(k)}$ of EM updates will converge to a stationary point of $L(\check{\Lambda}, \theta)$.

It is important to emphasize that in this context the EM algorithm yields the IFM estimate $\check{\theta}$ and not the MLE $\hat{\theta}$ as the EM algorithm is being applied to the pseudo log-likelihood $\ell(\check{\Lambda}, \theta)$ and not the true model log-likelihood $\ell(\Lambda, \theta)$.

In addition to the continuity condition previously described, other assumptions necessary for the convergence an EM sequence are outlined in Section 3.4.2 of McLachlan and Krishnan (2008). These assumptions include

(i) The parameter space Ω is a subset of the d-dimensional Euclidean space $\mathbb{R}^d$,

(ii) $\Omega_{\Psi_0} = \{\Psi \in \Omega : L(\Psi) \geqslant L(\Psi_0)\}$ is compact for any $L(\Psi_0) > -\infty$

(iii) $L(\Psi)$ is continuous in Ω and differentiable in the interior of Ω.

All these conditions are satisfied here.

As detailed in Chapter 3, the asymptotic variance of the IFM estimates can be expressed in terms of the Fisher information matrix as well as the univariate marginal Fisher information matrices. Let $\check{\Psi} = (\check{\lambda}_1, \check{\lambda}_2, \check{\theta})$ denote the IFM estimators and Ψ denote the true parameter values. Then, as $n \to \infty$, the limiting distribution is Gaussian, viz.

$$\sqrt{n}(\check{\Psi} - \Psi) \rightsquigarrow \mathcal{N}(0, V).$$

Let $\check{\boldsymbol{s}}$ denote the set of estimating equations used to derive the IFM estimates, namely

$$\check{\boldsymbol{s}}^{\top} = \left(\frac{\partial}{\partial \lambda_1} \ell_1(\lambda_1), \frac{\partial}{\partial \lambda_2} \ell_2(\lambda_2), \frac{\partial}{\partial \theta} \ell(\lambda_1, \lambda_2, \theta) \right).$$

Then, as shown in Chapter 3, the asymptotic variance V is defined by

$$V = (-D_{\check{s}}^{-1}) M_{\check{s}} (-D_{\check{s}}^{-1})^{\top},$$

where

$$M_{\check{s}} = \text{cov}\left\{\check{\boldsymbol{s}}(\mathbf{X}; \Psi)\right\}, \qquad D_{\check{s}} = \mathrm{E}\left\{ \frac{\partial \check{\boldsymbol{s}}(\mathbf{X}; \Psi)}{\partial \Psi^{\top}} \right\}.$$

Using similar notation as introduced in Chapter 3, let $\mathcal{J}_{jj}$ denote the information matrix associated with the jth margin and set $\mathcal{J}_{jk} = \text{cov}(\check{\text{s}}_j, \check{\text{s}}_k)$, for $j, k \in \{1, 2\}$. In the $\mathcal{BP}$ model, this reduces to

$$\begin{aligned}
\mathcal{J}_{jj} &= -\mathrm{E}\left\{ \frac{\partial^2}{\partial \lambda_j^2} \ell_j(\lambda_j) \right\} = -\mathrm{E}\left\{ \frac{\partial^2}{\partial \lambda_j^2} \ln g_{\lambda_j}(X_j) \right\} = \frac{1}{\lambda_j}, \\
\mathcal{J}_{jk} &= \text{cov}\left\{ \frac{\partial}{\partial \lambda_j} \ell_j(\lambda_j), \frac{\partial}{\partial \lambda_k} \ell_k(\lambda_k) \right\} = \text{cov}\left\{ \frac{\partial}{\partial \lambda_j} \ln g_{\lambda_j}(X_j), \frac{\partial}{\partial \lambda_k} \ln g_{\lambda_k}(X_k) \right\} \\
&= \text{cov}\left(-1 + \frac{X_j}{\lambda_j}, -1 + \frac{X_k}{\lambda_k} \right) = \frac{1}{\lambda_j \lambda_k} m_{\lambda_j, \lambda_k}(\theta),
\end{aligned}$$

where $m_{\lambda_j,\lambda_k}(\theta) = \text{cov}(X_j, X_k)$ for $(X_j, X_k) \sim \mathcal{BP}(\Lambda, \theta)$. Use $\mathcal{I}$ to denote the Fisher Information matrix, i.e.,

$$\mathcal{I} = -\text{E}\left\{ \frac{\partial^2}{\partial\Psi\partial\Psi^\top}\ \ell(\Lambda, \theta) \right\} = -\text{E}\left\{ \frac{\partial^2}{\partial\Psi\partial\Psi^\top}\ \ln f_{\Lambda,\theta}(X_1, X_2) \right\},$$

with $\Psi = (\lambda_1, \lambda_2, \theta)$. Further, define a_j to be equal to $-\mathcal{I}_{33}\mathcal{I}_{3j}\mathcal{J}_{jj}^{-1}$ for $j \in \{1, 2\}$, where $\mathcal{I}_{jk}$ represents the (j, k) element of the Fisher Information matrix. Note that in the $\mathcal{BP}$ model, each of the quantities $\mathcal{J}_{jk}, \mathcal{I}_{jk}, a_j$ is a scalar, for $j, k \in \{1, 2, 3\}$.

Following the simplifications for V derived in Joe (2005), the asymptotic variance of the IFM estimator $\breve{\Psi} = (\breve{\lambda}_1, \breve{\lambda}_2, \breve{\theta})$ in the $\mathcal{BP}$ model, for each (j, k) element denoted V_{jk}, is given by

$$\begin{aligned}
V_{11} &= \mathcal{J}_{11}^{-1} = \lambda_1, \\
V_{12} &= \mathcal{J}_{11}^{-1}\mathcal{J}_{12}\mathcal{J}_{22}^{-1} = \lambda_1 \left\{ \frac{1}{\lambda_1\lambda_2} m_{\lambda_1,\lambda_2}(\theta) \right\} \lambda_2 = m_{\lambda_1,\lambda_2}(\theta), \\
V_{13} &= \mathcal{J}_{11}^{-1} \sum_{k=1}^{2} \mathcal{J}_{1k} a_k^\top = \lambda_1 \left\{ \frac{1}{\lambda_1}(-\mathcal{I}_{33}^{-1}\mathcal{I}_{31})\lambda_1 + \frac{1}{\lambda_1\lambda_2} m_{\lambda_1,\lambda_2}(\theta)(-\mathcal{I}_{33}^{-1}\mathcal{I}_{32})\lambda_2 \right\} \\
&= -\mathcal{I}_{33}^{-1} \left\{ \lambda_1\mathcal{I}_{31} + m_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{32} \right\}, \\
V_{22} &= \mathcal{J}_{22}^{-1} = \lambda_2, \\
V_{23} &= \mathcal{J}_{22}^{-1} \sum_{k=1}^{2} \mathcal{J}_{2k} a_k^\top = \lambda_2 \left\{ \frac{1}{\lambda_1\lambda_2} m_{\lambda_1,\lambda_2}(\theta)(-\mathcal{I}_{33}^{-1}\mathcal{I}_{31})\lambda_1 + \frac{1}{\lambda_2}(-\mathcal{I}_{33}^{-1}\mathcal{I}_{32})\lambda_2 \right\} \\
&= -\mathcal{I}_{33}^{-1} \left\{ m_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{31} + \lambda_2\mathcal{I}_{32} \right\}, \\
V_{33} &= \mathcal{I}_{33}^{-1} + \sum_{j=1}^{2}\sum_{k=1}^{2} a_j \mathcal{J}_{jk} a_k^\top = \mathcal{I}_{33}^{-1} + (-\mathcal{I}_{33}^{-1}\mathcal{I}_{31}\lambda_1)\left(\frac{1}{\lambda_1}\right)(-\mathcal{I}_{33}^{-1}\mathcal{I}_{31}\lambda_1) \\
&\qquad + (-\mathcal{I}_{33}^{-1}\mathcal{I}_{31}\lambda_1)\left\{ \frac{1}{\lambda_1\lambda_2} m_{\lambda_1,\lambda_2}(\theta) \right\}(-\mathcal{I}_{33}^{-1}\mathcal{I}_{32}\lambda_2) \\
&\qquad + (-\mathcal{I}_{33}^{-1}\mathcal{I}_{32}\lambda_2)\left\{ \frac{1}{\lambda_1\lambda_2} m_{\lambda_1,\lambda_2}(\theta) \right\}(-\mathcal{I}_{33}^{-1}\mathcal{I}_{31}\lambda_1) \\
&\qquad + (-\mathcal{I}_{33}^{-1}\mathcal{I}_{32}\lambda_2)\left(\frac{1}{\lambda_2}\right)(-\mathcal{I}_{33}^{-1}\mathcal{I}_{32}\lambda_2) \\
&= \mathcal{I}_{33}^{-1} + (\mathcal{I}_{33}^{-1})^2 \left\{ \lambda_1\mathcal{I}_{31}^2 + 2m_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{31}\mathcal{I}_{32} + \lambda_2\mathcal{I}_{32}^2 \right\}.
\end{aligned}$$

Maximum likelihood estimators are asymptotically efficient, i.e., the asymptotic variance attains the Cramér–Rao lower bound; see, e.g., Theorem 10.1.12 of Casella and Berger (2002). As was discussed in Chapter 3, there is a loss in efficiency when IFM estimation is used rather than a full likelihood estimation procedure. For the marginal parameters, this loss in precision is due to the fact that estimation is based only on the marginal components $\ell_1(\lambda_1)$ and $\ell_2(\lambda_2)$, and systematically ignores the information in the dependence component $\ell_D(\Lambda, \theta)$. The form of V_{33} allows for a straightforward assessment of the loss in efficiency that ensues when $\breve{\theta}$, rather than $\hat{\theta}$, is used to estimate the dependence parameter θ. Indeed, the maximum likelihood estimator has asymptotic variance $\mathcal{I}_{33}^{-1}$ while that for the IFM estimator is V_{33} and the difference between the two is given by

$$V_{33} - \mathcal{I}_{33}^{-1} = (\mathcal{I}_{33}^{-1})^2 \left\{ \lambda_1 \mathcal{I}_{31}^2 + 2m_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{31}\mathcal{I}_{32} + \lambda_2 \mathcal{I}_{32}^2 \right\}.$$

As was previously noted, despite the differences in the various estimation procedures discussed throughout this section, namely the method of moments, maximum likelihood and inference functions for margins, all three result in consistent estimates. Consequently, as the sample size increases the differences between the three estimates $(\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\theta})$, $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta})$ and $(\breve{\lambda}_1, \breve{\lambda}_2, \breve{\theta})$ will be small. This leads to the next section, in which the results from various simulation studies are summarized.

4.4 Simulations

A set of simulations was carried out to assess the performance of the estimation techniques outlined in Section 4.3 for varying values of $\lambda_1 \in \{1, 2, 4, 10\}$, $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$, and varying sample size $n \in \{10, 50, 100, 1000\}$. Note that the value of λ_2 was held fixed at 5 throughout as the effect of the marginal parameter will be symmetric in the two components, thus testing the effects of λ_2 becomes redundant if λ_1 is allowed to vary. In each scenario resulting from a particular combination $(\lambda_1, \lambda_2, \theta, n)$, a random sample from the $\mathcal{BP}(\Lambda, \theta)$ model was generated. In particular, for $i \in \{1, \ldots, n\}$, three mutually independent $\mathcal{U}(0, 1)$ distributed random variables, denoted V_{i1},V_{i2} and U_i, were generated and the observations were set equal to

$$X_{i1} = G^{-1}_{(1-\theta)\lambda_1}(V_{i1}) + G^{-1}_{\theta\lambda_1}(U_i), \quad X_{i2} = G^{-1}_{(1-\theta)\lambda_2}(V_{i2}) + G^{-1}_{\theta\lambda_2}(U_i).$$

Each of the 80 unique scenarios was replicated 500 times resulting in a total of $40{,}000$ iterations. At every iteration, three sets of estimates were produced according to the method of moments (MM), denoted $(\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\theta})$, the maximum likelihood (ML) approach, denoted $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta})$, and the inference function for margins (IFM) approach, denoted $(\breve{\lambda}_1, \breve{\lambda}_2, \breve{\theta})$. Recall that it was shown that the EM algorithm could be used to determine the IFM estimate $\breve{\theta}$. In the practical implementation of the latter, however, it was found that the sequence $\theta^{(0)}, \theta^{(1)}, \ldots$ of EM estimates tended to be slower to converge than directly optimizing the pseudo log-likelihood $\ell(\breve{\Lambda}, \theta)$. As a result, when applying the IFM method in the simulations detailed here, $\breve{\theta}$ was established by directly optimizing the log-likelihood with marginal parameters fixed at their respective marginal MLEs, i.e.,

$$\breve{\theta} = \arg\max_{\theta} \ell(\breve{\Lambda}, \theta) = \arg\max_{\theta} \ell_D(\breve{\Lambda}, \theta).$$

This was done numerically using the `optim` function in R. Similarly, the maximum likelihood estimates were computed using the `optim` function. In both cases, the method of moments estimates were used as starting values. In order to avoid issues at the boundary of the parameter space, if $\tilde{\theta} = 0$ then a starting value of 0.01 was used while when $\tilde{\theta} = 1$ a starting value of 0.99 was used. Additionally, in cases where the sample covariance was negative, the corresponding MM estimate was set to NA and a starting value of 0.01 was used in the IFM and ML estimation procedures.

Out of the $40{,}000$ iterations, there were 2473 instances where $S_{12} < 0$ and accordingly the MM estimate was set to NA. This tended to occur for smaller sample sizes and for lower dependence levels, i.e., smaller values of θ, as one would expect. At the other extreme, there were 2670 iterations that yielded $\tilde{\theta} = 1$. Not surprisingly, this happened more often in smaller sample sizes and for higher values of θ.

There were a small number of iterations where the optimization procedure reached the maximum number of iterations of parameter updates without converging, which is set to 100 by default in R. For the maximum likelihood approach, this maximum was reached in 10 of the $40{,}000$ iterations while for the IFM method this occurred 187 times. This tended to occur when the sample size was small: Specifically all 10 occurrences for the ML method where in scenarios with $n = 10$, while for IFM estimation 173 occurred when $n = 10$, 12 when $n = 50$ and 2 times when $n = 100$. In the summaries provided hereafter, these cases of non-convergence are excluded as well as the instances where the method of moments estimate was set to NA, as previously described.

Figures 4.3 through 4.12 summarize the estimation results for the dependence parameter θ in the 80 distinct scenarios. As expected from asymptotic theory, the variability in the estimation of θ decreases as the sample size n increases; this holds for all estimation methods. It is also apparent from the boxplots that the true underlying value of the marginal parameter, λ_1, seemingly has no effect on the estimation of the dependence parameter. Note that the seemingly poor performance across all estimation techniques in smaller sample sizes and lower levels of θ is, in part, due to the issues previously discussed. Notably, several iterations were disregarded when $S_{12} < 0$ or either the optimization algorithms failed to converge.

As expected based on the theoretical results established in Section 4.3, the method of moments estimator tended to exhibit greater variability than both likelihood-based estimators. Additionally, the IFM method tended to produce slightly more variability in the estimate for the dependence parameter than the full maximum likelihood estimator, although both likelihood-based methods produced very similar results.

The figures further reveal that the method of moments approach seems to work better when the dependence is weaker. That is, the variability in the MM estimates $\tilde{\theta}$ is comparatively smaller when the true value of θ is smaller. In particular, in the scenarios when θ is smaller, e.g., $\theta = 0.1$, the spread in the boxplot for the MM estimates closely resembles that resulting from both the IFM and ML methods. Recall from Proposition 4.3 that the asymptotic variance of the method of moments estimator is given by $\{\gamma'(\sigma_{12})\}^2\, \sigma^2(\theta, \lambda_1, \lambda_2)$. It is not obvious from its general form whether the latter expression is an increasing function of θ. However, when $\lambda_1 = \lambda_2 = \lambda$, it was shown that the asymptotic variance of $\tilde{\theta}$ reduces to $(\lambda + \theta + \theta^2\lambda)/\lambda$, which is obviously an increasing function in θ. The simulations suggest that for any (λ_1, λ_2), the asymptotic variance for the method of moments estimator tends to be larger for larger values of θ.

By the same token, the plots indicate that the variability of the likelihood-based estimators (i.e., $\breve{\theta}$ and $\hat{\theta}$) shrinks as the value of θ increases. This phenomena is intuitive if the independent components (Y_1, Y_2) in Eq. (4.4) are regarded as random Poisson error terms to the comonotonic shock (Z_1, Z_2). Indeed, as θ increases the random error terms decrease so that (X_1, X_2) are predominantly generated by the comonotonic shock (Z_1, Z_2), which are themselves generated from a single standard Uniform random variable. Thus, as θ increases, the variability should decrease accordingly in the likelihood-based estimation approaches.

The estimates $\breve{\theta}$ obtained from the IFM method tended to closely resemble those resulting from maximum likelihood estimation, regardless of the sample size. Recall that the implementation of the IFM estimation is a two-step procedure wherein the marginal parameter estimates, $\breve{\Lambda}$, are set to their respective marginal maximum likelihood estimates while the dependence parameter is estimated according to

$$\breve{\theta} = \arg\max_{\theta} \ell(\breve{\Lambda}, \theta) = \arg\max_{\theta} \ell_D(\breve{\Lambda}, \theta).$$

In contrast, ML estimation simultaneously determines the parameter estimates $\hat{\Psi} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta})$ as

$$\hat{\Psi} = \arg\max_{\Lambda, \theta} \ell(\Lambda, \theta).$$

The simulation results suggest that holding the marginal parameters fixed at $\breve{\Lambda}$ has little effect on the estimation of θ. Further, this seems to indicate that there is very little information regarding the marginal parameters in the dependence portion of the log-likelihood $\ell_D(\Lambda, \theta)$. Figures 4.13 to 4.22 depict the estimation results for the marginal parameter λ_1. The plots provide further evidence of this and show very little difference between using the marginal log-likelihood $\ell_1(\lambda_1)$ for estimation of λ_1 as compared to using the full likelihood $\ell(\Lambda, \theta)$.

The plots in Figure 4.3 to 4.12 also reveal that there are some cases where the IFM and MLE methods resulted in outlier values for the estimate of θ. For example, this phenomena occurred when $(\lambda_1, \theta, n) = (4, 0.9, 1000)$, as well as several other scenarios. Based on the figures, outliers seemed to be produced more often when θ is large and tended to occur more often for the IFM estimation approach. As was previously mentioned, the method of moments estimates were used as starting values for the optimization procedures for both the IFM and MLE methods. In these extremal cases where the true value of θ was larger (e.g., 0.75 and 0.90), the MM estimate $\tilde{\theta}$ was larger than its true value. The results suggest that using $\tilde{\theta}$ as a starting value for the dependence parameter in the IFM and MLE optimization procedures could be inappropriate in some instances.

This led to a second, smaller simulation study in which the effects of the starting values were investigated. Focusing on the scenario where $(\lambda_1, \theta, n) = (4, 0.9, 1000)$, with $\lambda_2 = 5$, both IFM and MLE estimates for θ were produced using the following starting values: (1) the MM estimate $\tilde{\theta}$, (2) 0.10, (3) 0.90 (the true value) and (4) 0.99. This was repeated 500 times. As shown in Figure 4.2, these outlier estimates for θ seem to be produced when the starting value is high. In particular, for the IFM estimation approach, the extreme values for $\check{\theta}$ are produced when either the MM estimate is used as a starting value, or the value 0.99 is used to initiate the optimization algorithm. For the MLE, this was only seen when $\tilde{\theta}$ was used as a starting value. In both approaches, when the MM starting value led to outliers, the value of $\tilde{\theta}$ was quite large (e.g., 0.997 for the IFM approach and 0.998 in the ML estimation). Seemingly, starting too close to the boundary of the parameter space can yield poor results, which is not surprising.

Overall, the implementation of the MM method tended to run the fastest, on average, followed by the IFM approach and last the full maximum likelihood estimation approach. Although this is somewhat inconsequential in the bivariate case as all three methods had relatively similar running times, in higher dimensions this will likely have a much greater impact. This will be discussed in Chapter 6.

Finally, Figure 4.23 provides 2-D plots of the log-likleihood for a random sample of size $n = 1000$ for varying values of θ and (λ_1, λ_2) fixed at $(4, 5)$. In addition, a 3-D plot of the log-likelihood is shown for the case where $\theta = 0.5$ and again $(\lambda_1, \lambda_2) = (4, 5)$.

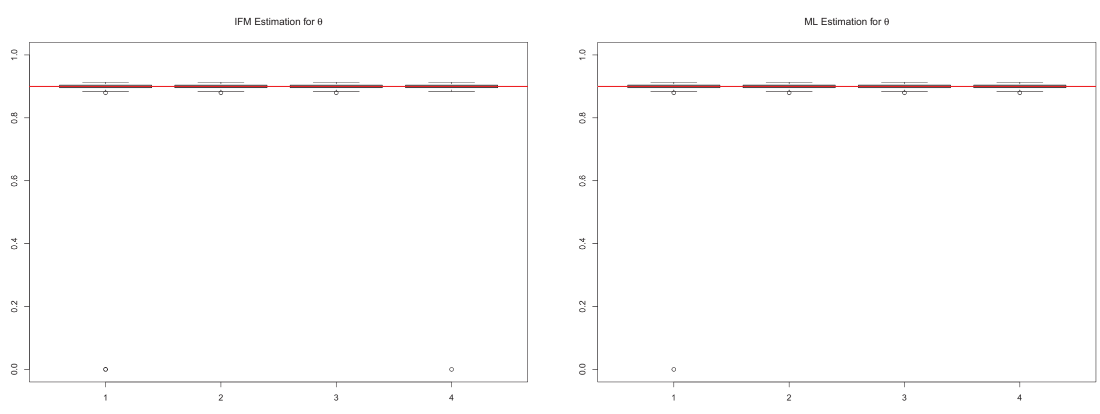

Figure 4.2: Estimation results for θ using different starting values: (1) the method of moments estimate (2) 0.01 (3) 0.99 (4) 0.90 (the true value). The inference for margins results are shown on the left and maximum likelihood estimate on the right.

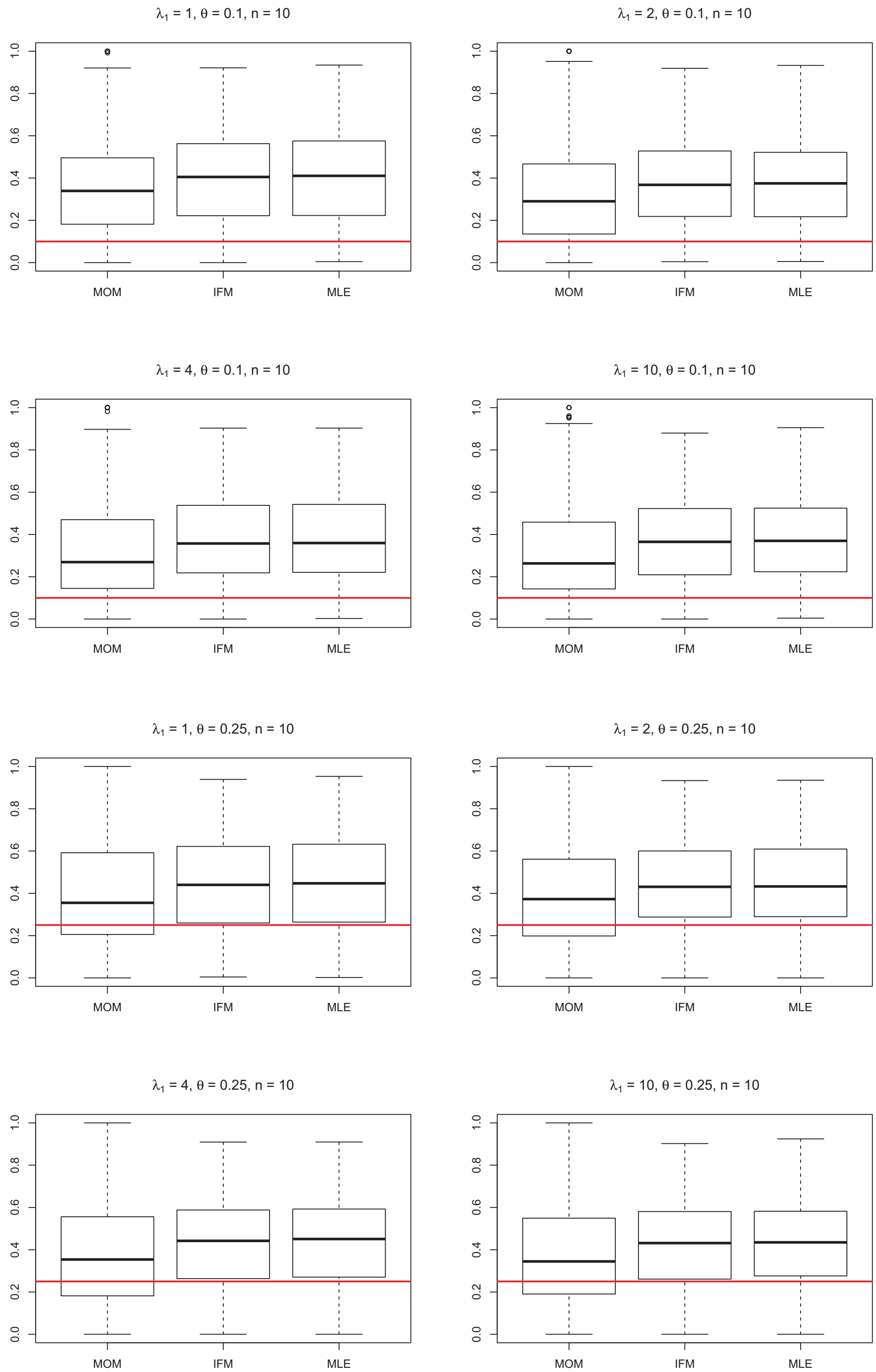

Figure 4.3: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

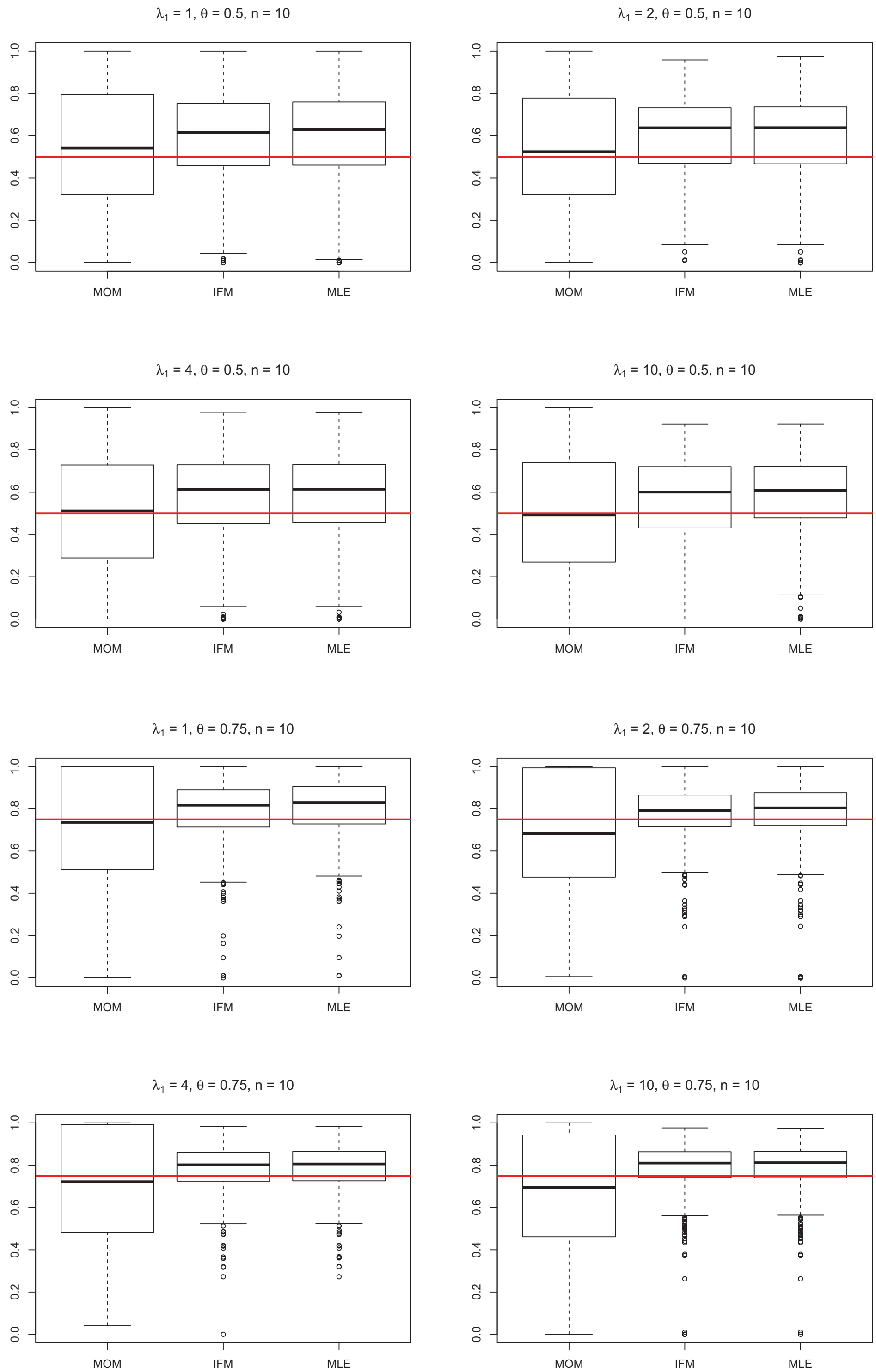

Figure 4.4: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

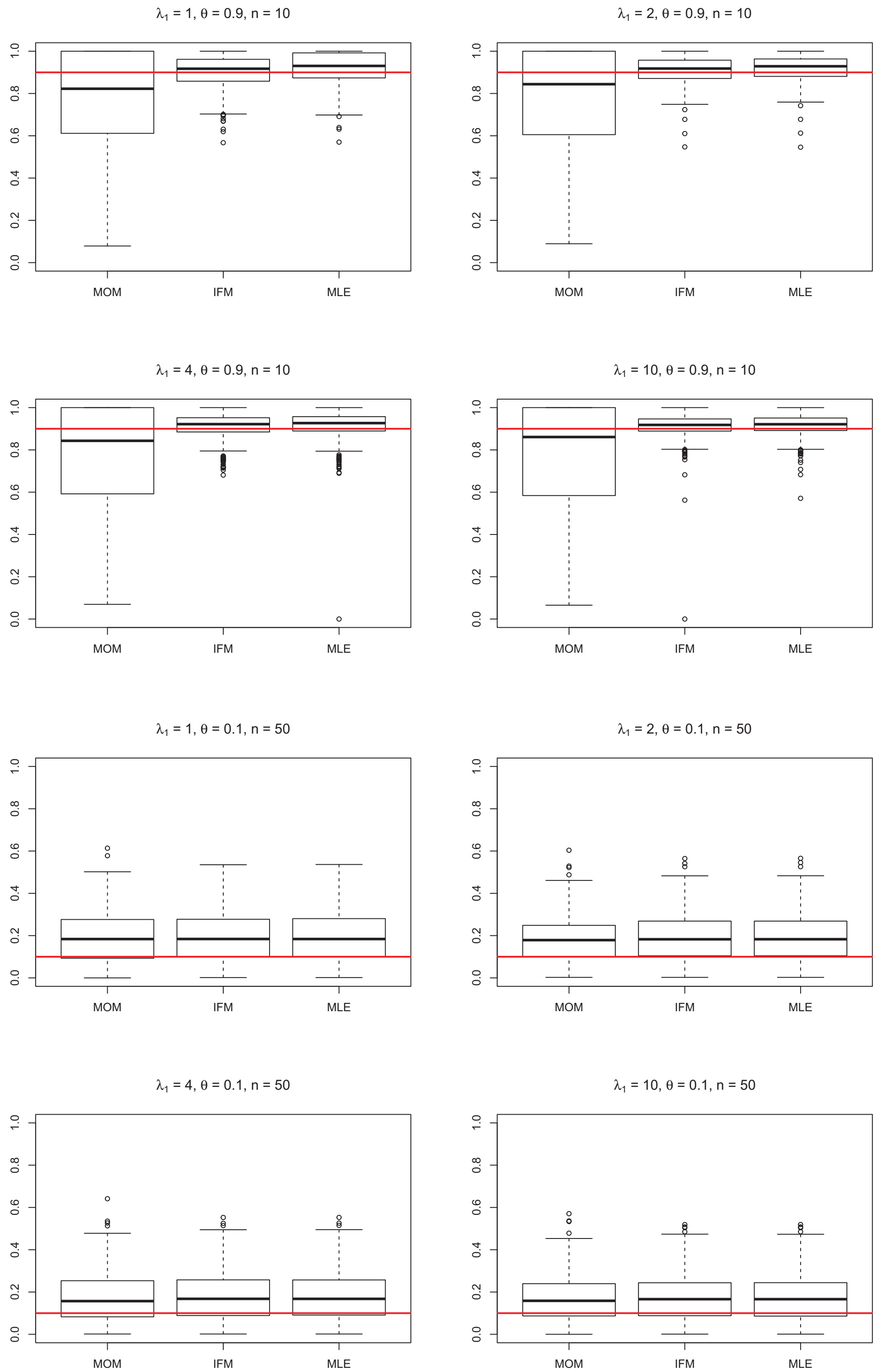

Figure 4.5: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

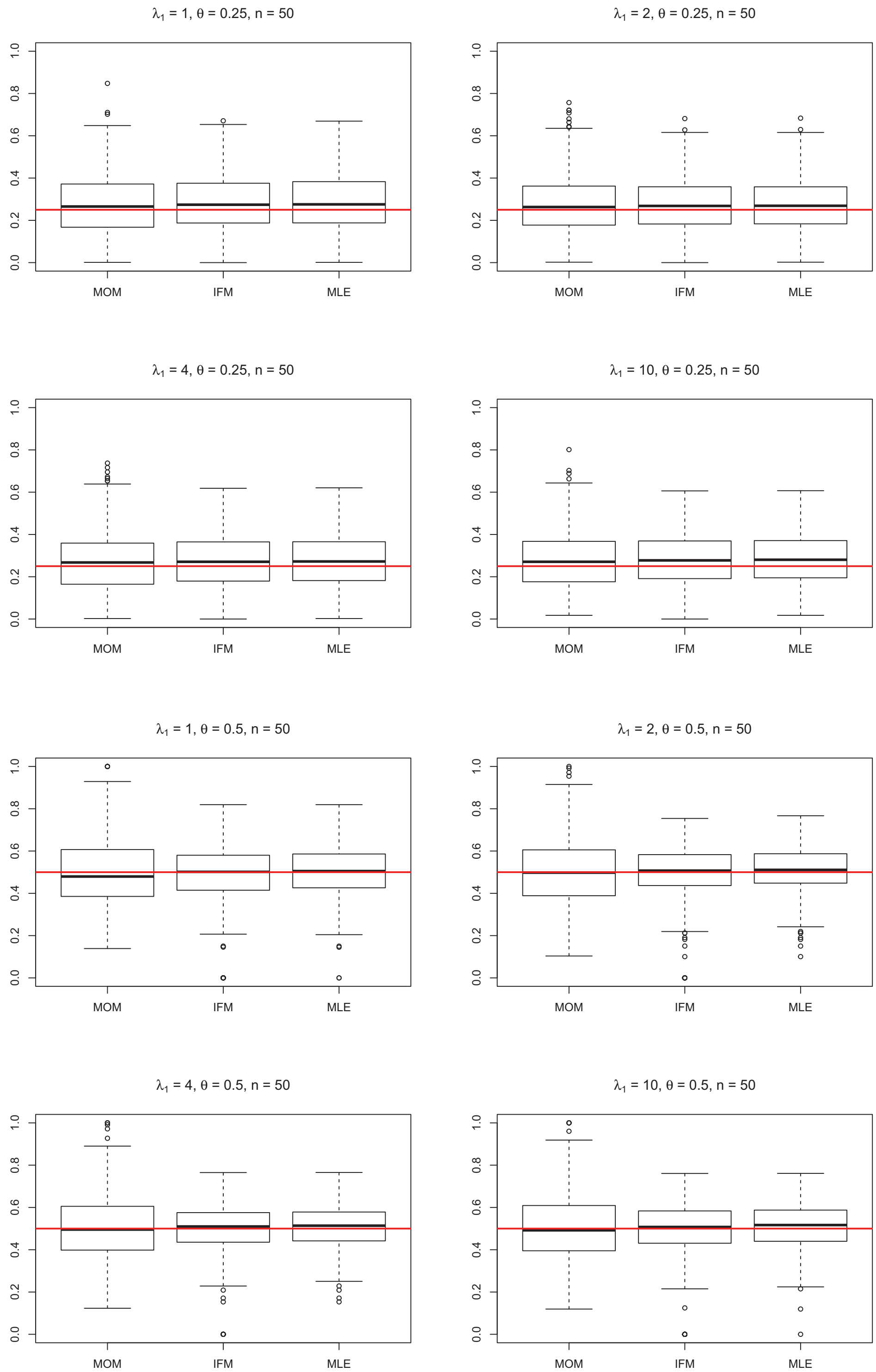

Figure 4.6: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

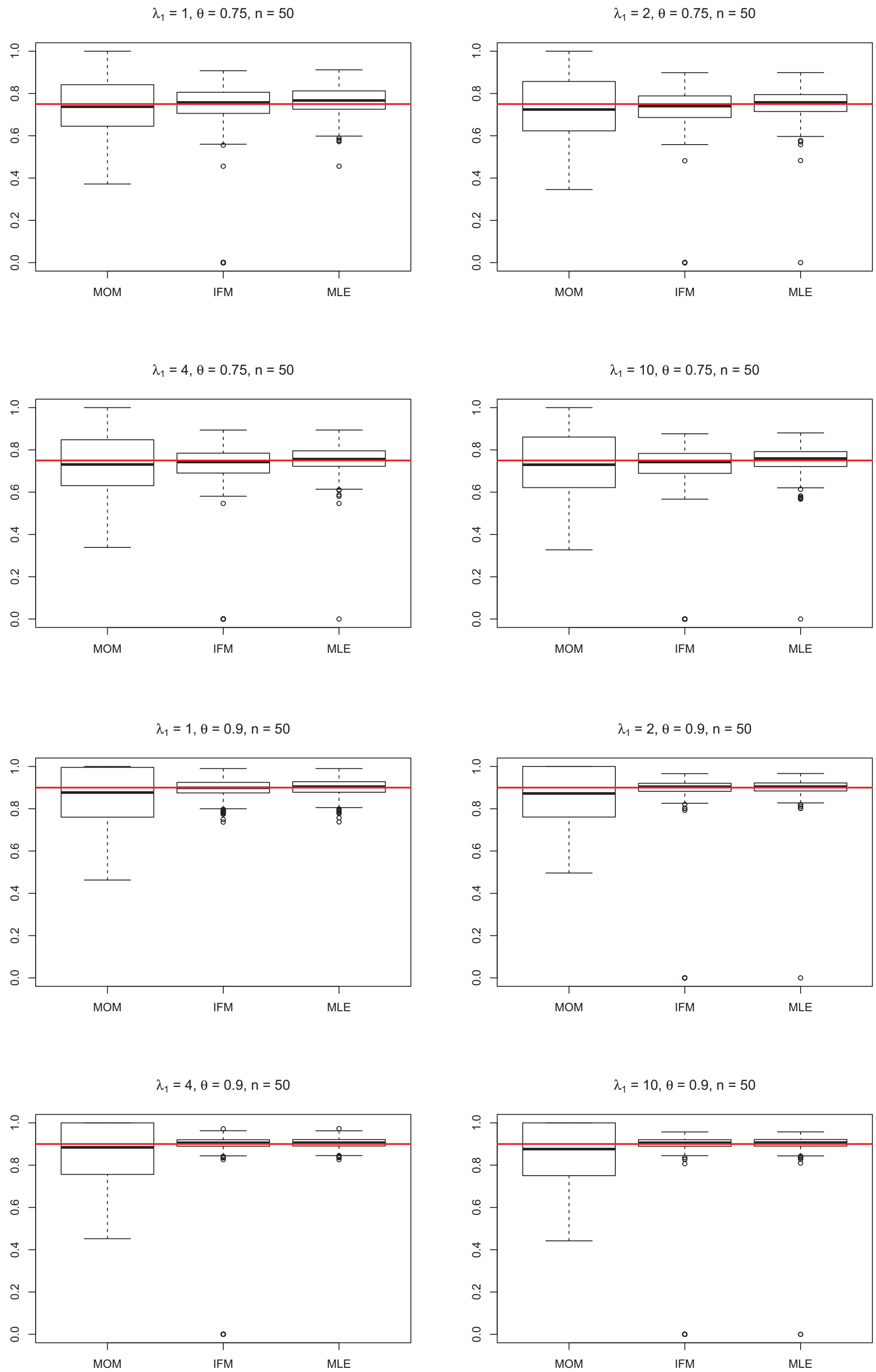

Figure 4.7: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

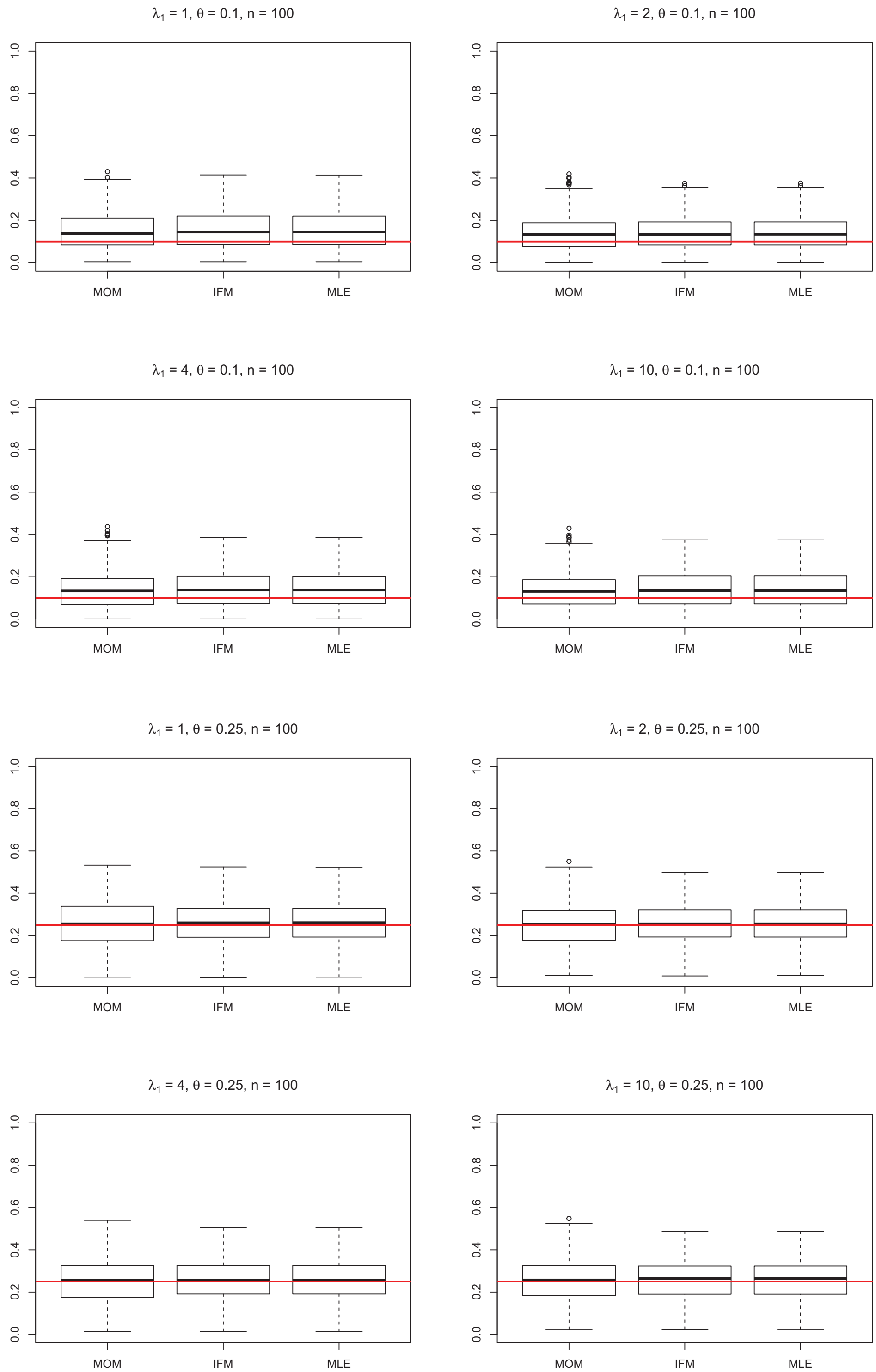

Figure 4.8: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

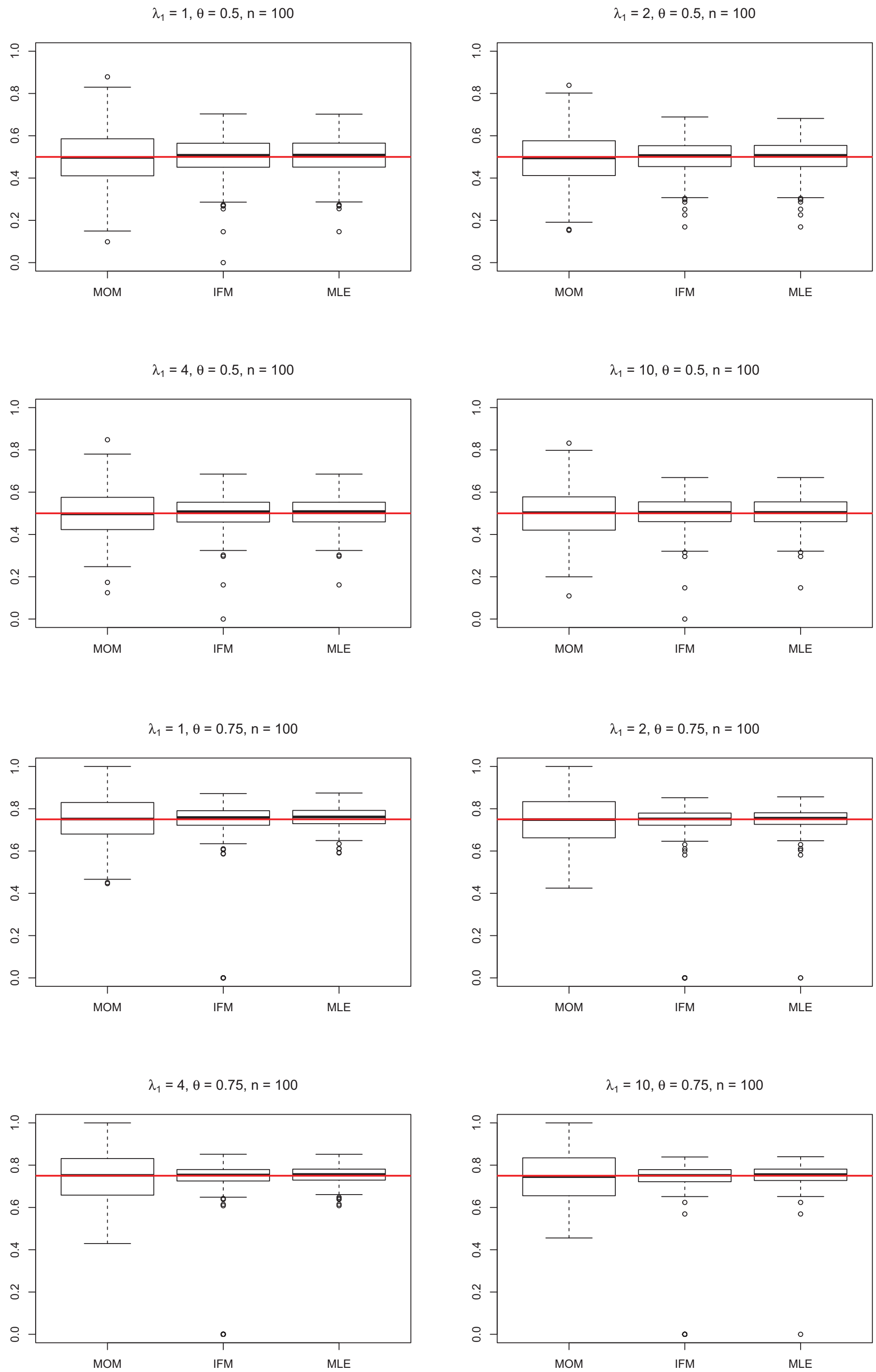

Figure 4.9: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

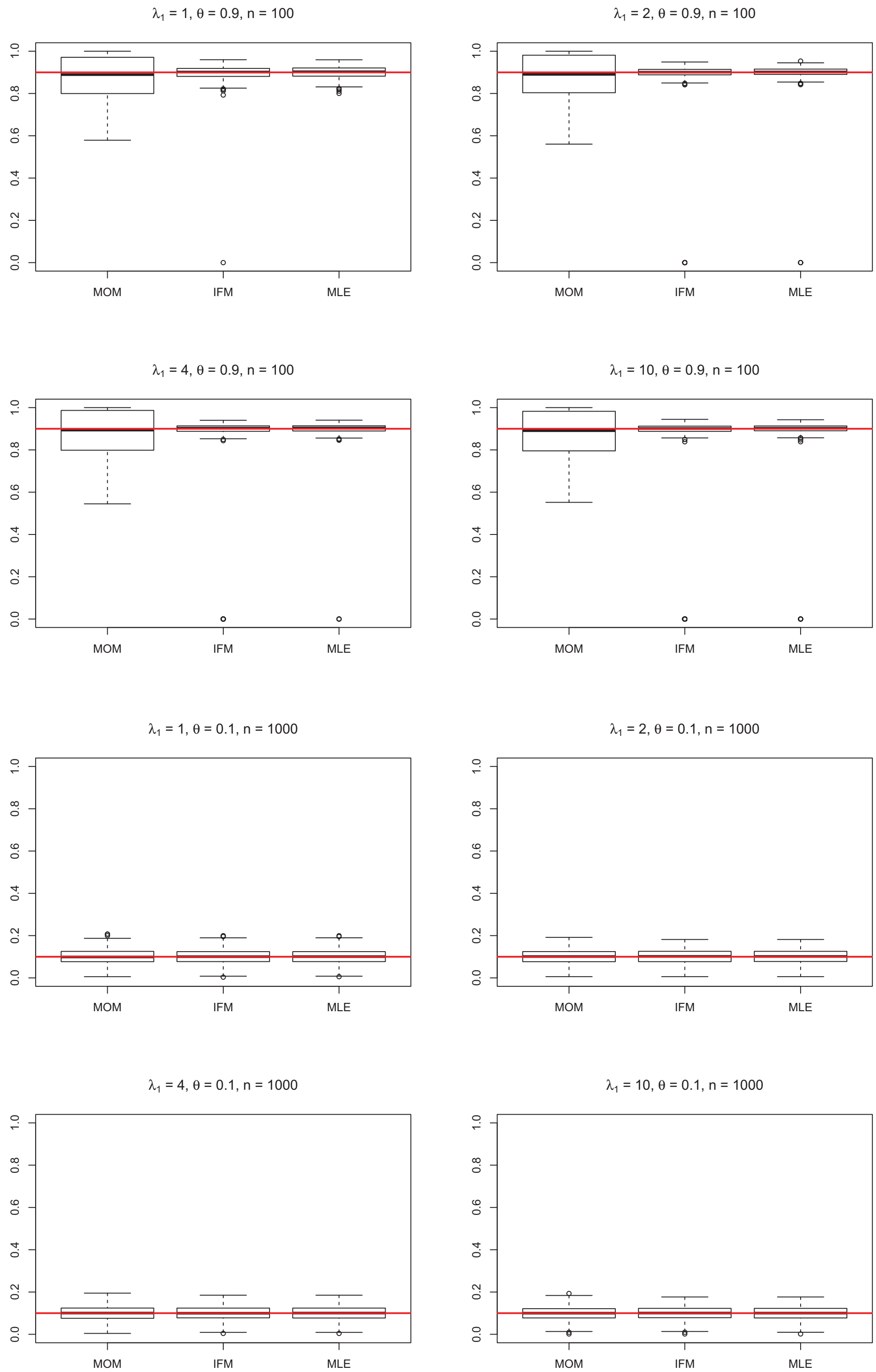

Figure 4.10: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

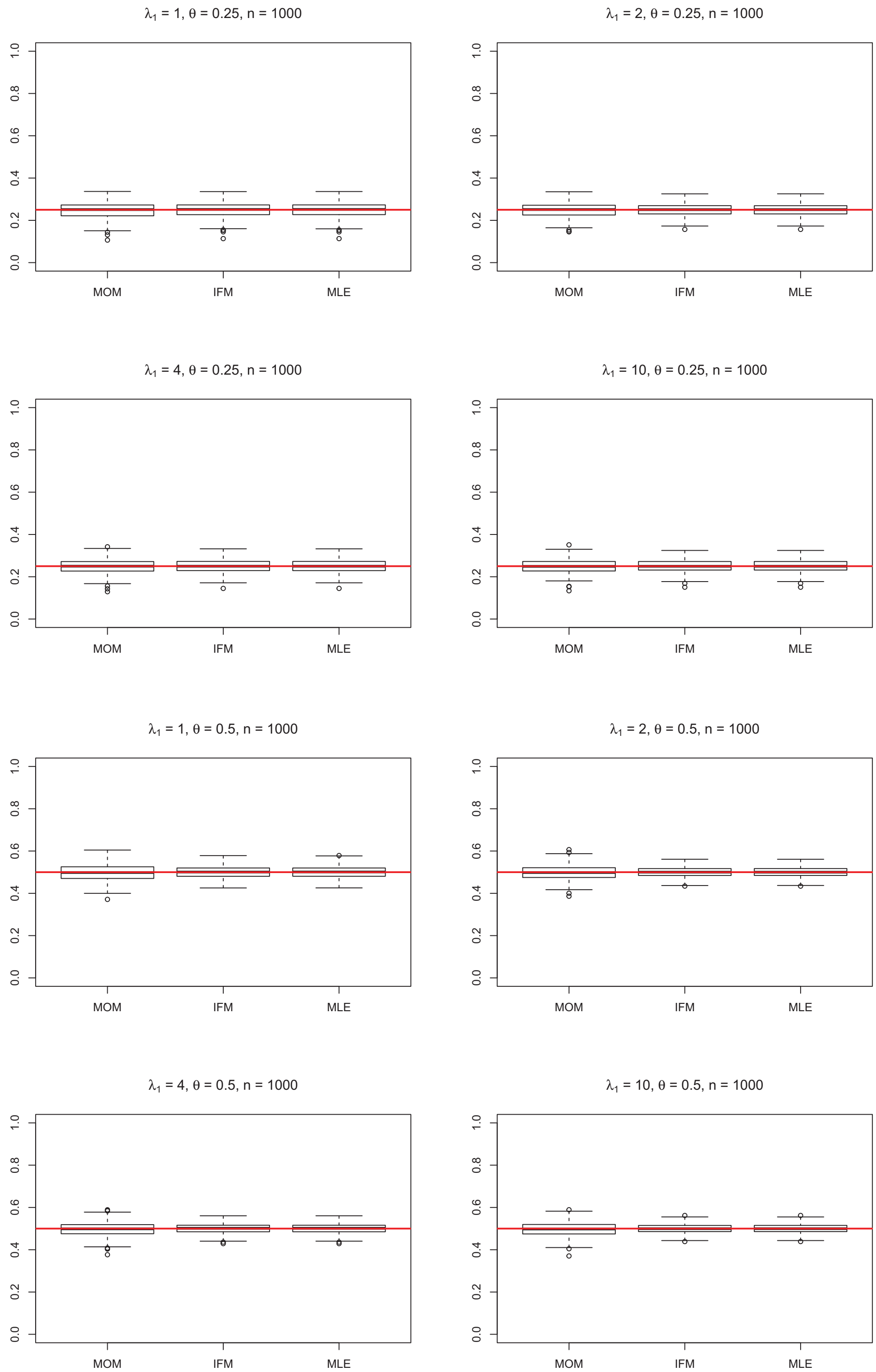

Figure 4.11: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

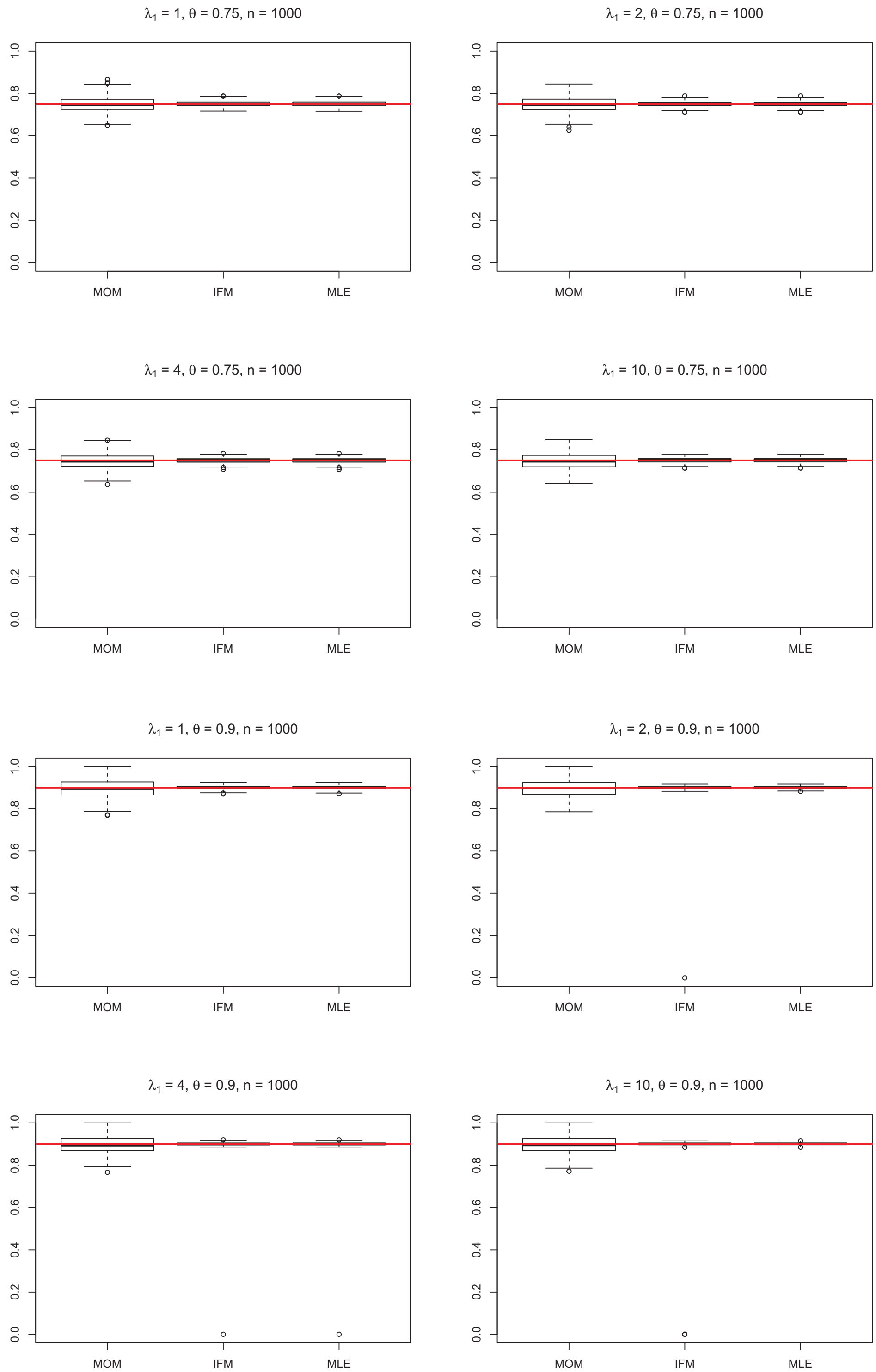

Figure 4.12: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

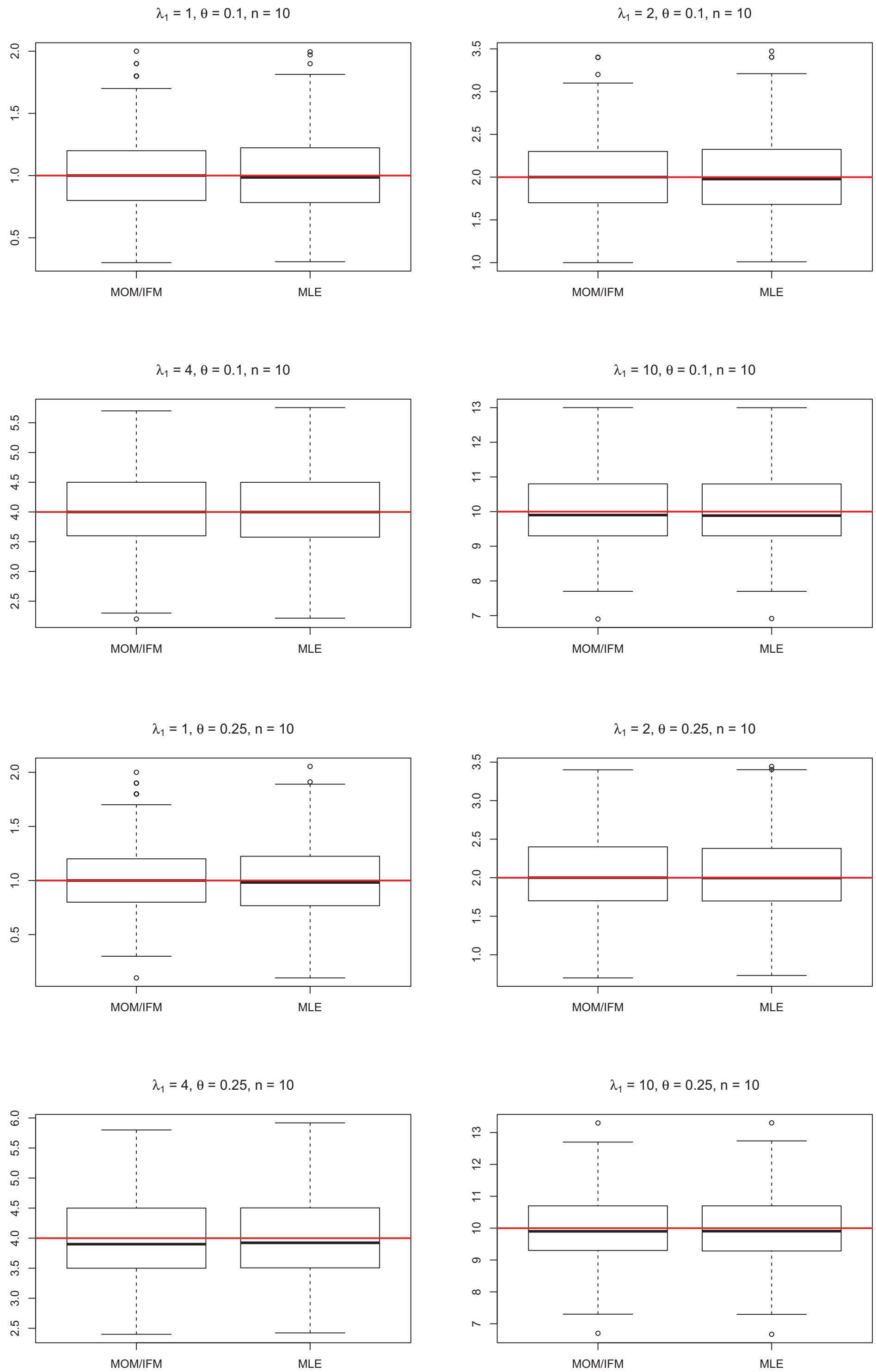

Figure 4.13: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

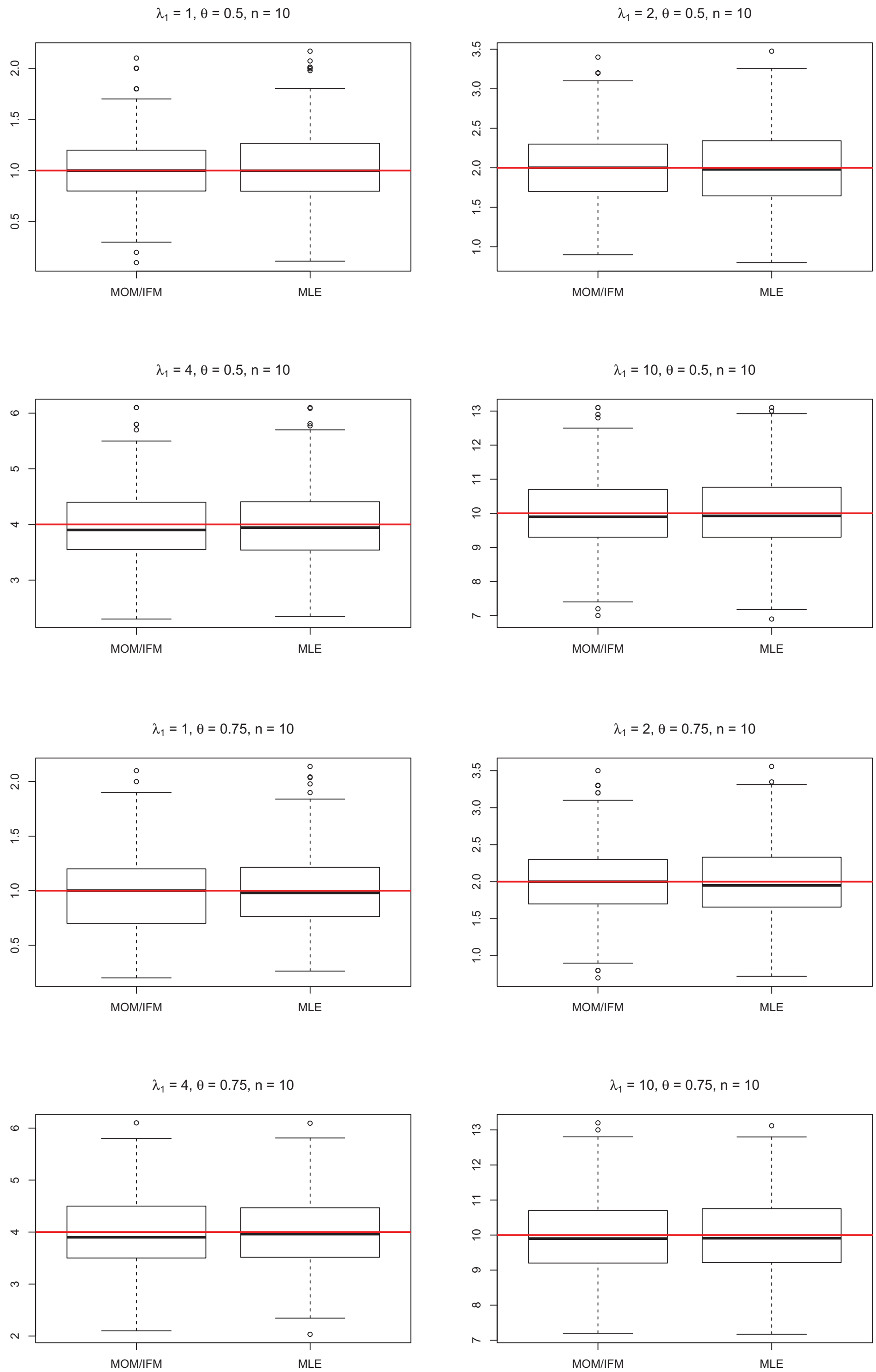

Figure 4.14: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

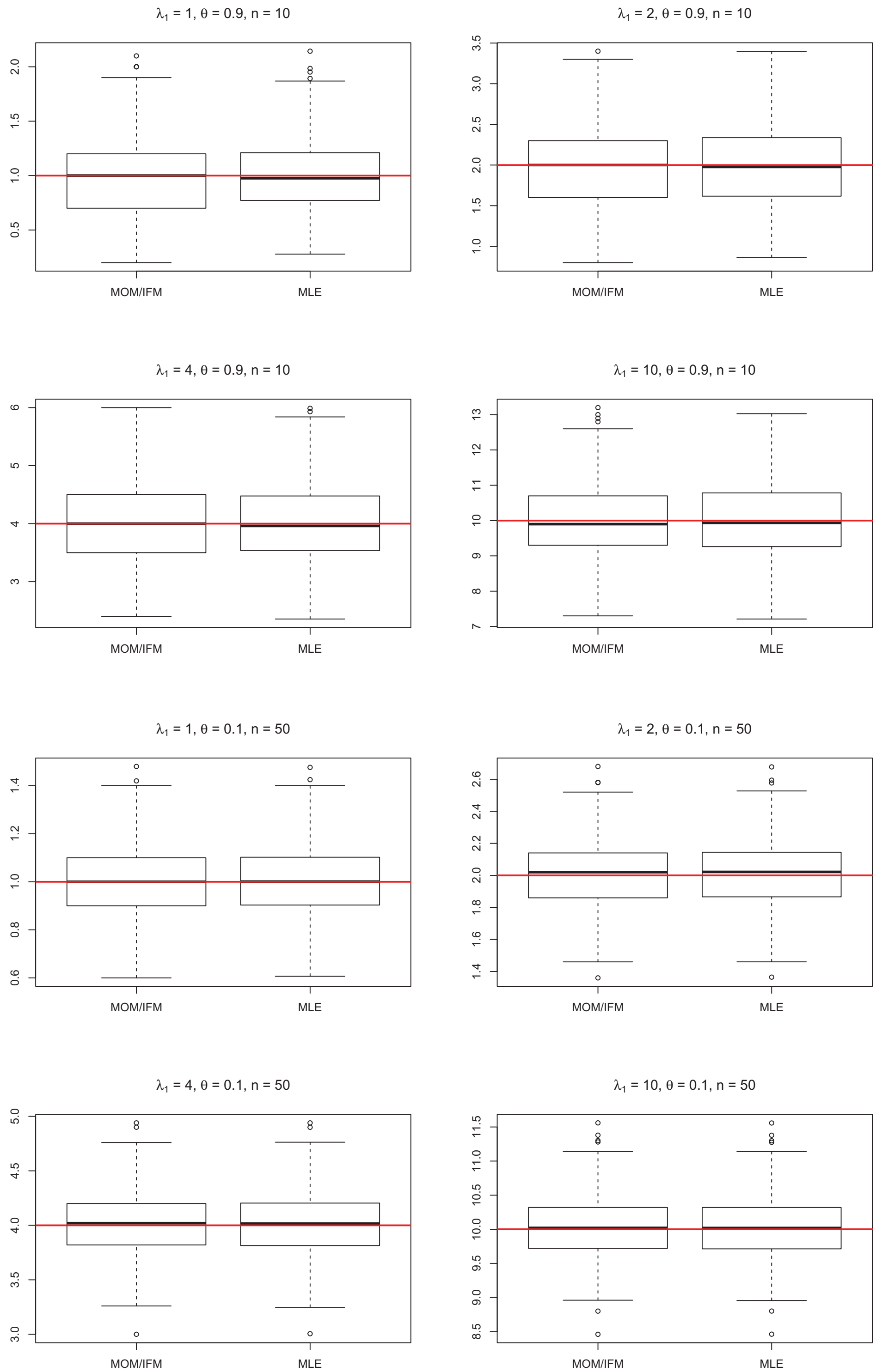

Figure 4.15: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

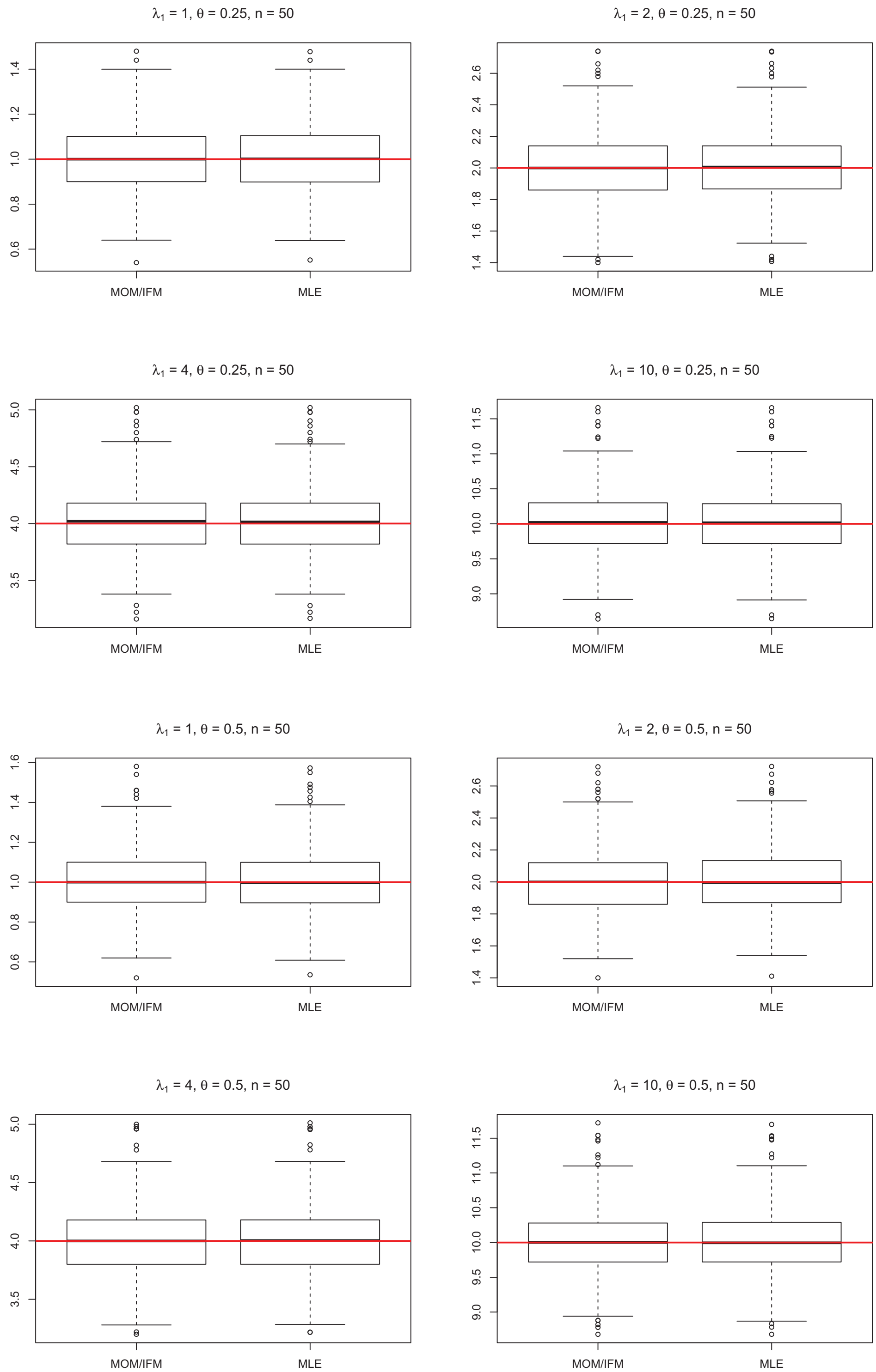

Figure 4.16: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

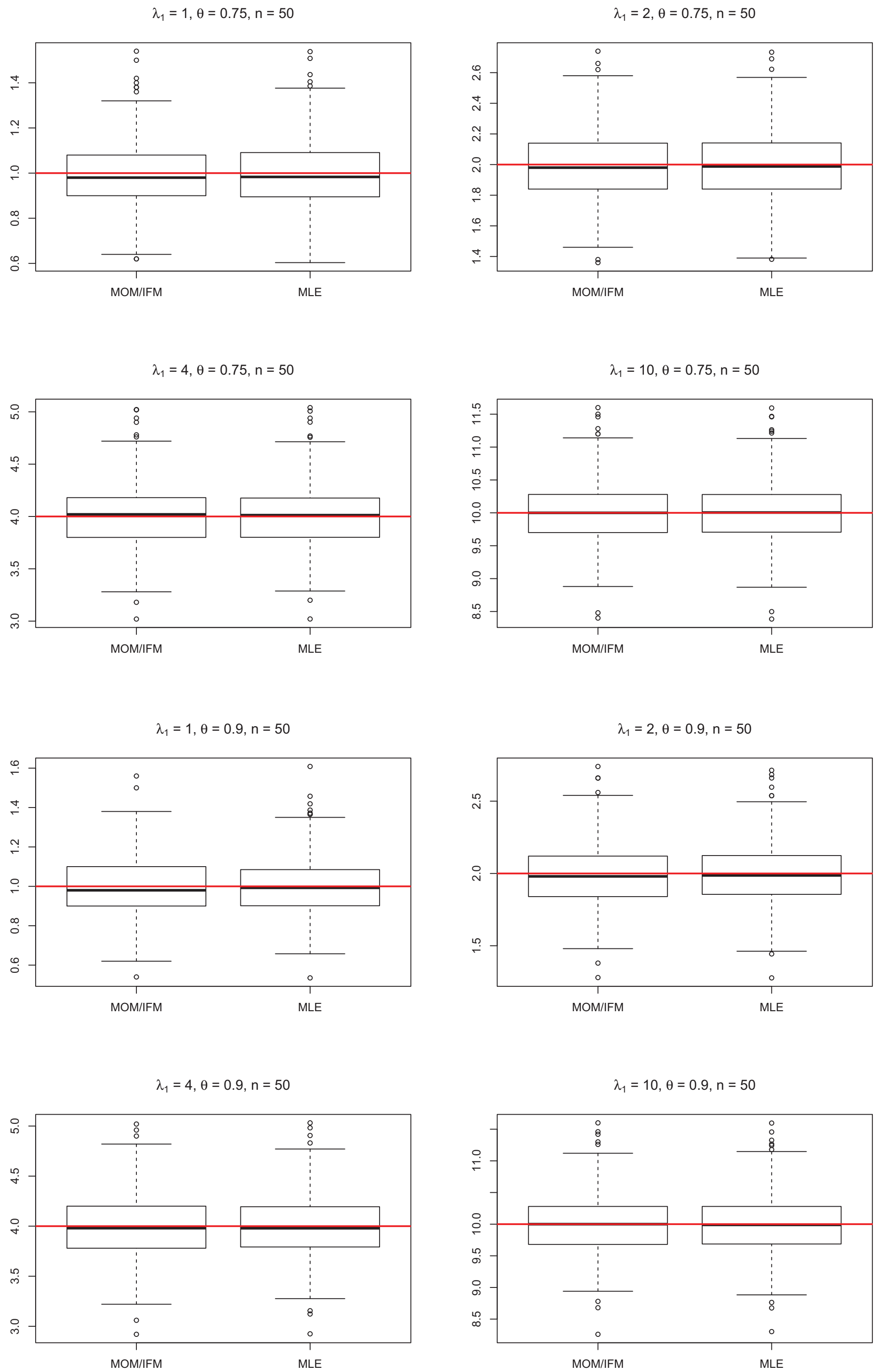

Figure 4.17: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

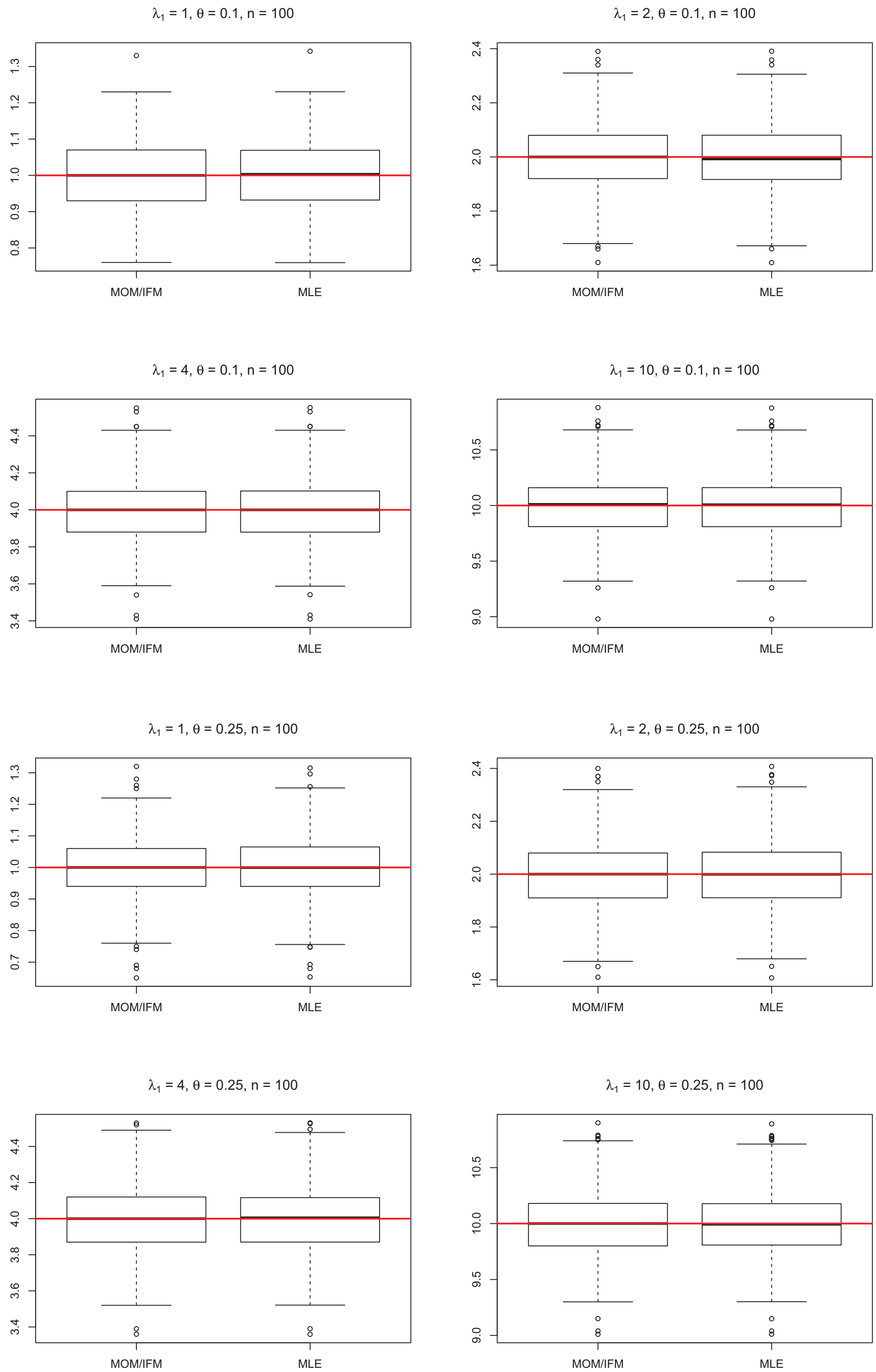

Figure 4.18: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

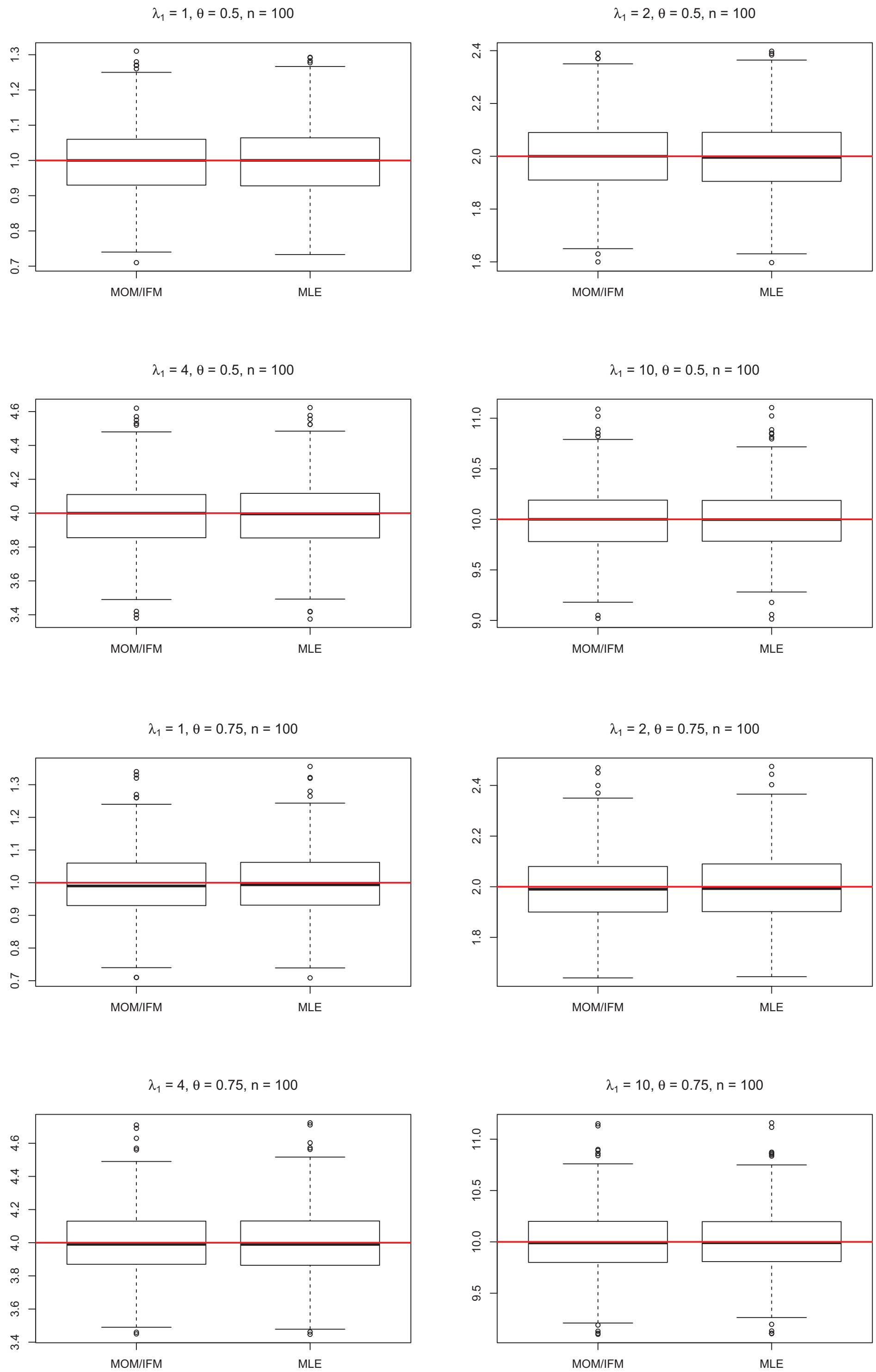

Figure 4.19: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

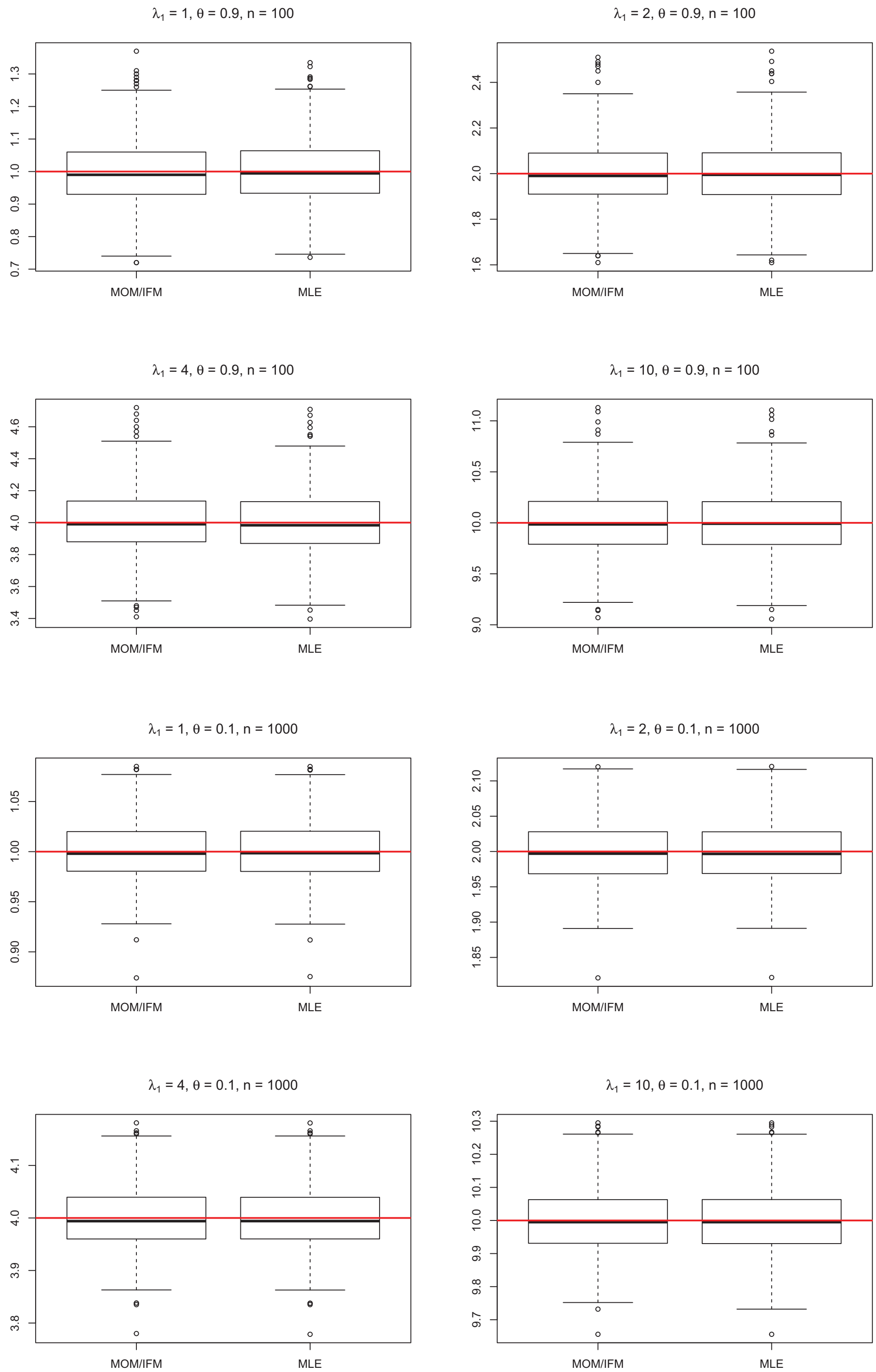

Figure 4.20: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

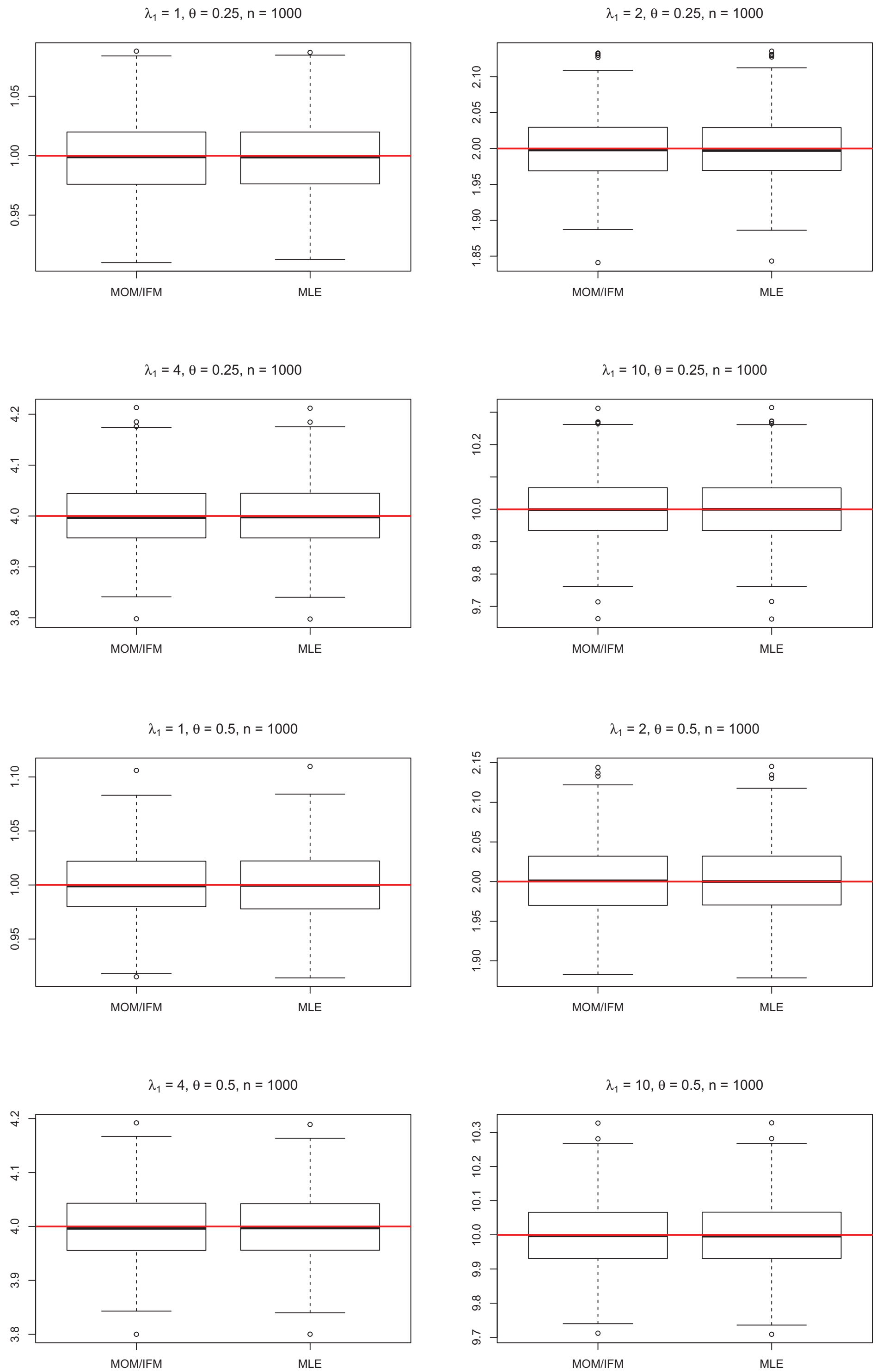

Figure 4.21: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

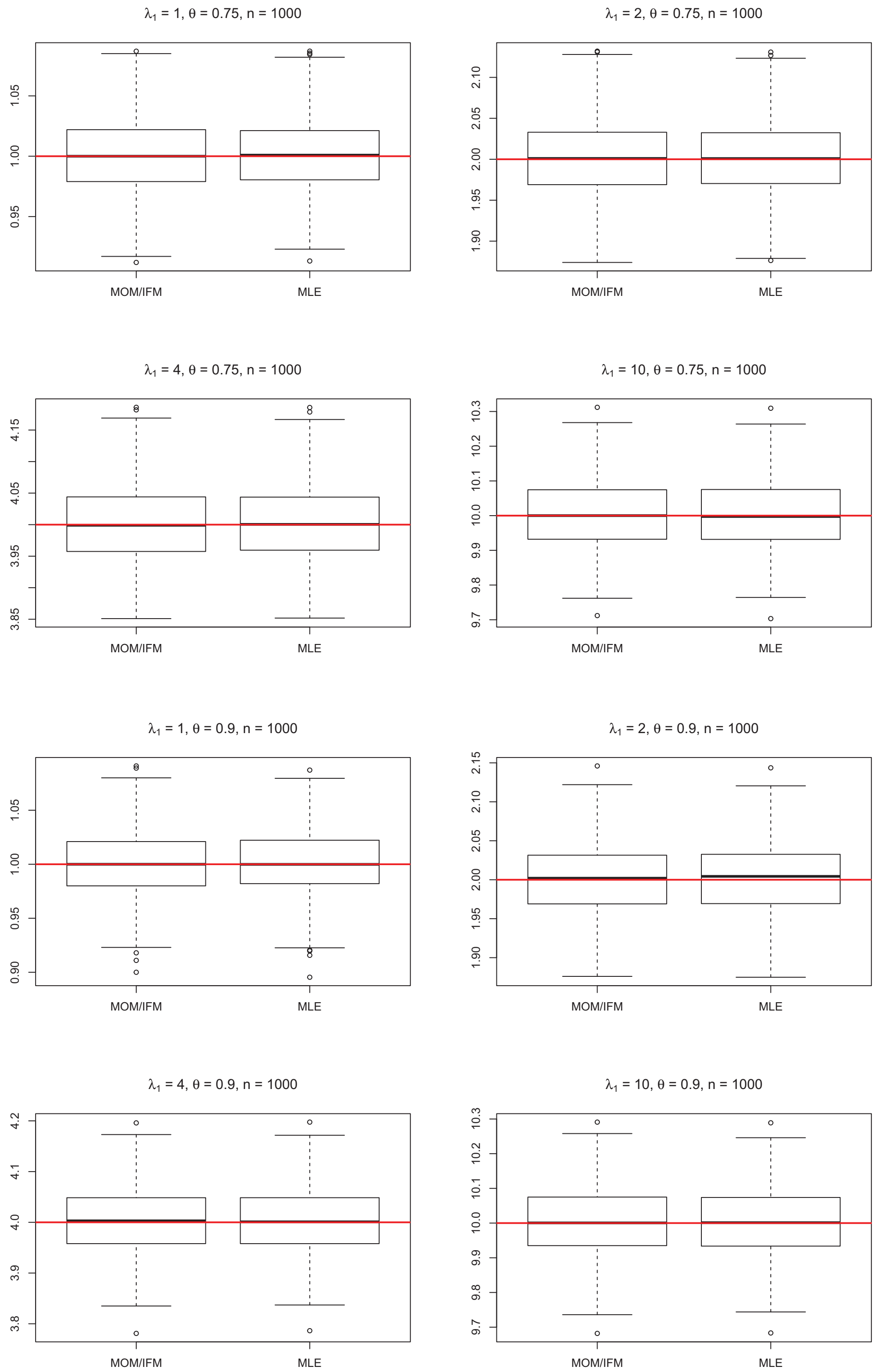

Figure 4.22: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

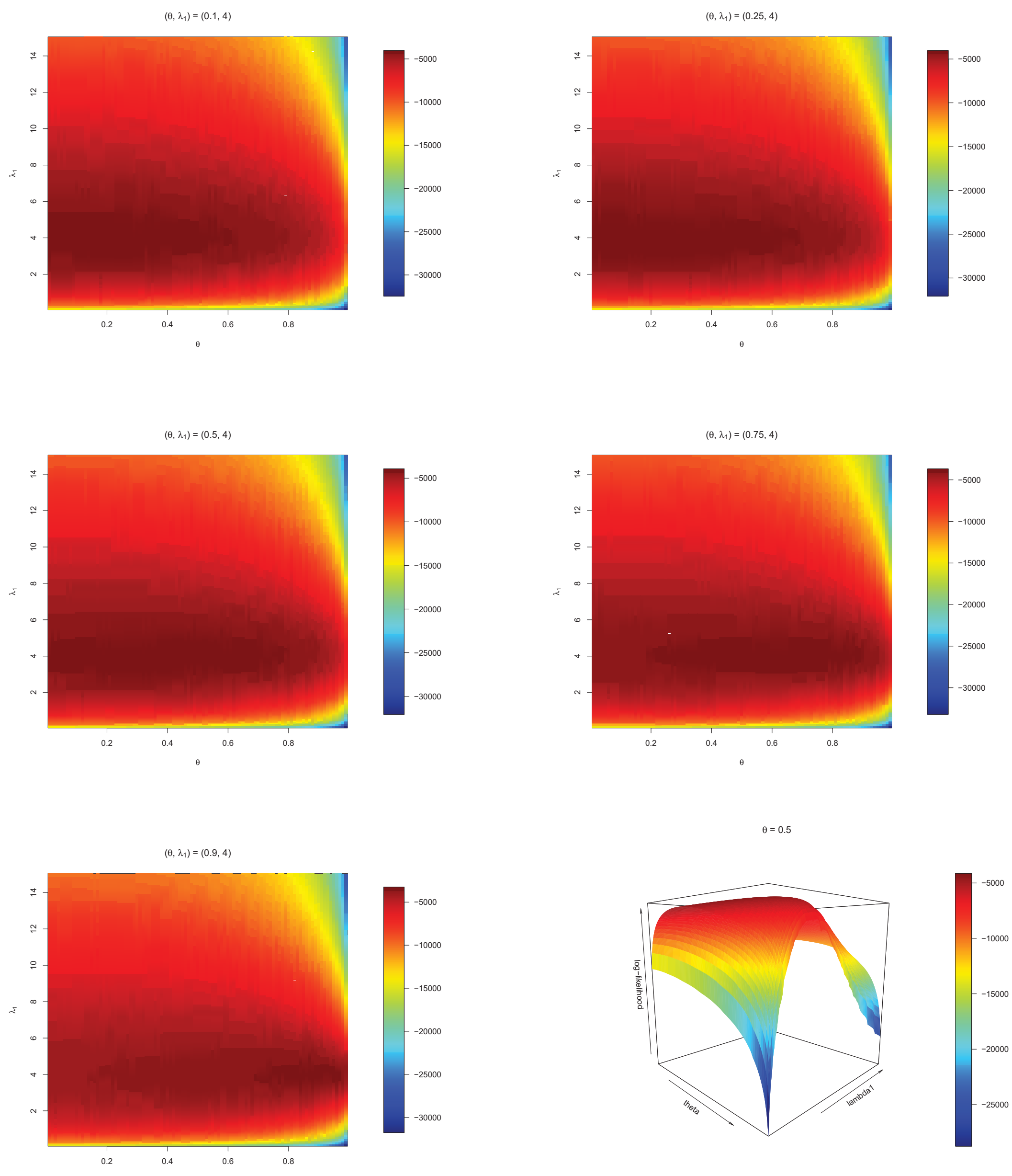

Figure 4.23: Log-likelihood plots for varying values of θ, with $(\lambda_1, \lambda_2) = (4, 5)$ and $n = 1000$.

4.5 Data illustration

In order to further explore the merits of the proposed bivariate Poisson family, the $\mathcal{BP}$ model was applied to environmental data pertaining to two US weather stations, namely La Guardia Airport in New York and Newark Liberty International Airport in New Jersey. The raw data are publicly available from the US National Centers for Environmental Information and consist of daily precipitation recorded at the aforementioned weather stations. The bivariate counts are then established as the number of days per year where the total rainfall recorded at the weather station exceeded a pre-specified threshold. For the La Guardia station, the threshold was taken to be the 95th percentile, specifically 18.54 mm. At the Newark station, the threshold was set as the 99th percentile given by 41.40 mm. The yearly number of exceedances at La Guardia is defined as the variable X_1 while that at Newark is defined as X_2. The data included observations from the year 1940 to 2017, for a total of 78 paired observations denoted (x_{i1}, x_{i2}) for $i \in \{1, \ldots, 78\}$.

In the proposed $\mathcal{BP}$ model, both margins are assumed to follow a univariate Poisson distribution. This assumption can easily be checked using standard statistical tests. In particular, a chi-square goodness-of-fit test yields p-values of 0.31 for the first component (La Guardia) and 0.88 for the second component (Newark). This was done in R using the `gofstat` function. The QQ-plots for both the data at La Guardia and Newark, shown in Figure 4.24, provide additional evidence of Poisson margins. Further theoretical justification for the marginal Poisson assumption stems from extreme-value theory. As detailed in e.g., Corollary 4.19 of Resnick (2008), the number of exceedances above a high threshold can be shown to be approximately Poisson distributed.

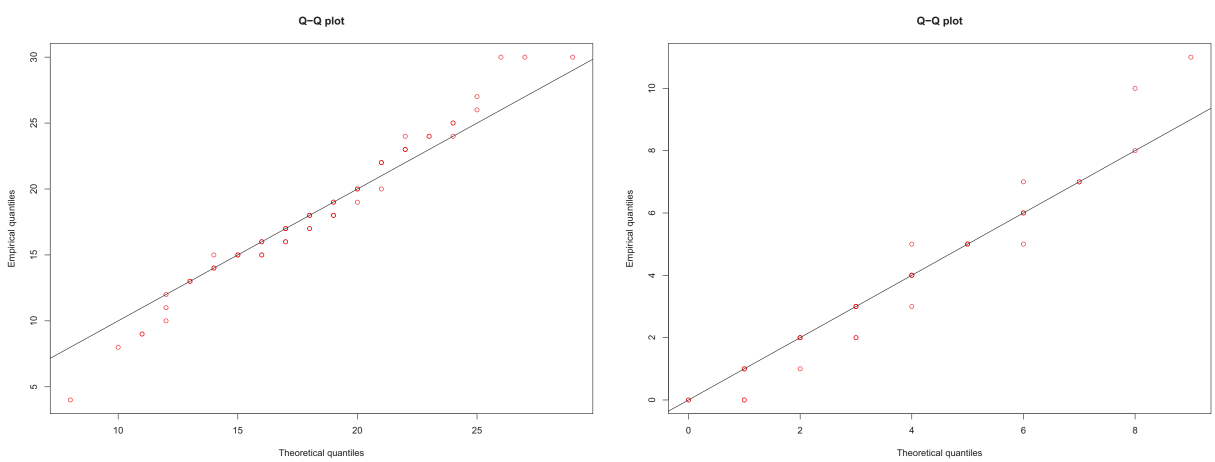

Figure 4.24: QQ-plots assessing the marginal Poisson assumption for variables X_1 (left) and X_2 (right) in the data illustration. Theoretical and empirical quantiles are on the x- and y-axis, respectively.

The counts (X_1, X_2) are derived from rainfall data recorded at two nearby weather stations. As a result of their spatial disposition, it is likely that X_1 and X_2 exhibit a strong association. Indeed, the sample covariance is $S_{12} = 4.22$, whereas the mean number of exceedances at the La Guardia station is $\bar{X}_1 = 17.99$ and that at the Newark station is $\bar{X}_2 = 3.64$. Using the sample means, $\bar{X}_1$ and $\bar{X}_2$, as the marginal parameter estimates, the implied sample correlation is then given by $R_{12} = S_{12}/\sqrt{\bar{X}_1\bar{X}_2} = 0.52$.

In the classical bivariate Poisson model, the covariance ξ must fall between 0 and $\min(\lambda_1, \lambda_2)$. As was outlined in Chapter 2, both the method of moments and maximum likelihood estimation procedures result in marginal parameter estimates equal to their respective sample means. Provided that $S_{12} > 0$, the method of moments estimator for the covariance parameter is $\tilde{\xi} = \min\{S_{12}, \min(\bar{X}_1, \bar{X}_2)\}$. For the data considered here, this results in $\tilde{\xi} = \min(\bar{X}_1, \bar{X}_2) = 3.64$ implying a correlation of $\tilde{R} = \tilde{\xi}/\sqrt{\bar{X}_1\bar{X}_2} = 0.45$. Maximum likelihood estimation requires the use of numerical techniques. In particular, using the `optim` function in R, the MLE $\hat{\xi}$ is obtained by maximizing the log-likelihood in the classical bivariate Poisson model subject to the constraint that $\xi \in [0, \min(\lambda_1, \lambda_2)]$, with (λ_1, λ_2) fixed at $(\bar{X}_1, \bar{X}_2)$, as the latter represent the MLEs of the marginal parameters. Using the method of moments estimate as a starting value (with slight adjustments when necessary so as to avoid issues at the boundary of the parameter space), this resulted in $\hat{\xi} = 2.64$ thus implying a correlation of $\hat{R} = \hat{\xi}/\sqrt{\bar{X}_1\bar{X}_2} = 0.33$.

Based on 1000 bootstrap replications, 95% confidence intervals for ξ resulting from the method of moments and maximum likelihood approaches were found to be $(1.44, 4.09)$ and $(1.23, 3.73)$, respectively. The method of moments approach yielded a 95% bootstrap confidence interval for the implied correlation of $(0.18, 0.47)$, while in the case of maximum likelihood estimation this interval was given by $(0.16, 0.45)$. Note that in both estimation approaches, neither the observed sample covariance nor the corresponding correlation is contained in their respective confidence interval.

The construction of the proposed bivariate Poisson model allows for greater flexibility in the dependence structure between positively correlated count data. For the rainfall data considered here, all three estimation approaches discussed in Section 4.3 were applied, the results from which are summarized in Table 4.1.

The numerical techniques used in the optimization for both the ML and IFM methods used the method of moments estimates as starting values. In order to avoid issues at the boundary of the parameter space, if the method of moments resulted in $\tilde{\theta} = 0$, the starting value was set to 0.01 while if $\tilde{\theta} = 1$ a starting value of 0.99 was used. For this particular data, the bootstrap replications never resulted in $\tilde{\theta} = 0$. Note that there were a total of six replications where the MM estimate was found to be 1. In each of these 6 occurrences, the sample covariance attained the upper bound $S_{12} = m_{\bar{X}_1,\bar{X}_2}(1)$, thus appropriately yielding $\tilde{\theta} = 1$. All of the computations were done in R; the

MM estimation used the `uniroot` function while both IFM and ML results were obtained using `optim`. Note that the IFM estimation results presented in Table 4.1 were established by optimizing the log-likelihood in the proposed $\mathcal{BP}$ model, holding (λ_1, λ_2) fixed at their respective marginal MLEs, namely $(\bar{X}_1, \bar{X}_2)$. The EM algorithm was not used to find $\check{\theta}$ as it tended to be quite slow to converge. Nonetheless, applying the IFM method with the EM algorithm yields a similar estimate for the dependence parameter, specifically, $\check{\theta} = 0.36$.

The results shown in Table 4.1 suggest that both likelihood-based methods (IFM and MLE) yield a much smaller estimate for θ and consequently imply weaker dependence. This is also the case in the classical bivariate Poisson model, where the maximum likelihood estimate for θ corresponded to a much lower correlation than that resulting from the method of moments approach. The method of moments estimator leads to a substantially wider confidence interval for both θ and the implied correlation as compared to the likelihood-based methods. This is perhaps an indicator of the poor performance of the moment-based estimator for the given sample.

Notice, however, that only the bootstrap confidence interval for ρ derived from the method of moments estimates contains the observed sample correlation $R_{12} = S_{12}/\sqrt{\bar{X}_1\bar{X}_2} = 0.52$. For this particular data, despite both the theoretical and empirical justification for Poisson margins, there is slight evidence of overdispersion. Indeed, the sample variance for X_1 and X_2 are $S_1^2 = 25.23$ and $S_2^2 = 4.60$, respectively, both of which are larger than the corresponding marginal sample means $\bar{X}_1 = 17.99$ and $\bar{X}_2 = 3.64$. Defining the sample correlation as $R_{12}^* = S_{12}/\sqrt{S_1^2 S_2^2}$ yields an observed value of 0.39. This version of the sample correlation is contained in the bootstrap confidence intervals for ρ implied by both the IFM and MLE methods.

It may be that the proposed bivariate Poisson model is not entirely appropriate for the rainfall data under question. However, these issues could be due to the small sample size as the data under question consist of only 78 pairs. Despite the issues with the estimation of the dependence parameter, the results for the marginal parameters were less problematic. As expected based on the simulation results, there is very little difference in the marginal parameter estimates from the various estimation approaches.

Note that different starting values for θ were tested for the IFM and MLE methods in the

Table 4.1: Estimation results for the proposed bivariate Poisson model. Estimates are given with 95% bootstrap confidence intervals in parentheses.

	λ_1	λ_2	θ	ρ
MM	$17.99\ (16.81, 19.15)$	$3.64\ (3.19, 4.14)$	$0.54\ (0.20, 0.90)$	$0.52\ (0.18, 0.89)$
IFM	$17.99\ (16.81, 19.15)$	$3.64\ (3.19, 4.14)$	$0.35\ (0.00, 0.50)$	$0.33\ (0.00, 0.48)$
MLE	$17.98\ (16.81, 19.15)$	$3.62\ (3.19, 4.14)$	$0.35\ (0.16, 0.52)$	$0.33\ (0.14, 0.50)$

Table 4.2: Estimation results using different starting values for θ.

Starting Value	IFM	MLE
0.10	0.35	0.35
0.25	0.35	0.35
0.50	0.35	0.35
0.75	0.00	0.35
0.90	0.00	0.35

$\mathcal{BP}(\Lambda, \theta)$ model, the results for which are summarized in Table 4.2. Ignoring the poor results for the starting values of 0.75 and 0.90 in the IFM approach, the likelihood-based methods seem to be more or less unaffected by the starting value for θ in the optimization procedure. This was also observed in the simulation studies.

Recall that the main motivation in defining the proposed bivariate Poisson model was to address the shortcomings of the classical common shock model. Specifically, the classical model cannot accurately account for strong degrees of dependence. The proposed bivariate Poisson model, in contrast, can span the full spectrum of possible correlation, ranging from 0 to $\rho_{\max}$. In this particular data illustration, the margins exhibit a moderate degree of dependence, with the likelihood-based methods estimating θ at 0.35 and the MM approach yielding 0.54. In comparing the implied correlations in the proposed and classical models, it is clear that there are only slight differences between the two. Nonetheless, the two models do have different implications in terms of the resulting probability distribution.

In the context of these data, one may be interested in estimating the probability of certain extremal events. For example, consider the event $\{X_1 > 30, X_2 > 30\}$, i.e., the event that the total daily rainfall amounts at both the La Guardia and Newark stations exceed their respective thresholds (18.54 mm and 41.40 mm) for more than 30 days in the year, i.e., roughly one month. Under the classical bivariate Poisson model, the probability associated with this event is estimated at roughly 0.330% while in the proposed model the probability is approximately 0.328%. Note that both computations were done using the MLEs obtained in the respective models. Seemingly, the classical model overestimates the probability of this extremal event, although the difference is quite small.

As another example, consider the event $\{X_1 > 30, X_2 > 7\}$. Under the classical model, the probability associated with this event is 3.52%, whereas in the proposed model this probability becomes 3.43% (where both computations are based on the MLEs.) Again, the classical model overestimates the probability of this event. Note that based on the sample, the empirical probability of both of the aforementioned events is 0.

5

Inducing Negative Dependence in the Proposed Bivariate Poisson Model

5.1 Introduction

By construction, the proposed $\mathcal{BP}$ model is only appropriate for positively correlated Poisson random variables since, as was previously shown, for $(X_1, X_2) \sim \mathcal{BP}(\Lambda, \theta)$, $\text{corr}(X_1, X_2) \in [0, \rho_{\max}(\lambda_1, \lambda_2)]$. Nonetheless, the stochastic representation as given in (4.4) allows for a natural extension to the case of negative dependence. Consider a similar set-up wherein the margins (X_1, X_2) are decomposed into an independent pair (Y_1, Y_2). However, now let (Z_1, Z_2) denote a counter-monotonic shock. More specifically, suppose that Y_1 and Y_2 are mutually independent Poisson random variables with rates $(1-\theta)\lambda_1$ and $(1-\theta)\lambda_2$, respectively. Further assume that (Y_1, Y_2) is independent of the counter-monotonic pair (Z_1, Z_2), which can be written as

$$(Z_1, Z_2) = (G^{-1}_{\theta\lambda_1}(U), G^{-1}_{\theta\lambda_2}(1-U)) \tag{5.1}$$

for $U \sim \mathcal{U}(0,1)$. Then

$$(X_1, X_2) = (Y_1 + Z_1, Y_2 + Z_2) \tag{5.2}$$

follows a bivariate Poisson counter-monotonic shock model, which will be denoted as $(X_1, X_2) \sim \mathcal{BP}^-(\Lambda, \theta)$. Similarly to the model for positive dependence, the parameter constraints in the proposed $\mathcal{BP}^-(\Lambda, \theta)$ model are $\lambda_1, \lambda_2 \in (0, \infty)$ and $\theta \in [0,1]$. Note that this proposed bivariate Poisson model for negatively correlated margins is similar to that proposed by Cuenin et al. (2016); see Chapter 2 for more details.

Many of the properties in the $\mathcal{BP}^-$ family parallel those in the proposed bivariate Poisson model for positive dependence. Similarly to what was shown in the $\mathcal{BP}$ model, θ can be regarded as a dependence parameter in that it regulates the strength of the association between the margins. In particular, setting $\theta = 0$ yields independence while $\theta = 1$ results in perfect negative dependence.

In the latter case, (X_1, X_2) represent a counter-monotonic pair with correlation equal to the lower bound $\rho_{\min}(\lambda_1, \lambda_2)$ given in (4.1). In this sense, the proposed $\mathcal{BP}^-(\Lambda, \theta)$ model provides a fully flexible bivariate Poisson model for negative dependence. The rest of this chapter is dedicated to exploring the distributional properties as well as estimation techniques for the proposed $\mathcal{BP}^-(\Lambda, \theta)$ model. As a result of the similarity in the stochastic representations in (4.4) and (5.2), much of what will be presented in this chapter directly parallels that of Chapter 4. As such, the present chapter will be kept relatively brief in comparison to the previous.

5.2 The proposed model for negative dependence

The construction of the proposed model for negative dependence, as given in (5.2), allows to express both the probability mass function and distribution function in an analogous manner to that of the $\mathcal{BP}$ model. Let $c^-_{\Lambda,\theta}$ denote the probability mass function of the counter-monotonic pair $(Z_1, Z_2) \sim \mathcal{W}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\}$. By (5.1), it follows that for all $z_1, z_2 \in \mathbb{N}$,

$$\begin{aligned}
c^-_{\Lambda,\theta}(z_1, z_2) &= \Pr\{G^{-1}_{\theta\lambda_1}(U) = z_1, G^{-1}_{\theta\lambda_2}(1-U) = z_2\} \\
&= \Pr\left\{G_{\theta\lambda_1}(z_1 - 1) < U \leqslant G^{-1}_{\theta\lambda_1}(z_1), G^{-1}_{\theta\lambda_2}(z_2 - 1) < 1 - U \leqslant G_{\theta\lambda_2}(z_2)\right\} \\
&= \Pr\{G_{\theta\lambda_1}(z_1 - 1) < U \leqslant G^{-1}_{\theta\lambda_1}(z_1), \bar{G}_{\theta\lambda_2}(z_2) \leqslant U < \bar{G}_{\theta\lambda_2}(z_2 - 1)\} \\
&= \Pr\left[\max\{G_{\theta\lambda_1}(z_1 - 1), \bar{G}_{\theta\lambda_2}(z_2)\right\} \leqslant U \leqslant \min\left\{G_{\theta\lambda_1}(z_1), \bar{G}_{\theta\lambda_2}(z_2 - 1)\}\right] \\
&= \left[\min\{G_{\theta\lambda_1}(z_1), \bar{G}_{\theta\lambda_2}(z_2 - 1)\} - \max\{G_{\theta\lambda_1}(z_1 - 1), \bar{G}_{\theta\lambda_2}(z_2)\}\right]_+
\end{aligned}$$

where again $[x]_+ = x\mathbf{1}(x \geqslant 0)$. As was shown in Chapter 3, counter-monotonicity implies that the pair (Z_1, Z_2) has distribution function given by the lower Fréchet–Hoeffding bound, i.e.,

$$\Pr(Z_1 \leqslant z_1, Z_2 \leqslant z_2) = \max\{0, G_{\theta\lambda_1}(z_1) + G_{\theta\lambda_2}(z_2) - 1\}.$$

Accordingly, the probability mass function in the proposed negative dependence model, denoted as $f^-_{\Lambda,\theta}$, can be written as

$$f^-_{\Lambda,\theta}(x_1, x_2) = \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) c^-_{\Lambda,\theta}(z_1, z_2),$$

for $x_1, x_2 \in \mathbb{N}$. This leads to the following equivalent forms for the distribution function $F^-_{\Lambda,\theta}$:

$$\begin{aligned}
F^-_{\Lambda,\theta}(x_1, x_2) &= \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) \max\{0, G_{\theta\lambda_1}(z_1) + G_{\theta\lambda_2}(z_2) - 1\} \\
&= \sum_{z_1=0}^{x_1} \sum_{z_2=0}^{x_2} G_{(1-\theta)\lambda_1}(x_1 - z_1) G_{(1-\theta)\lambda_2}(x_2 - z_2) c^-_{\Lambda,\theta}(z_1, z_2).
\end{aligned}$$

5.2.1 PQD ordering

The concept of positive quadrant dependence (PQD) ordering was introduced in Chapter 4 as a means of quantifying the nature of the dependence. Recall that for an arbitrary joint distribution function F the notion of PQD ordering ensures that

$$F_L \prec_{PQD} F \prec_{PQD} F_U, \tag{5.3}$$

where F_L and F_U represent the lower and upper Fréchet–Hoeffding boundary distributions, respectively, and F, F_L and F_U all have the same marginal distributions. As was the case in the proposed bivariate Poisson model with positively correlated margins, the strength of the dependence in the $\mathcal{BP}^-(\Lambda, \theta)$ family is regulated by θ in terms of PQD ordering. This is formally stated in the following lemma, analogous to Lemma 4.1.

Lemma 5.1 ***PQD ordering in the*** $\mathcal{BP}^-(\Lambda, \theta)$ ***family***
Let $(X_1, X_2) \sim \mathcal{BP}^-(\Lambda, \theta)$ *and* $(X_1', X_2') \sim \mathcal{BP}^-(\Lambda, \theta')$. *Then* $\theta < \theta' \Rightarrow (X_1', X_2') \prec_{PQD} (X_1, X_2)$.

Proof. The proof is analogous to that of Lemma 4.1. Fix λ_1, λ_2 and $\theta < \theta'$. By definition of the $\mathcal{BP}^-(\Lambda, \theta)$ family, (X_1, X_2) can be written as

$$X_1 = Y_1 + Z_1, \quad X_2 = Y_2 + Z_2,$$

with $Y_1 \sim \mathcal{P}\{(1-\theta)\lambda_1\}$ independent of $Y_2 \sim \mathcal{P}\{(1-\theta)\lambda_2\}$ and (Y_1, Y_2) independent of the counter-monotonic pair $(Z_1, Z_2) \sim \mathcal{W}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\}$. Similarly, write

$$X_1' = Y_1' + Z_1', \quad X_2' = Y_2' + Z_2',$$

with $Y_1' \sim \mathcal{P}\{(1-\theta')\lambda_1\}$ independent of $Y_2' \sim \mathcal{P}\{(1-\theta')\lambda_2\}$ and (Y_1', Y_2') independent of $(Z_1', Z_2') \sim \mathcal{W}\{\mathcal{P}(\theta'\lambda_1), \mathcal{P}(\theta'\lambda_2)\}$. The pair (Y_1, Y_2) can be expressed as

$$Y_1 = T_1 + S_1, \quad Y_2 = T_2 + S_2,$$

where $T_1 \sim \mathcal{P}\{(1-\theta')\lambda_1\}$, $T_2 \sim \mathcal{P}\{(1-\theta')\lambda_2\}$, $S_1 \sim \mathcal{P}\{(\theta'-\theta)\lambda_1\}$, $S_2 \sim \mathcal{P}\{(\theta'-\theta)\lambda_2\}$ are mutually independent. Since (T_1, T_2) and (Y_1', Y_2') consist of independent pairs with identical marginal distributions, it follows that $(Y_1', Y_2') \prec_{PQD} (T_1, T_2)$. The pairs $(S_1 + Z_1, S_2 + Z_2)$ and (Z_1', Z_2') have identical marginal distributions, respectively given by $\mathcal{P}(\theta'\lambda_1)$ and $\mathcal{P}(\theta'\lambda_2)$ for the first and second components. By definition, (Z_1', Z_2') is counter-monotonic and thus (5.3) ensures that $(Z_1', Z_2') \prec_{PQD} (S_1 + Z_1, S_2 + Z_2)$. By the closure properties of the PQD ordering, as given in Theorem 4.1, it follows that

$$(Y_1' + Z_1', Y_2' + Z_2') \prec_{PQD} (T_1 + S_1 + Z_1, T_2 + S_2 + Z_2).$$

By construction, the above is equivalent to the desired result, i.e., $(X_1', X_2') \prec_{PQD} (X_1, X_2)$. This concludes the argument. □

5.2.2 Moments and measures of dependence

As was explored in Chapter 4, the PQD ordering implies an ordering in certain measures of dependence. Accordingly, Lemma 5.1 ensures that for fixed marginal parameters λ_1 and λ_2, both the covariance and the correlation in the proposed $\mathcal{BP}^-(\Lambda, \theta)$ model are decreasing functions of θ.

By construction, $\theta = 0$ corresponds to independence. As θ increases to 1, the resulting distribution $F^-_{\Lambda,\theta}$ reaches the Fréchet–Hoeffding lower bound F_L, in which case the implied correlation coincides with $\rho_{\min}$, where the latter denotes the correlation lower bound given in (4.1). This will be shown explicitly using the probability generating function in the $\mathcal{BP}^-(\Lambda, \theta)$ model, which is derived in the following proposition.

Proposition 5.1 *The pair* $(X_1, X_2) \sim \mathcal{BP}^-(\Lambda, \theta)$ *has probability generating function given by*

$$\mathrm{E}(s_1^{X_1} s_2^{X_2}) = \exp\left\{(1-\theta)\lambda_1(s_1 - 1) + (1-\theta)\lambda_2(s_2 - 1)\right\} \varrho^-_{\Lambda,\theta}(s_1, s_2),$$

where

$$\varrho^-_{\Lambda,\theta}(s_1, s_2) = 1 + (s_1 - 1)\sum_{i=0}^{\infty} s_1^i \bar{G}_{\theta\lambda_1}(i) + (s_2 - 1)\sum_{j=0}^{\infty} s_2^j \bar{G}_{\theta\lambda_2}(j) + (s_1 - 1)(s_2 - 1)\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \max\{0, \bar{G}_{\theta\lambda_1}(i) + \bar{G}_{\theta\lambda_2}(j) - 1\}.$$

Proof. The construction given in (5.2) allows to breakdown the probability generating fucntion as

$$\mathrm{E}(s_1^{X_1} s_2^{X_2}) = \mathrm{E}(s_1^{Y_1})\mathrm{E}(s_2^{Y_2})\mathrm{E}(s_1^{Z_1} s_2^{Z_2}).$$

Letting $\varrho^-_{\Lambda,\theta}(s_1, s_2)$ denote the probability generating function of the counter-monotonic pair (Z_1, Z_2), the above further simplifies to

$$\mathrm{E}(s_1^{X_1} s_2^{X_2}) = \exp\left\{(1-\theta)\lambda_1(s_1 - 1) + (1-\theta)\lambda_2(s_2 - 1)\right\} \varrho^-_{\Lambda,\theta}(s_1, s_2).$$

For $(Z_1, Z_2) \sim \mathcal{W}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\}$, it follows that

$$\begin{aligned}
\Pr(Z_1 \geqslant z_1, Z_2 \geqslant z_2) &= \Pr\{G^{-1}_{\theta\lambda_1}(U) \geqslant z_1, G^{-1}_{\theta\lambda_2}(1-U) \geqslant z_2\} \\
&= \Pr\{U > G_{\theta\lambda_1}(z_1 - 1), 1 - U > G_{\theta\lambda_2}(z_2 - 1)\} \\
&= \Pr\{G_{\theta\lambda_1}(z_1 - 1) < U < \bar{G}_{\theta\lambda_2}(z_2 - 1)\} \\
&= \max\{0, \bar{G}_{\theta\lambda_2}(z_2 - 1) - G_{\theta\lambda_1}(z_1 - 1)\} \\
&= \max\{0, \bar{G}_{\theta\lambda_1}(z_1 - 1) + \bar{G}_{\theta\lambda_2}(z_2 - 1) - 1\}.
\end{aligned}$$

Writing the probability $\Pr(Z_1 = z_1, Z_2 = z_2)$ as the difference

$$\Pr(Z_1 \geqslant z_1, Z_2 \geqslant z_2) - \Pr(Z_1 \geqslant z_1 + 1, Z_2 \geqslant z_2) \\
- \Pr(Z_1 \geqslant z_1, Z_2 \geqslant z_2 + 1) + \Pr(Z_1 \geqslant z_1 + 1, Z_2 \geqslant z_2 + 1)$$

allows for an alternative formulation of the probability mass function of (Z_1, Z_2) given by

$$\max\{0, \bar{G}_{\theta\lambda_1}(z_1 - 1) + \bar{G}_{\theta\lambda_2}(z_2 - 1) - 1\} - \max\{0, \bar{G}_{\theta\lambda_1}(z_1) + \bar{G}_{\theta\lambda_2}(z_2 - 1) - 1\} \\
- \max\{0, \bar{G}_{\theta\lambda_1}(z_1 - 1) + \bar{G}_{\theta\lambda_2}(z_2) - 1\} + \max\{0, \bar{G}_{\theta\lambda_1}(z_1) + \bar{G}_{\theta\lambda_2}(z_2) - 1\}.$$

Working with the above form, the probability generating function $\varrho^{-}_{\Lambda,\theta}(s_1, s_2)$ can be expressed as

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \big[\max\{0, \bar{G}_{\theta\lambda_1}(i-1) + \bar{G}_{\theta\lambda_2}(j-1) - 1\} - \max\{0, \bar{G}_{\theta\lambda_1}(i) + \bar{G}_{\theta\lambda_2}(j-1) - 1\} \\
- \max\{0, \bar{G}_{\theta\lambda_1}(i-1) + \bar{G}_{\theta\lambda_2}(j) - 1\} + \max\{0, \bar{G}_{\theta\lambda_1}(i) + \bar{G}_{\theta\lambda_2}(j) - 1\}\big].$$

Noting that $\bar{G}_\lambda(-1) = 1$, the above can be rewritten as

$$1 + \sum_{i=1}^{\infty} s_1^i \bar{G}_{\theta\lambda_1}(i-1) + \sum_{j=1}^{\infty} s_2^j \bar{G}_{\theta\lambda_2}(j-1) + \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} s_1^i s_2^j \max\{0, \bar{G}_{\theta\lambda_1}(i-1) + \bar{G}_{\theta\lambda_2}(j-1) - 1\} \\
- \sum_{i=0}^{\infty} s_1^i \bar{G}_{\theta\lambda_1}(i) - \sum_{i=0}^{\infty}\sum_{j=1}^{\infty} s_1^i s_2^j \max\{0, \bar{G}_{\theta\lambda_1}(i) + \bar{G}_{\theta\lambda_2}(j-1) - 1\} \\
- \sum_{j=0}^{\infty} s_2^j \bar{G}_{\theta\lambda_2}(j) - \sum_{i=1}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \max\{0, \bar{G}_{\theta\lambda_1}(i-1) + \bar{G}_{\theta\lambda_2}(j) - 1\} \\
+ \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^i s_2^j \max\{0, \bar{G}_{\theta\lambda_1}(i) + \bar{G}_{\theta\lambda_2}(j) - 1\}.$$

Manipulating the indices then yields

$$1+\sum_{i=0}^{\infty} s_1^{i+1}\bar{G}_{\theta\lambda_1}(i)+\sum_{j=0}^{\infty} s_2^{j+1}\bar{G}_{\theta\lambda_2}(j)+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^{i+1}s_2^{j+1}\max\{0,\bar{G}_{\theta\lambda_1}(i)+\bar{G}_{\theta\lambda_2}(j)-1\}$$
$$-\sum_{i=0}^{\infty} s_1^{i}\bar{G}_{\theta\lambda_1}(i)-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^{i}s_2^{j+1}\max\{0,\bar{G}_{\theta\lambda_1}(i)+\bar{G}_{\theta\lambda_2}(j)-1\}$$
$$-\sum_{j=0}^{\infty} s_2^{j}\bar{G}_{\theta\lambda_2}(j)-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^{i+1}s_2^{j}\max\{0,\bar{G}_{\theta\lambda_1}(i)+\bar{G}_{\theta\lambda_2}(j)-1\}$$
$$+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} s_1^{i}s_2^{j}\max\{0,\bar{G}_{\theta\lambda_1}(i)+\bar{G}_{\theta\lambda_2}(j)-1\}.$$

Upon further simplifications, the following is obtained:

$$1+\sum_{i=0}^{\infty}(s_1^{i+1}-s_1^{i})\bar{G}_{\theta\lambda_1}(i)+\sum_{j=0}^{\infty}(s_2^{j+1}-s_2^{j})\bar{G}_{\theta\lambda_2}(j)$$
$$+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(s_1^{i+1}s_2^{j+1}-s_1^{i}s_2^{j+1}-s_1^{i+1}s_2^{j}+s_1^{i}s_2^{j})\max\{0,\bar{G}_{\theta\lambda_1}(i)+\bar{G}_{\theta\lambda_2}(j)-1\},$$

which factors to the desired result. □

It is simple to obtain an expression for the covariance in the proposed counter-monotonic shock model from the probability generating function. Let $w_{\lambda_1,\lambda_2}(\theta)$ denote the covariance for a pair $(X_1,X_2)\sim\mathcal{BP}^{-}(\Lambda,\theta)$. Since $\text{cov}(X_1,X_2)=\text{cov}(Z_1,Z_2)$, it follows that

$$w_{\lambda_1,\lambda_2}(\theta)=\text{E}(Z_1Z_2)-\theta^2\lambda_1\lambda_2.$$

Working with the probability generating function of the counter-monotonic pair (Z_1,Z_2), it can be shown that

$$\text{E}(Z_1Z_2)=\frac{\partial^2}{\partial s_1\partial s_2}\,\varrho^{-}_{\Lambda,\theta}(s_1,s_2)\Big|_{s_1=s_2=1}=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\max\{0,\bar{G}_{\theta\lambda_1}(i)+\bar{G}_{\theta\lambda_2}(j)-1\}.$$

For any $a\in\mathbb{R}$, it is easy to verify that $\max(0,a)=-\min(0,-a)$. Thus, the mixed moment above can be rewritten as

$$\text{E}(Z_1Z_2)=-\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\min\{0,G_{\theta\lambda_1}(i)+G_{\theta\lambda_2}(j)-1\}.$$

From this expression it follows that the covariance in the proposed $\mathcal{BP}^{-}(\Lambda, \theta)$ family is given by

$$w_{\lambda_1,\lambda_2}(\theta) = -\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \min\{0, G_{\theta\lambda_1}(i) + G_{\theta\lambda_2}(j) - 1\} - \theta^2\lambda_1\lambda_2$$

with corresponding correlation

$$\rho_\theta^- = \frac{1}{\sqrt{\lambda_1\lambda_2}}\left[-\theta^2\lambda_1\lambda_2 - \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \min\{0, G_{\theta\lambda_1}(i) + G_{\theta\lambda_2}(j) - 1\}\right].$$

It is clear from the above form that $\theta = 0$ yield $\rho_\theta^- = 0$ while when $\theta = 1$ the correlation coincides with $\rho_{\min}$. By Lemma 5.1, ρ_θ^- decreases as θ increases. Consequently, the proposed counter-monotonic shock bivariate Poisson model provides a fully flexible model for negative dependence in the sense that it spans the full spectrum of negative correlation in the interval $[\rho_{\min}, 0]$.

5.2.3 Recurrence relations

As shown in Chapter 4, the derivation of recurrence relations in the $\mathcal{BP}(\Lambda, \theta)$ model relies on the univariate Poisson distribution recursive relation, namely for $k \in \{1, 2, \ldots\}$,

$$g_\lambda(k) = \frac{\lambda}{k}\, g_\lambda(k-1).$$

Note that the contribution towards $f_{\Lambda,\theta}$ due to the comonotonic shock, i.e., $c_{\Lambda,\theta}$, has no direct effect on the simplifications and rather only influences the form of the conditional expectations $\mathrm{E}(Z_1|x_1, x_2)$ and $\mathrm{E}(Z_2|x_1, x_2)$. As such, it is straightforward to see that the same set of recurrence relations will hold in the $\mathcal{BP}^{-}(\Lambda, \theta)$ model, the only difference being the form of the conditional expectations. Thus, the following three recurrence relations hold in the counter-monotonic shock model:

$$\begin{aligned}
f_{\Lambda,\theta}^-(x_1 - 1, x_2) &= f_{\Lambda,\theta}^-(x_1, x_2)\,\{(1-\theta)\lambda_1\}^{-1}\,\{x_1 - \mathrm{E}(Z_1|x_1, x_2)\}\,,\\
f_{\Lambda,\theta}^-(x_1, x_2 - 1) &= f_{\Lambda,\theta}^-(x_1, x_2)\,\{(1-\theta)\lambda_2\}^{-1}\,\{x_2 - \mathrm{E}(Z_2|x_1, x_2)\}\,,\\
f_{\Lambda,\theta}^-(x_1 - 1, x_2 - 1) &= f_{\Lambda,\theta}^-(x_1, x_2)\,\left\{(1-\theta)^2\lambda_1, \lambda_2\right\}^{-1}\Big[x_2\,\{x_1 - \mathrm{E}(Z_1|x_1, x_2)\}\\
&\quad + x_1\,\{x_2 - \mathrm{E}(Z_2|x_1, x_2)\} - \{x_1x_2 - \mathrm{E}(Z_1Z_2|x_1, x_2)\}\Big],
\end{aligned}$$

where the conditional expectations $\mathrm{E}(Z_1|x_1, x_2)$ and $\mathrm{E}(Z_2|x_1, x_2)$ are computed according to the conditional probability mass function

$$\frac{g_{(1-\theta)\lambda_1}(x_1 - z_1)g_{(1-\theta)\lambda_2}(x_2 - z_2)c_{\Lambda,\theta}^-(z_1, z_2)}{\sum_{z_1=0}^{x_1}\sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1)g_{(1-\theta)\lambda_2}(x_2 - z_2)c_{\Lambda,\theta}^-(z_1, z_2)}.$$

5.2.4 Convolutions in the $\mathcal{BP}^-$ family

As was the case in the comonotonic shock model, the $\mathcal{BP}^-(\Lambda, \theta)$ family is not closed under convolutions. Analogously to what was shown in Section 4.2.4, this is due to the fact that sums of counter-monotonic random variables do not retain the property of counter-monotonicity.

Consider a similar set up as in Section 4.2.4. Namely, let (Z_{11}, Z_{21}) and (Z_{12}, Z_{22}) denote independent pairs with distribution $\mathcal{W}\{\mathcal{P}(1), \mathcal{P}(2)\}$. Suppose two independent uniform random variables are observed, specifically $u_1 = 0.1$ and $u_2 = 0.6$. This leads to

$$\begin{aligned} Z_{11} &= G_1^{-1}(0.1) = 0, \quad & Z_{21} &= G_2^{-1}(0.1) = 0, \\ Z_{12} &= G_1^{-1}(0.6) = 1, \quad & Z_{22} &= G_2^{-2}(0.6) = 2. \end{aligned}$$

Then setting $(Z_1, Z_2) = (Z_{11} + Z_{12}, Z_{21} + Z_{22})$ yields an observed pair $(z_1, z_2) = (1, 2)$. Writing $Z_1 = G_2^{-1}(U)$ implies that a value of 1 corresponds to U in the interval $(G_2(0), G_2(1)] = (0.135, 0.406]$. Similarly, for $Z_2 = G_4^{-1}(1 - U)$, observing a value of 2 implies $1 - U$ falls in the interval $(G_4(1), G_4(2)] = (0.092, 0.238]$ so that U falls in the interval $(0.762, 0.908]$. Since $(0.135, 0.406] \cap (0.762, 0.908] = \varnothing$, it follows that (Z_1, Z_2) are not counter-monotonic. Accordingly, the $\mathcal{BP}^-(\Lambda, \theta)$ family is not closed under convolutions.

5.3 Estimation

Let $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2}) \overset{\text{iid}}{\sim} \mathcal{BP}^-(\Lambda, \theta)$. As was the case in the comonotonic shock model, there are several methods that can be used for estimating the parameters $(\lambda_1, \lambda_2, \theta)$. In particular, the method of moments (MM), maximum likelihood (ML) and inference functions for margins (IFM) will be discussed in this section. Naturally, many of the results established in Section 4.3 readily extend to the counter-monotonic shock model, as will be shown in what follows.

5.3.1 Method of moments

Method of moments estimation in the proposed $\mathcal{BP}^-(\Lambda, \theta)$ model is essentially identical to that in the comonotonic shock model for positive dependence. Analogously to what was done in the $\mathcal{BP}(\Lambda, \theta)$ model, the method of moments estimates are established by matching theoretical moments to sample moments. In particular, the marginal parameters are consistently estimated by their respective sample means, i.e., $\tilde{\lambda}_1 = \bar{X}_1$ and $\tilde{\lambda}_2 = \bar{X}_2$. As previously detailed, the Central Limit Theorem ensures that, as $n \to \infty$,

$$\sqrt{n}\,(\tilde{\lambda}_1 - \lambda_1) \rightsquigarrow \mathcal{N}(0, \lambda_1), \quad \sqrt{n}\,(\tilde{\lambda}_2 - \lambda_2) \rightsquigarrow \mathcal{N}(0, \lambda_2).$$

Estimation of the dependence parameter is based on the sample covariance S_{12}. The results in Section 5.2.1 ensure that for fixed λ_1 and λ_2, the covariance of a pair $(X_1, X_2) \sim \mathcal{BP}^-(\Lambda, \theta)$

is a decreasing function of θ. In particular, the covariance will always be negative with minimum possible value give by $w_{\lambda_1,\lambda_2}(1) = \rho_{\min}(\lambda_1,\lambda_2)\sqrt{\lambda_1\lambda_2}$. Accordingly, provided that S_{12} falls within the appropriate interval, there will be a unique value of θ that will yield $w_{\lambda_1,\lambda_2}(\theta) = S_{12}$. The MM estimator $\tilde{\theta}$ is then established by solving

$$w_{\bar{X}_1,\bar{X}_2}(\theta) = S_{12},$$

for $S_{12} \in [\rho_{\min}(\lambda_1,\lambda_2)\sqrt{\bar{X}_1\bar{X}_2}, 0]$.

Recall from Chapter 4 that the sample covariance is consistent and asymptotically Gaussian. Specifically, as $n \to \infty$,

$$\sqrt{n}\,\{S_{12} - w_{\lambda_1,\lambda_2}(\theta)\} \rightsquigarrow \mathcal{N}[0, \sigma^2(\theta,\lambda_1,\lambda_2)],$$

where $\sigma^2(\theta,\lambda_1,\lambda_2) = \text{var}\,\{(X_1-\lambda_1)(X_2-\lambda_2)\}$.

As was established in Proposition 4.3, a straightforward application of the Delta Method then ensures that the method of moments estimator $\tilde{\theta}$ is consistent and asymptotically Gaussian. This is formally stated in the following proposition, analogous to that in Chapter 4.

Proposition 5.2 ***Asymptotic normality of the method of moments estimator***
Let $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ denote a random sample from the $\mathcal{BP}^-(\Lambda,\theta)$ family and let S_{12} denote the sample covariance given by

$$\frac{1}{n-1}\left\{\sum_{\ell=1}^{n}(X_{\ell 1}-\bar{X}_1)(X_{\ell 2}-\bar{X}_2)\right\}.$$

The method of moments estimator is the unique solution of the equation

$$w_{\bar{X}_1,\bar{X}_2}(\theta) = S_{12}.$$

Moreover, for $S_{12} \in [\rho_{\min}(\bar{X}_1,\bar{X}_2)\sqrt{\bar{X}_1\bar{X}_2}, 0]$, as $n \to \infty$,

$$\sqrt{n}\,(\tilde{\theta}-\theta) \rightsquigarrow \mathcal{N}\left[0, \{\delta'(\sigma_{12})\}^2\,\sigma^2(\theta,\lambda_1,\lambda_2)\right]$$

where σ_{12} denotes the covariance $w_{\lambda_1,\lambda_2}(\theta)$ and δ denotes the inverse of the function $w_{\lambda_1,\lambda_2}(\theta)$, i.e., $\delta : \theta \mapsto w^{-1}_{\lambda_1,\lambda_2}(\theta)$, with corresponding derivative δ'.

As was shown in Section 4.3.1, the asymptotic variance of the MM estimator $\tilde{\theta}$ may be simplified. In particular, it was established in Eq. (4.13) that

$$\sigma^2(\theta,\lambda_1,\lambda_2) = (1-\theta)^2\lambda_1\lambda_2 + 2(1-\theta)\theta\lambda_1\lambda_2 + \text{var}\,\{(Z_1-\theta\lambda_1)(Z_2-\theta\lambda_2)\}\,. \tag{5.4}$$

As was the case for the $\mathcal{BP}(\Lambda, \theta)$ family, when $\theta = 0$ the independence model ensues yielding $\sigma^2(0, \lambda_1, \lambda_2) = \lambda_1\lambda_2$. Note, however, that it is not straightforward to further simplify the expression in (5.4) in the special case where $\lambda_1 = \lambda_2 = \lambda$. In the $\mathcal{BP}(\Lambda, \theta)$ model, when the marginal parameters coincide the comonotonic shock is equivalent to a common shock since $Z_1 = Z_2$ almost surely. As a result, Eq. (4.13) further simplifies to yield a simple expression for the asymptotic variance of the MM estimator in the $\mathcal{BP}(\lambda, \lambda, \theta)$ model. This is not the case in the counter-monotonic shock model. Recall that counter-monotonicity implies that Z_1 is generated from an underlying uniform random variable U whereas Z_2 is generated from $1 - U$ and thus Z_1 and Z_2 are not identical shock variables when $\lambda_1 = \lambda_2$. Consequently, computing the term var $\{(Z_1 - \theta\lambda_1)(Z_2 - \theta\lambda_2)\}$ is more involved and requires evaluating higher mixed moments. As such, no simple expression results for the asymptotic variance of $\tilde{\theta}$ when $\lambda_1 = \lambda_2$.

In practice, the implementation of the method of moments estimation may lead to issues, particularly when the sample covariance does not fall within the appropriate range. Similarly to what was done in Section 4.3.1, if $S_{12} < w_{\bar{X}_1, \bar{X}_2}(1)$ the convention will be to set $\tilde{\theta} = 1$ while if $S_{12} > 0$ the model in Chapter 4 should be used instead.

5.3.2 Maximum likelihood estimation

Let $\mathbf{X}$ denote a random sample $(X_{11}, X_{12}), \ldots, (X_{n1}, X_{n2})$ from the proposed $\mathcal{BP}^-(\Lambda, \theta)$ model. For $\mathbf{X} = \mathbf{x}$, the log-likelihood is given by

$$\begin{aligned} \ell^-(\Lambda, \theta; \mathbf{x}) &= \sum_{i=1}^{n} \ln f^-_{\Lambda,\theta}(x_{i1}, x_{i2}) \\ &= \sum_{i=1}^{n} \ln \left\{ \sum_{z_1=0}^{x_{i1}} \sum_{z_2=0}^{x_{i2}} g_{(1-\theta)\lambda_1}(x_{i1} - z_1) g_{(1-\theta)\lambda_2}(x_{i2} - z_2) c^-_{\Lambda,\theta}(z_1, z_2) \right\}, \end{aligned}$$

with

$$c^-_{\Lambda,\theta}(z_1, z_2) = \left[\min\{G_{\theta\lambda_1}(z_1), \bar{G}_{\theta\lambda_2}(z_2 - 1)\} - \max\{G_{\theta\lambda_1}(z_1 - 1), \bar{G}_{\theta\lambda_2}(z_2)\}\right]_+ .$$

By definition, the maximum likelihood estimates, $\hat{\Psi} = (\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta})$, are given by

$$\hat{\Psi} = \arg\max_{\Psi} \ell^-(\Lambda, \theta; \mathbf{x})$$

with the constraints $\lambda_1, \lambda_2 \in (0, \infty)$, $\theta \in [0, 1]$. This amounts to solving the set of score equations:

$$\left(\frac{\partial}{\partial\lambda_1} \ell^-(\Lambda, \theta), \frac{\partial}{\partial\lambda_2} \ell^-(\Lambda, \theta), \frac{\partial}{\partial\theta} \ell^-(\Lambda, \theta) \right) = \mathbf{0}^\top.$$

In Section 4.3.2, the score equations were further simplified by differentiating the univariate Poisson probability mass function contributions in the log-likelihood. These contributions are due to the independent components (Y_1, Y_2) and are identical in the counter-monotonic shock model. Note that throughout the computations in Section 4.3.2, the partial derivatives involving the comonotonic shock mass function, $c_{\Lambda,\theta}(z_1, z_2)$, were not further simplified due to their inconvenient form. As such, the derivations established in Section 4.3.2 are easily adapted for the $\mathcal{BP}^-(\Lambda, \theta)$ model to yield the following set of score equations:

$$\begin{aligned}
\frac{\partial}{\partial \lambda_1}\,\ell(\Lambda,\theta) &= \frac{n}{\lambda_1}(\bar{x}_1 - \lambda_1) - \frac{n}{\lambda_1}\left\{\bar{q}_1(\Lambda,\theta) - \theta\lambda_1\right\} + \sum_{i=1}^{n} \mathrm{E}\left\{\frac{\partial}{\partial \lambda_1}\,\ln c^{-}_{\Lambda,\theta}(Z_1, Z_1)\middle| x_{i1}, x_{i2}\right\},\\
\frac{\partial}{\partial \lambda_2}\,\ell(\Lambda,\theta) &= \frac{n}{\lambda_2}(\bar{x}_2 - \lambda_2) - \frac{n}{\lambda_2}\left\{\bar{q}_2(\Lambda,\theta) - \theta\lambda_2\right\} + \sum_{i=1}^{n} \mathrm{E}\left\{\frac{\partial}{\partial \lambda_2}\,\ln c^{-}_{\Lambda,\theta}(Z_1, Z_1)\middle| x_{i1}, x_{i2}\right\},\\
\frac{\partial}{\partial \theta}\,\ell(\Lambda,\theta) &= n(\lambda_1 + \lambda_2) - \frac{n}{(1-\theta)}\left\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda,\theta) - \bar{q}_2(\Lambda,\theta)\right\}\\
&\qquad + \sum_{i=1}^{n} \mathrm{E}\left\{\frac{\partial}{\partial \theta}\,\ln c^{-}_{\Lambda,\theta}(Z_1, Z_2)\middle| x_{i1}, x_{i2}\right\},
\end{aligned}$$

where

$$\bar{q}_1(\Lambda,\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(Z_1 | x_{i1}, x_{i2}), \quad \bar{q}_2(\Lambda,\theta) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(Z_2 | x_{i1}, x_{i2}).$$

Note that although the same notation for $\bar{q}_1(\Lambda, \theta)$ and $\bar{q}_2(\Lambda, \theta)$ is used for both the $\mathcal{BP}$ and $\mathcal{BP}^-$ models, the actual calculation of these two expressions is according to the appropriate conditional probability mass function. In the present case, this corresponds to

$$\frac{g_{(1-\theta)\lambda_1}(x_1 - z_1)\, g_{(1-\theta)\lambda_2}(x_2 - z_2)\, c^{-}_{\Lambda,\theta}(z_1, z_2)}{\sum_{z_1=0}^{x_1}\sum_{z_2=0}^{x_2} g_{(1-\theta)\lambda_1}(x_1 - z_1)\, g_{(1-\theta)\lambda_2}(x_2 - z_2)\, c^{-}_{\Lambda,\theta}(z_1, z_2)}.$$

In the actual implementation of the maximum likelihood estimation, the log-likelihood function $\ell^-(\Lambda, \theta)$ can be directly optimized using numerical procedures. As was mentioned in Chapter 4, a reparameterization of the log-likelihood can be used wherein the parameter constraints are removed. This is achieved by setting

$$\zeta_1 = \ln(\lambda_1), \quad \zeta_2 = \ln(\lambda_2), \quad \eta = \ln\left(\frac{\theta}{1-\theta}\right).$$

By the invariance property of MLEs it follows that

$$(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta}) = (e^{\hat{\zeta}_1}, e^{\hat{\zeta}_2}, e^{\hat{\eta}}/(1 + e^{\hat{\eta}})).$$

As was outlined in Chapter 4, standard maximum likelihood theory ensures that the MLE $\hat{\Psi}$ is consistent and asymptotically Gaussian. Specifically

$$\sqrt{n}\,(\hat{\Psi} - \Psi) \rightsquigarrow \mathcal{N}(0, \mathcal{I}^{-1}),$$

where $\mathcal{I}$ denotes the Fisher Information. The asymptotic variance is thus given by

$$\mathcal{I}^{-1} = \left[\mathrm{E}\left[\left\{ \frac{\partial}{\partial \Psi} \ln f^{-}_{\Lambda,\theta}(X_1, X_2) \right\}^2 \right] \right]^{-1} = \left[-\mathrm{E}\left\{ \frac{\partial^2}{\partial \Psi \partial \Psi^\top} \ln f^{-}_{\Lambda,\theta}(X_1, X_2) \right\} \right]^{-1}.$$

The EM algorithm

In the comonotonic shock model, it was shown that using the EM algorithm to determine the MLEs lead to non-identifiability. This problem is also inherent in the proposed $\mathcal{BP}^-$ model. By definition, the counter-monotonic shock (Z_1, Z_2) can be expressed as $\big(G^{-1}_{\theta\lambda_1}(U), G^{-1}_{\theta\lambda_2}(1-U)\big)$. Treating the underlying uniform variable U as the missing data yields the same E-step as produced in the comonotonic shock model, viz.

$$\begin{aligned} Q(\Psi|\Psi^{(k)}) = n\left\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)}) - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)}\right\} \ln\{(1-\theta)(\lambda_1+\lambda_2)\} \\ - n(1-\theta)(\lambda_1+\lambda_2) + n\left\{\bar{x}_1 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)})\right\} \ln\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right) \\ + n\left\{\bar{x}_2 - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\right\} \ln\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right) - R(\mathbf{x}, \Psi^{(k)}). \end{aligned}$$

Setting $\alpha = \lambda_1/(\lambda_1+\lambda_2)$ and $\beta = (1-\theta)(\lambda_1+\lambda_2)$, the above can be rewritten as

$$\begin{aligned} Q(\Psi|\Psi^{(k)}) = -n\beta + n\left\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)}) - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\right\} \ln(\beta) \\ + n\left\{\bar{x}_1 - \bar{q}_1(\Lambda^{(k)}, \theta^{(k)})\right\} \ln(\alpha) + n\left\{\bar{x}_2 - \bar{q}_2(\Lambda^{(k)}, \theta^{(k)})\right\} \ln(1-\alpha) \\ - R(\mathbf{x}, \Psi^{(k)}). \end{aligned}$$

Note again that in the above expression the terms $\bar{q}_1(\Lambda, \theta)$ and $\bar{q}_2(\Lambda, \theta)$ are computed under the $\mathcal{BP}^-(\Lambda, \theta)$ model.

Reiterating what was already established in Chapter 4, the form of $Q(\Psi|\Psi^{(k)})$ only allows to identify (α, β). Hence, any set of values $(\lambda_1, \lambda_2, \theta)$ that yield the same (α, β) are indistinguishable and thus the EM algorithm will be unable to update the parameter estimates.

5.3.3 Inference functions for margins

Let $p^{-}_{\Lambda,\theta}$ denote the probability mass function of the quadruple (Z_1, Z_2, X_1, X_2), with components as defined in (5.1) and (5.2). Analogously to what was established in Section 4.3.3, it can be shown that

$$p^{-}_{\Lambda,\theta}(z_1, z_2, x_1, x_2) = g_{\lambda_1}(x_1)g_{\lambda_2}(x_2)b_{x_1,\theta}(z_1)b_{x_2,\theta}(z_2)c^{-}_{\Lambda,\theta}(z_1, z_2)/\left\{g_{\theta\lambda_1}(z_1)g_{\theta\lambda_2}(z_2)\right\},$$

where

$$b_{x_k,\theta}(z_k) = \binom{x_k}{z_k}\theta^{z_k}(1-\theta)^{x_k-z_k}.$$

Further define weights $\omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2)$ as

$$\omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2) = \frac{b_{x_1,\theta}(z_1)b_{x_2,\theta}(z_2)}{g_{\theta\lambda_1}(z_1)g_{\theta\lambda_2}(z_2)}.$$

Following the same steps as in Section 4.3.3, it can be shown that the probability mass function in the $\mathcal{BP}^{-}(\Lambda,\theta)$ model can be rewritten as

$$f^{-}_{\Lambda,\theta}(x_1, x_2) = g_{\lambda_1}(x_1)g_{\lambda_2}(x_2) \times \left[\sum_{z_1=0}^{x_1}\sum_{z_2=0}^{x_2} \omega_{\Lambda,\theta}(z_1, z_2; x_1, x_2)c^{-}_{\Lambda,\theta}(z_1, z_2)\right]. \tag{5.5}$$

As was the case for $f_{\Lambda,\theta}$, this formulation of $f^{-}_{\Lambda,\theta}$ resembles the construction considered in Lakshminarayana et al. (1999) wherein the bivariate probability mass function is comprised of the product of the marginal Poisson mass functions along with a multiplicative factor which encompasses the dependence structure.

Based on the form in (5.5), the log-likelihood in the $\mathcal{BP}^{-}(\Lambda,\theta)$ model can be expressed as

$$\ell^{-}(\Lambda,\theta;\mathbf{x}) = \ell_1(\lambda_1;\mathbf{x}_1) + \ell_2(\lambda_2;\mathbf{x}_2) + \ell^{-}_{D}(\Lambda,\theta;\mathbf{x}_1,\mathbf{x}_2),$$

where $\mathbf{x}_j = \{x_{1j}, \ldots, x_{nj}\}$ for $j \in \{1, 2\}$. The marginal contributions to the log-likelihood simplify, for $j \in \{1, 2\}$, as follows

$$\ell_j(\lambda_j;\mathbf{x}_j) = \sum_{i=1}^{n} \ln g_{\lambda_j}(x_{ij}) = -n\lambda_j + n\bar{x}_j \ln(\lambda_j) - \sum_{i=1}^{n} \ln(x_{ij}!).$$

The remaining portion of the log-likelihood, namely the dependence term, consists of

$$\ell^{-}_{D}(\Lambda,\theta;\mathbf{x}) = \sum_{i=1}^{n} \ln\left\{\sum_{z_1=0}^{x_{i1}}\sum_{z_2=0}^{x_{i2}} \omega_{\Lambda,\theta}(z_1, z_2; x_{i1}, x_{i2})c^{-}_{\Lambda,\theta}(z_1, z_2)\right\}.$$

The application of the IFM estimation approach in the $\mathcal{BP}^{-}(\Lambda,\theta)$ model parallels that in the comonotonic shock model. Indeed, the IFM estimates, denoted as $(\check{\lambda}_1, \check{\lambda}_2, \check{\theta})$, are found according to

$$\left(\frac{\partial}{\partial\lambda_1}\,\ell_1(\lambda_1), \frac{\partial}{\partial\lambda_2}\,\ell_2(\lambda_2), \frac{\partial}{\partial\theta}\,\ell^{-}_{D}(\lambda_1,\lambda_2,\theta)\right) = \mathbf{0}^{\top}.$$

As discussed in Chapter 3, implementation of the above consists of a two-stage procedure. As a first step, the marginal parameters are estimated according to their marginal MLEs. Subsequently, the dependence parameter is estimated using the pseudo log-likelihood $\ell^-(\check{\Lambda}, \theta)$, i.e., the log-likelihood with the marginal parameters held fixed at their respective marginal MLEs. As was the case in the $\mathcal{BP}$ model, the marginal parameters are estimated by their respective sample means: $\check{\lambda}_1 = \bar{X}_1$, $\check{\lambda}_2 = \bar{X}_2$. The dependence parameter is then determined as

$$\check{\theta} = \arg\max_{\theta} \ell^-(\check{\Lambda}, \theta) = \arg\max_{\theta} \ell^-_D(\check{\Lambda}, \theta).$$

As was outlined in Chapter 3, the IFM estimates $\check{\Psi} = (\check{\lambda}_1, \check{\lambda}_2, \check{\theta})$ have a joint limiting Gaussian distribution, specifically, as $n \to \infty$,

$$\sqrt{n}\,(\check{\Psi} - \Psi) \rightsquigarrow \mathcal{N}(0, V).$$

Following the definition of V given in Section 4.3.3, it can be shown that the elements of the asymptotic variance simplify to

$$\begin{aligned}
&V_{11} = \lambda_1, \quad V_{12} = w_{\lambda_1,\lambda_2}(\theta), \quad V_{13} = -\mathcal{I}_{33}^{-1}\{\lambda_1 \mathcal{I}_{31} + w_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{32}\},\\
&V_{22} = \lambda_2, \quad V_{23} = -\mathcal{I}_{33}^{-1}\{w_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{31} + \lambda_2 \mathcal{I}_{32}\},\\
&V_{33} = \mathcal{I}_{33}^{-1} + (\mathcal{I}_{33}^{-1})^2\left\{\lambda_1 \mathcal{I}_{31}^2 + 2w_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{31}\mathcal{I}_{32} + \lambda_2 \mathcal{I}_{32}^2\right\},
\end{aligned}$$

where $w_{\lambda_1,\lambda_2}(\theta) = \text{cov}(X_1, X_2)$ in the $\mathcal{BP}^-(\Lambda, \theta)$ model. As was the case in the comonotonic shock model, the IFM approach yields estimators that are less efficient than the MLEs. In a similar manner, this loss in efficiency associated with the dependence parameter θ is quantified as

$$V_{33} - \mathcal{I}_{33}^{-1} = (\mathcal{I}_{33}^{-1})^2\left\{\lambda_1 \mathcal{I}_{31}^2 + 2w_{\lambda_1,\lambda_2}(\theta)\mathcal{I}_{31}\mathcal{I}_{32} + \lambda_2 \mathcal{I}_{32}^2\right\}.$$

This will be further studied in the simulations in the next section.

As was pointed out in Section 4.3.3, the EM algorithm can be used as an iterative procedure for finding $\check{\theta}$. Using the underlying uniform variables as the missing data, the complete data pseudo log-likelihood has a similar form as (4.25), viz.

$$\begin{aligned}
\ell^-_C(\check{\Lambda}, \theta; \mathbf{x}, \mathbf{u}) = \sum_{i=1}^{n} \Big[&-(1-\theta)(\bar{x}_1 + \bar{x}_2) + \ln(1-\theta)\left\{x_{i1} + x_{i2} - G^{-1}_{\theta\bar{x}_1}(u_i) - G^{-1}_{\theta\bar{x}_2}(1-u_i)\right\}\\
&+ \ln(\bar{x}_1)\left\{x_{i1} - G^{-1}_{\theta\bar{x}_1}(u_i)\right\} + \ln(\bar{x}_2)\left\{x_{i2} - G^{-1}_{\theta\bar{x}_2}(1-u_i)\right\}\\
&- \ln\left[\left\{x_{i1} - G^{-1}_{\theta\bar{x}_1}(u_i)\right\}!\left\{x_{i2} - G^{-1}_{\theta\bar{x}_2}(1-u_i)\right\}!\right]\Big].
\end{aligned}$$

As before, let

$$q_{i1}(\check{\Lambda}, \theta^{(k)}) = \mathrm{E}\left\{G^{-1}_{\theta\bar{x}_1}(U_i)|x_{i1}, x_{i2}, \theta^{(k)}\right\} = \mathrm{E}\left\{Z_1|x_{i1}, x_{i2}, \theta^{(k)}\right\},$$
$$q_{i2}(\check{\Lambda}, \theta^{(k)}) = \mathrm{E}\left\{G^{-1}_{\theta\bar{x}_2}(1-U_i)|x_{i1}, x_{i2}, \theta^{(k)}\right\} = \mathrm{E}\left\{Z_2|x_{i1}, x_{i2}, \theta^{(k)}\right\},$$

and correspondingly define

$$\bar{q}_1(\check{\Lambda}, \theta^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} q_{i1}(\check{\Lambda}, \theta^{(k)}), \quad \bar{q}_2(\check{\Lambda}, \theta^{(k)}) = \frac{1}{n}\sum_{i=1}^{n} q_{i2}(\check{\Lambda}, \theta^{(k)}),$$

where the conditional expectations are under the $\mathcal{BP}^{-}(\check{\Lambda}, \theta)$ model. The E-step is then identical to that derived in Chapter 4, viz.

$$Q(\theta|\theta^{(k)}) = -n(1-\theta)(\bar{x}_1 + \bar{x}_2) + n\ln(1-\theta)\left\{\bar{x}_1 + \bar{x}_2 - \bar{q}_1(\check{\Lambda}, \theta^{(k)}) - \bar{q}_2(\check{\Lambda}, \theta^{(k)})\right\}$$
$$+ n\ln(\bar{x}_1)\left\{\bar{x}_1 - \bar{q}_1(\check{\Lambda}, \theta^{(k)})\right\} + n\ln(\bar{x}_2)\left\{\bar{x}_2 - \bar{q}_2(\check{\Lambda}, \theta^{(k)})\right\} - R(\mathbf{x}, \theta^{(k)}).$$

Accordingly, this leads to the same updates as obtained in Chapter 4, viz.

$$\theta^{(k+1)} = \frac{\bar{q}_1(\check{\Lambda}, \theta^{(k)}) + \bar{q}_2(\check{\Lambda}, \theta^{(k)})}{\bar{x}_1 + \bar{x}_2}.$$

5.4 Simulations

In order to evaluate the estimation procedures outlined in Section 5.3, a series of simulations were carried out. The three estimation methods were compared for varying values of $\lambda_1 \in \{1, 2, 4, 10\}$, $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and $n \in \{10, 50, 100, 1000\}$, while holding $\lambda_2 = 5$ fixed throughout. As explained in Chapter 4, λ_2 was held fixed as it is redundant to test the effects of both marginal parameters on the estimation procedures.

The simulations were set-up in a similar manner to those detailed in Section 4.4. For each scenario defined by a unique set $(\lambda_1, \lambda_2, \theta, n)$, a set of n mutually independent copies of $\mathcal{U}(0, 1)$ random variables (U_i, V_{i1}, V_{i2}) was generated and for each $i \in \{1, \ldots, n\}$, observations from the proposed $\mathcal{BP}^{-}(\Lambda, \theta)$ model were generated according to

$$X_{i1} = G^{-1}_{(1-\theta)\lambda_1}(V_{i1}) + G^{-1}_{\theta\lambda_1}(U_i), \quad X_{i2} = G^{-1}_{(1-\theta)\lambda_2}(V_{i1}) + G^{-1}_{\theta\lambda_2}(1-U_i).$$

Each of the 80 unique scenarios was repeated for 500 replications, thus running a total of 40,000 iterations.

Initial tests revealed that both the IFM and MLE optimization procedures were relatively sensitive to the choice of starting values. Thus, rather than using the method of moments estimates as the initial values, the true parameter values were used. As was the case in the $\mathcal{BP}$ model, the EM algorithm within the IFM framework tended to be slow to converge in comparison to directly optimizing the pseudo log-likelihood $\ell^{-}(\check{\Lambda}, \theta)$. As such, the simulations did not include the EM-IFM estimation procedure. Both the IFM and MLE estimates were produced using the `optim` function in R with the BFGS method, while the MM estimates were obtained from the `uniroot` function.

Out of the $40{,}000$ iterations, there were 3861 occurrences where the method of moments estimation procedure failed due to a positive observed sample covariance. As detailed in Section 5.3.1, the MM estimator $\tilde{\theta}$ is determined as the unique solution of $w_{\bar{X}_1, \bar{X}_2}(\theta) = S_{12}$. By design, a random pair $(X_1, X_2) \sim \mathcal{BP}^{-}(\Lambda, \theta)$ has covariance in the interval $[w_{\lambda_1, \lambda_2}(1), 0]$. Thus, if $S_{12} > 0$, the estimation procedure will fail and set $\tilde{\theta} = NA$. Not surprisingly, this issue arose more frequently for smaller values of θ and smaller sample sizes. The IFM method produced substantially fewer errors, with only 436 of the $40{,}000$ iterations failing. Most often, this was due to inappropriate starting values, where $\ell(\check{\Lambda}, \theta_0) = -\infty$. There were nine instances where the error message indicated "non-finite finite-difference value", apparently due to issues that arise when trying to calculate the gradient. The latter issue tended to occur for higher values of θ, when the $\mathcal{BP}^{-}$ model approaches the lower Fréchet–Hoeffding boundary distribution. Maximum likelihood estimation worked best, with only six iterations producing errors. In all six of these instances, the errors were due to "non-finite finite-difference value". Again, this error only occurred for larger values of θ, specifically when $\theta = 0.90$. As will be seen later, these issues are likely due to the behavior of the log-likelihood for larger values of θ; see Figures 5.22 through 5.26. Naturally, the MM and IFM estimators for the marginal parameters never posed any issues as both methods estimate λ_1 and λ_2 by $\bar{X}_1$ and $\bar{X}_2$, respectively. Of course, the maximum likelihood estimation failed for λ_1 and λ_2 in the instances previously described in terms of θ.

In addition to these errors, there were several instances where the optimization procedure for both the IFM and MLE methods reached the maximum number of iterations, which is 100 by default. This occurred in 76 iterations for maximum likelihood estimation and 64 iterations for the IFM method. While for maximum likelihood estimation this only occurred when the sample size was 10, the IFM estimation reached the maximum number of iterations for sample sizes of 100 and under. Seemingly, this issue is a small-sample problem.

Figures 5.1 through 5.10 display the estimation results for the dependence parameter θ from the three methods. Note that iterations that resulted in issues, as previously discussed, were omitted from the boxplots. Similar patterns as observed in the comonotonic shock model in Chapter 4 are also apparent in these graphs. In particular, the method of moments seems to perform comparatively better for weaker levels of dependence, i.e., for smaller values of θ. Moreover, the variability in the likelihood-based estimates seems to decrease as θ increases. Both likelihood-based methods, namely the IFM and MLE approaches, produce similar results. This suggests that there is little information concerning the marginal parameters Λ in the dependence portion of the log-likelihood $\ell_D^-(\Lambda, \theta)$. Recall that the implementation of the IFM approach yields $\check{\theta}$ as the maximizer of the pseudo log-likelihood $\ell_D^-(\check{\Lambda}, \theta)$. Maximum likelihood estimation, in contrast, simultaneously solves for $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\theta})$ that maximizes the full log-likelihood $\ell^-(\Lambda, \theta)$. Additionally, the magnitude of the marginal parameter seems to have no effect on the estimation of the dependence parameter. The figures also depict a decrease in variability of the estimators as the sample size increases.

In the simulations for the $\mathcal{BP}^-$ model, it was found that the method of moments estimation procedure tended to run the fastest, followed by the IFM and finally MLE methods. The differences between the MM and IFM methods was very small, and even the MLE approach completed in only slightly longer running times.

Figures 5.11 through 5.20 show the estimation results for λ_1. Both approaches yielded similar results, for all degrees of dependence and sample size. This provides further evidence that there is little information concerning λ_1 in the dependence portion of the log-likelihood $\ell_D^-(\Lambda, \theta)$.

A second set of simulations was implemented in which the effects of the starting values were tested, similarly to what was done in Chapter 4. Holding the parameters fixed at $(\lambda_1, \lambda_2, \theta) = (4, 5, 0.9)$ and the sample size at $n = 1000$, four different starting values were considered for the IFM and MLE estimation methods:

(1) $\theta_1^{(0)} = \tilde{\theta}$, i.e., the MM estimate
(2) $\theta_2^{(0)} = 0.10$
(3) $\theta_3^{(0)} = \theta_0$, i.e., the true value 0.90
(4) $\theta_4^{(0)} = 0.99$.

A total of 500 replications were considered, each of which generating four sets of estimates, $\check{\theta}$ and $\hat{\theta}$, obtained from initializing the IFM and MLE estimation procedures at each of $\theta_1^{(0)}$, $\theta_2^{(0)}$, $\theta_3^{(0)}$ and $\theta_4^{(0)}$. Note that for the MLE approach, the initial values for the marginal parameters were taken to be the marginal sample means, i.e., $(\lambda_1^{(0)}, \lambda_2^{(0)}) = (\bar{X}_1, \bar{X}_2)$, throughout.

For the particular scenario of $(\lambda_1, \lambda_2, \theta, n)$ considered in this second simulation study, the method of moments estimation procedure worked without error for each of the 500 replications. Overall, there were 809 instances out of the total of 2000 iterations where the IFM estimation failed and 890 error occurrences in the ML estimation. In the iterations where the estimation procedures did not yield errors, there were no instances where the optimization algorithm reached the maximum number of iterations before converging for both the IFM and MLE methods.

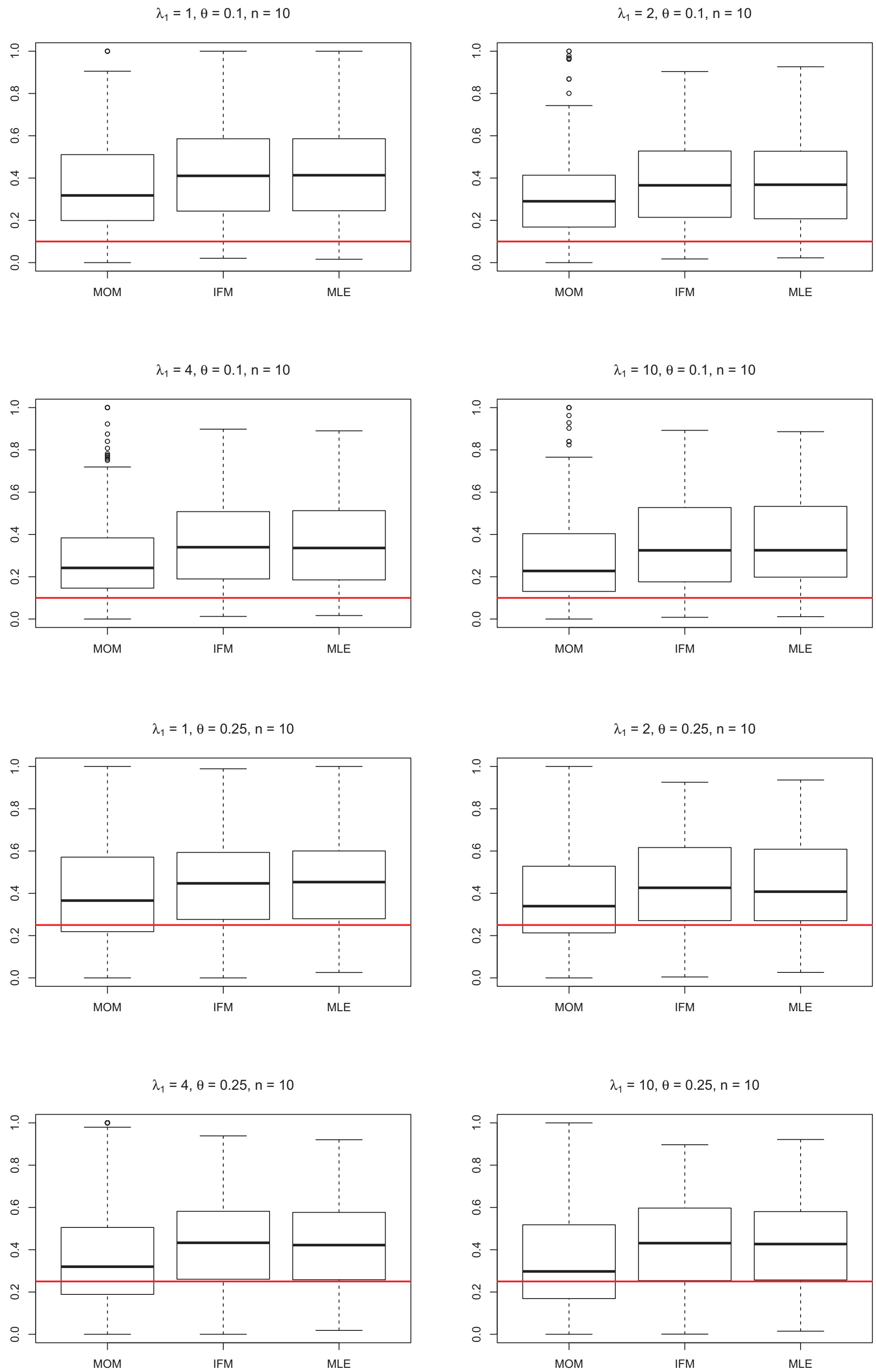

Figure 5.1: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

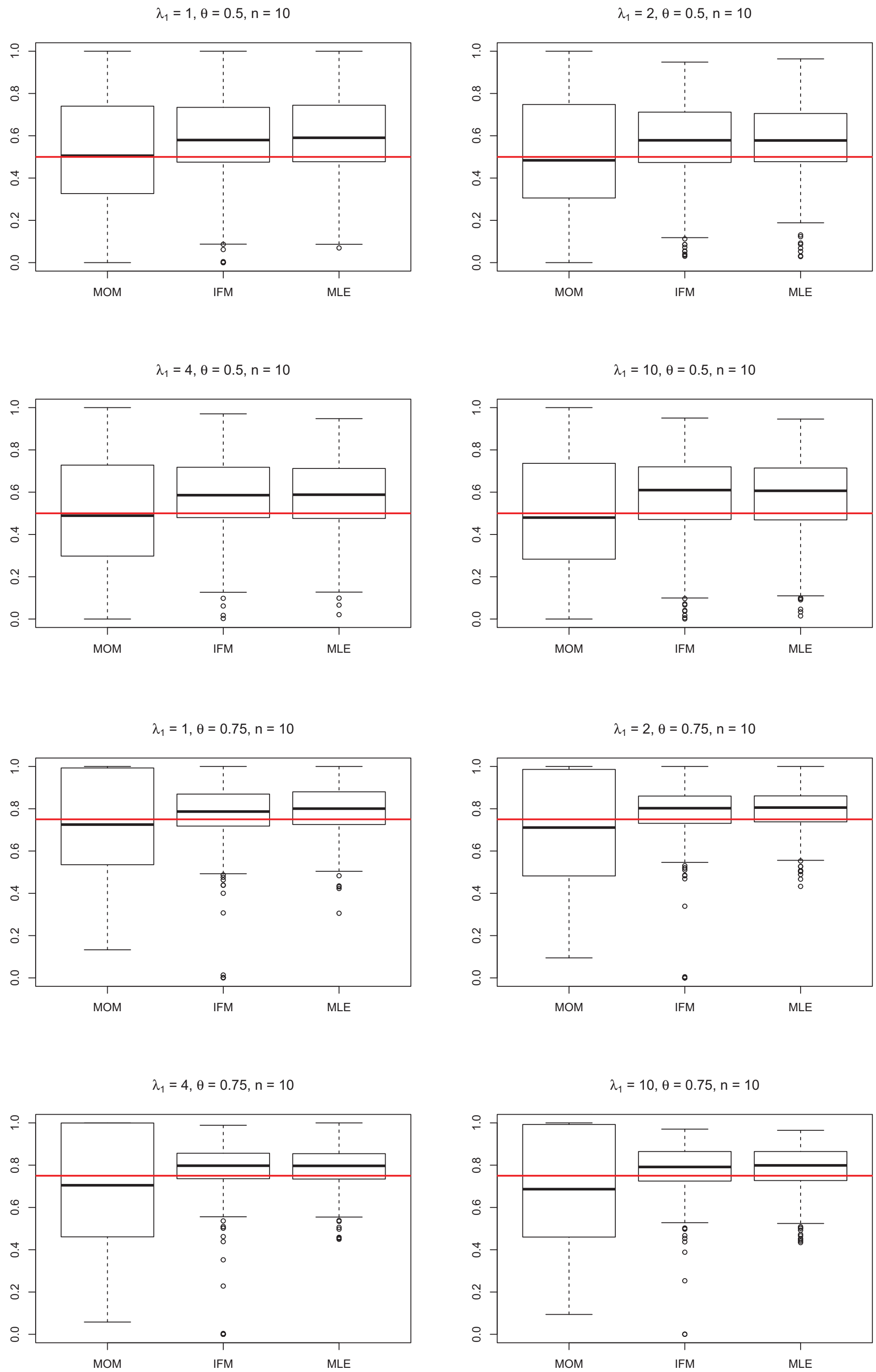

Figure 5.2: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

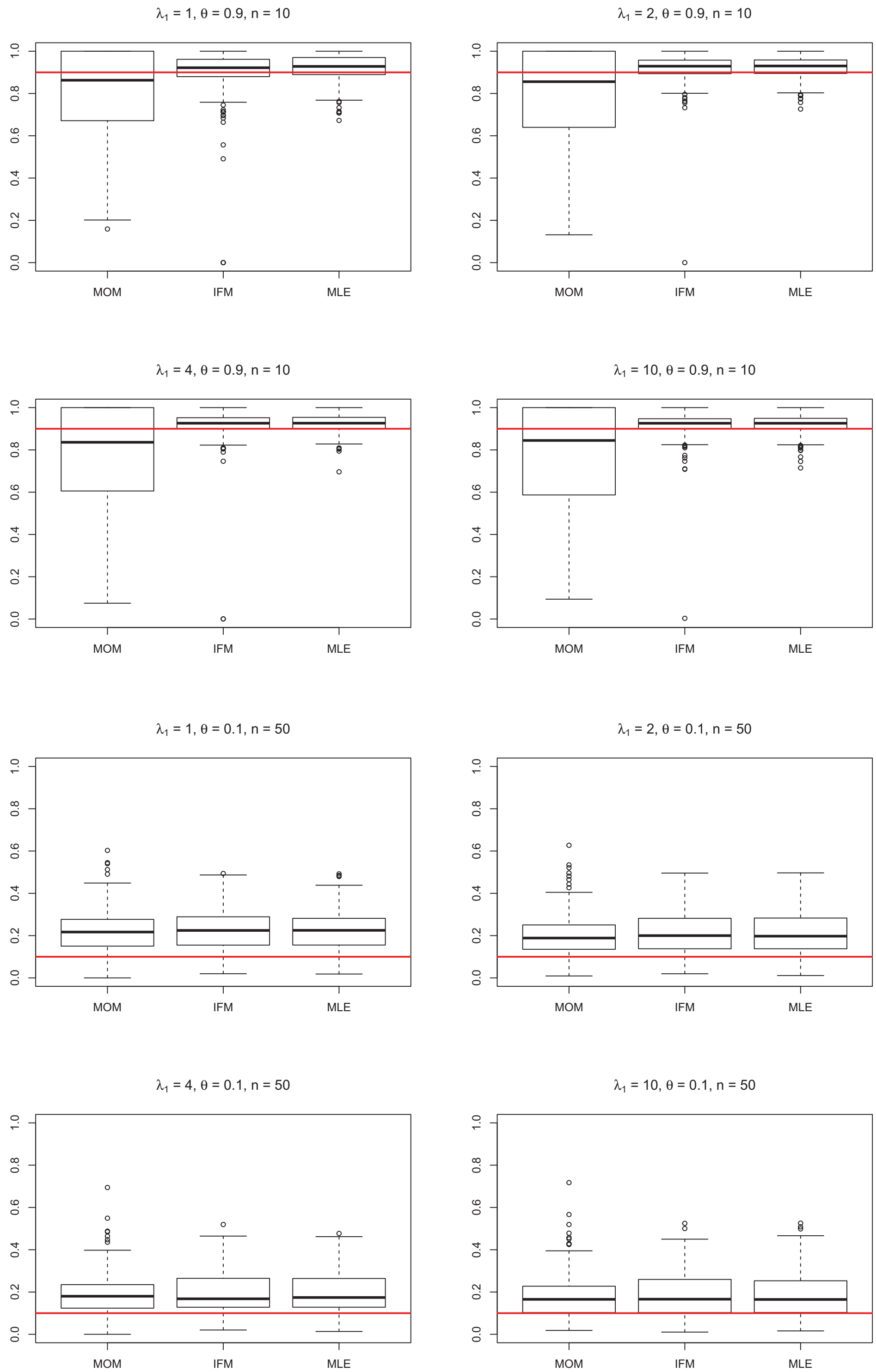

Figure 5.3: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

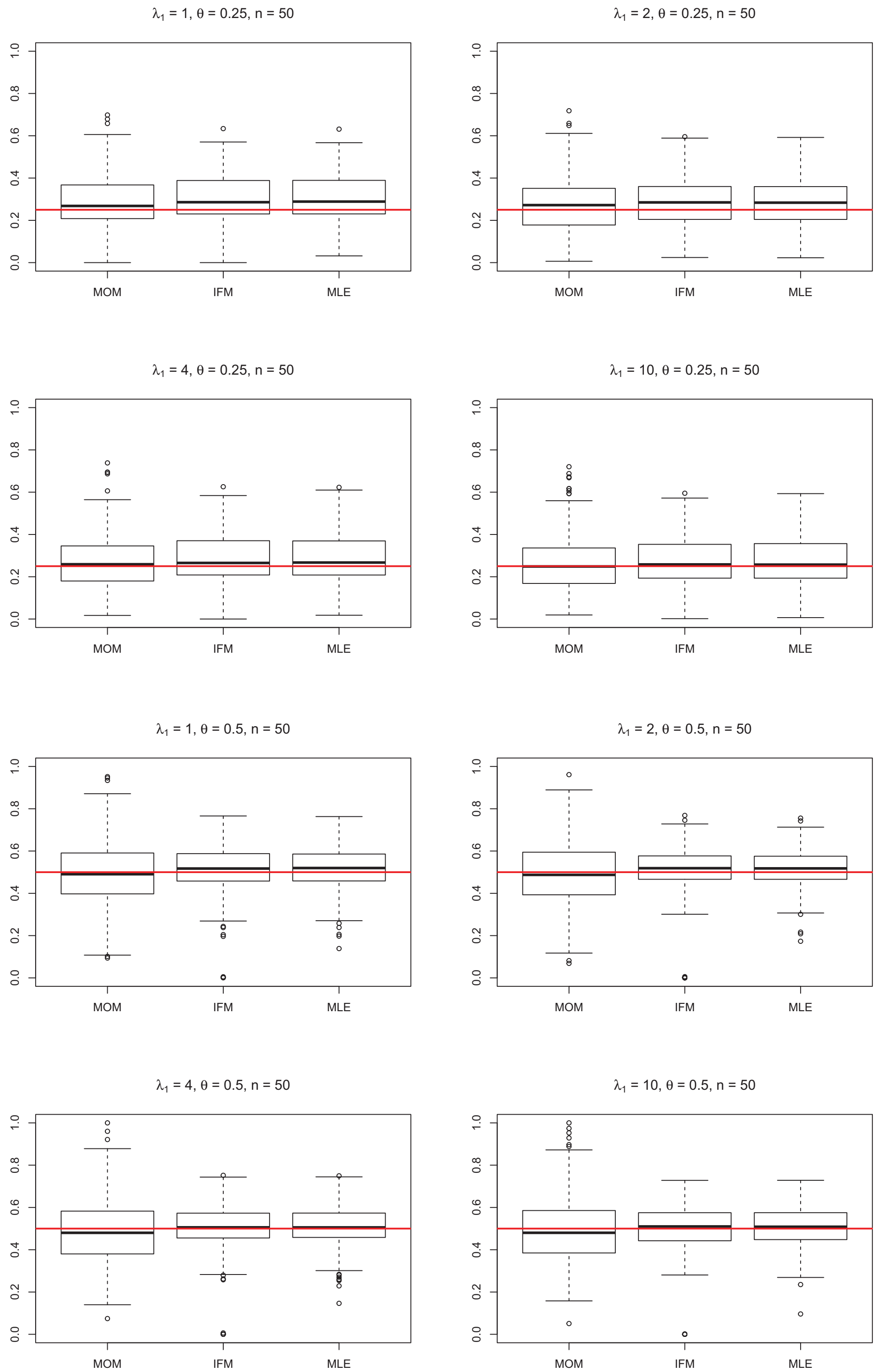

Figure 5.4: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

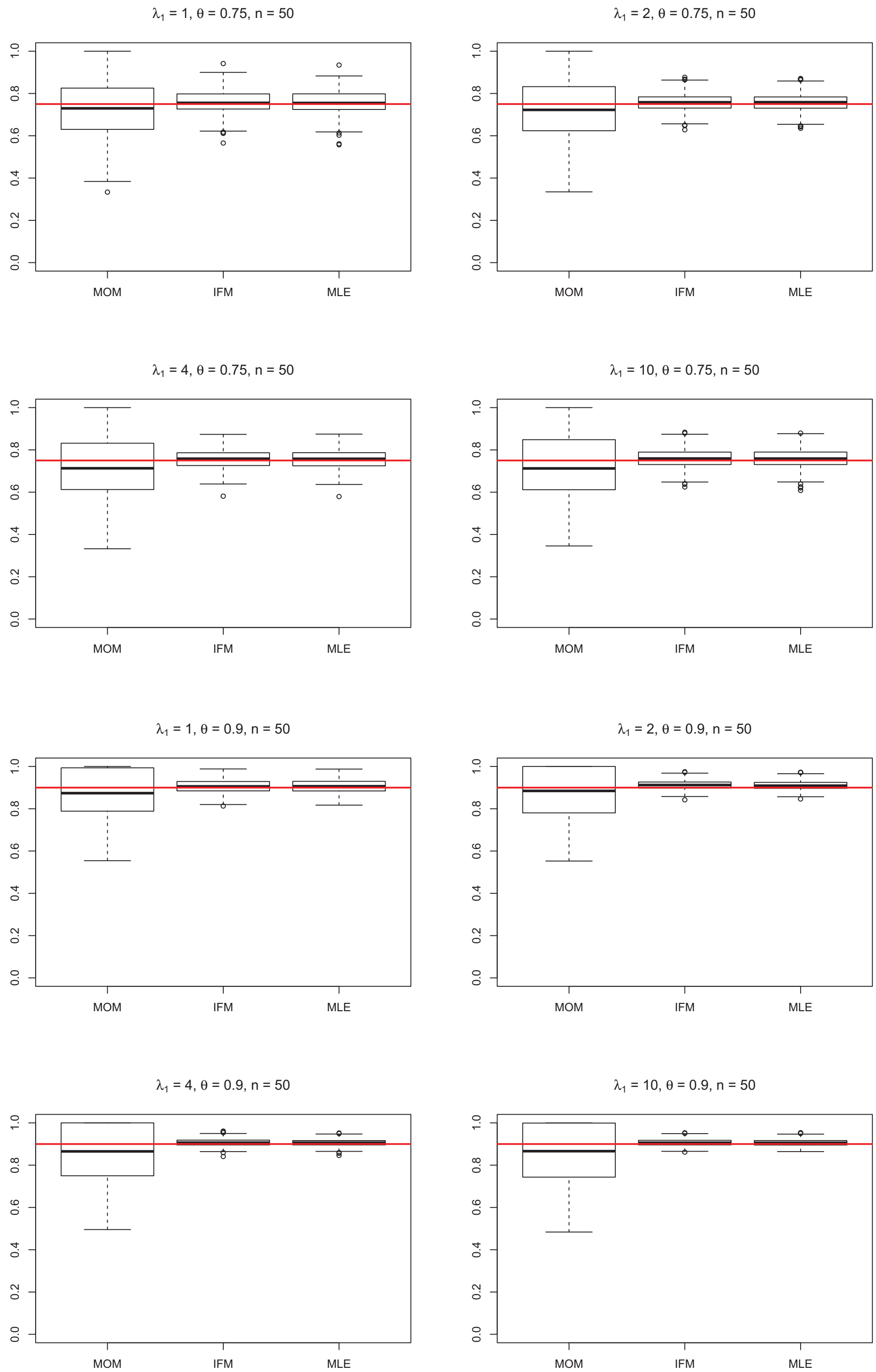

Figure 5.5: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

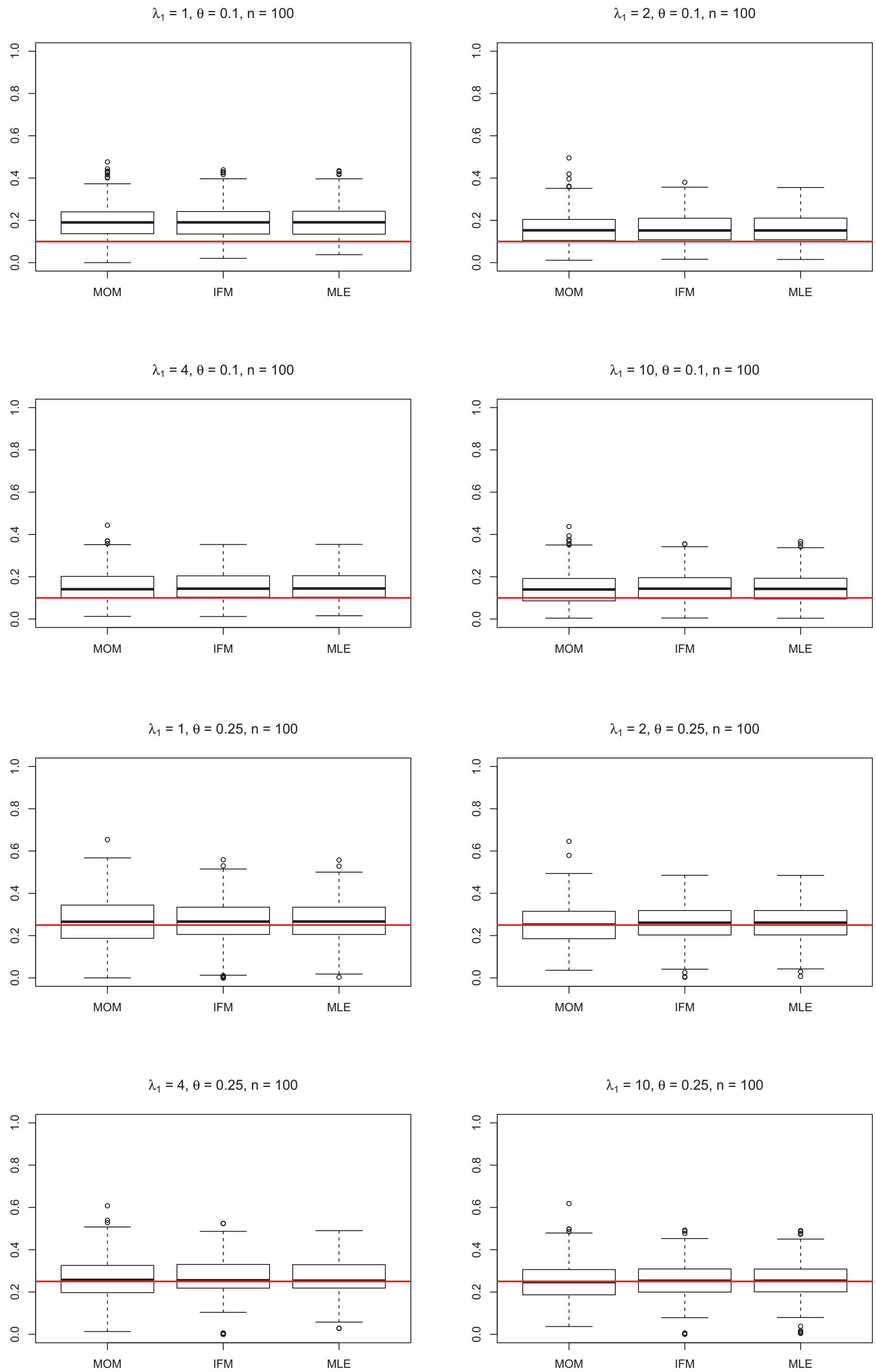

Figure 5.6: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

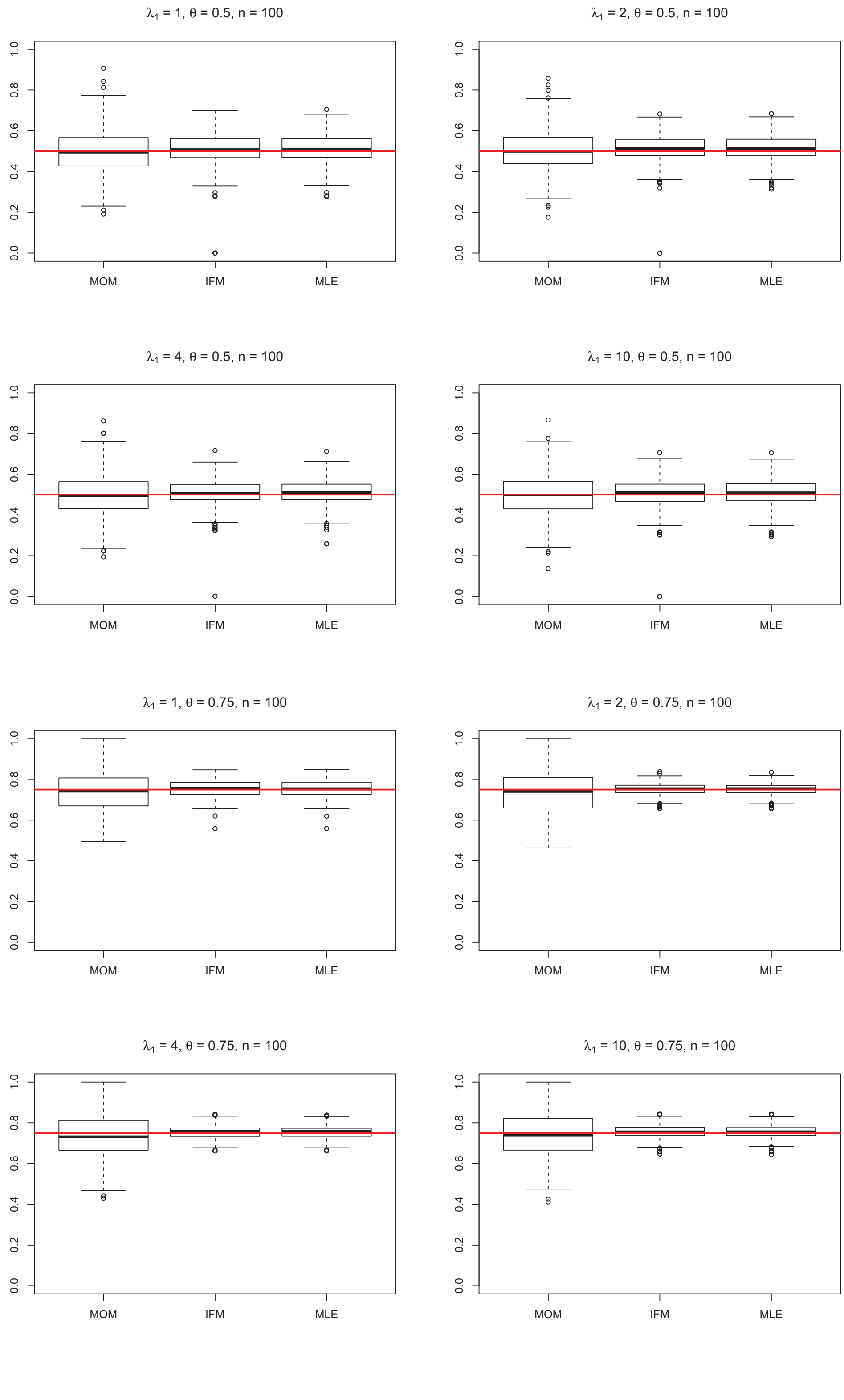

Figure 5.7: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

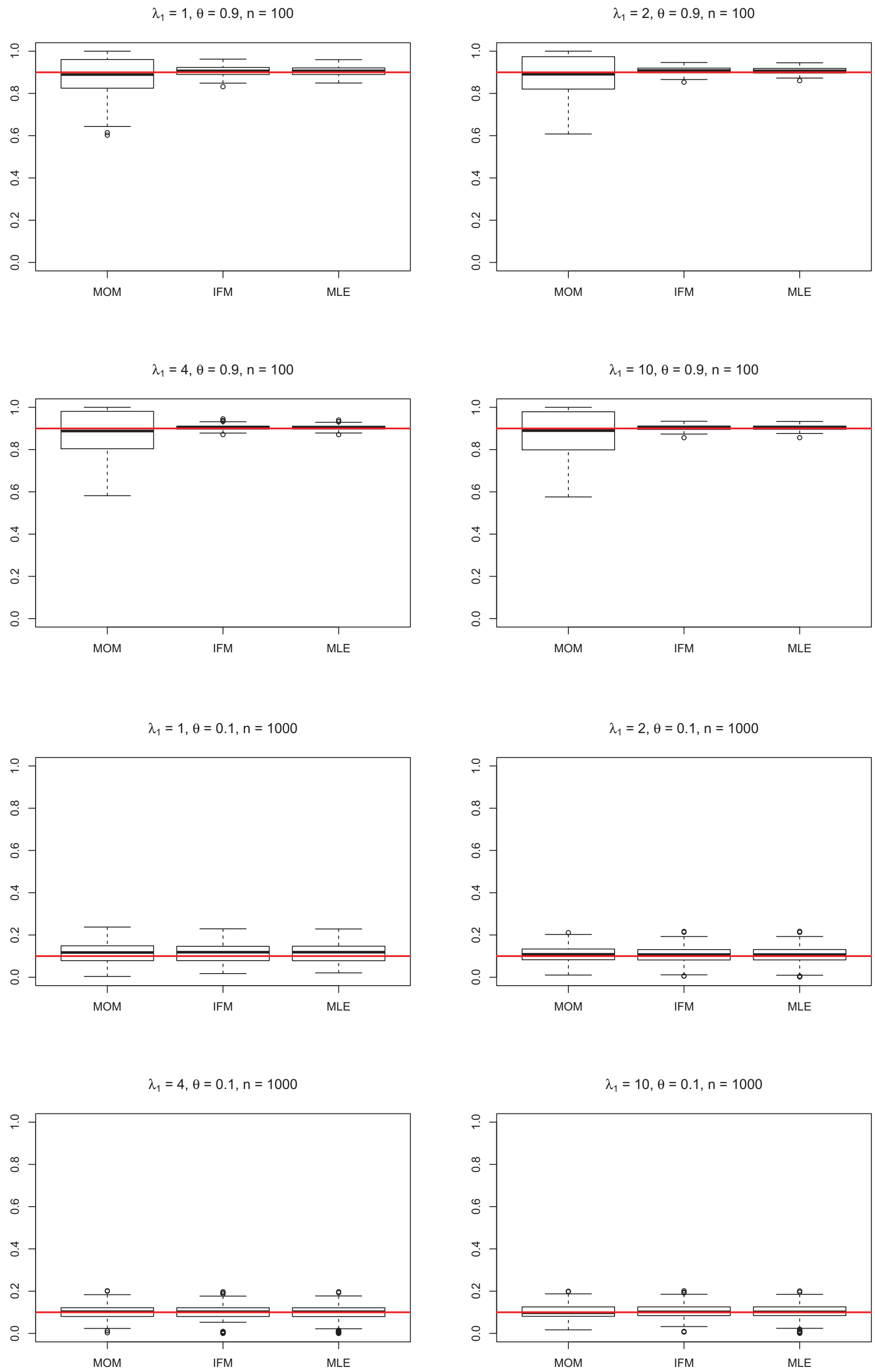

Figure 5.8: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

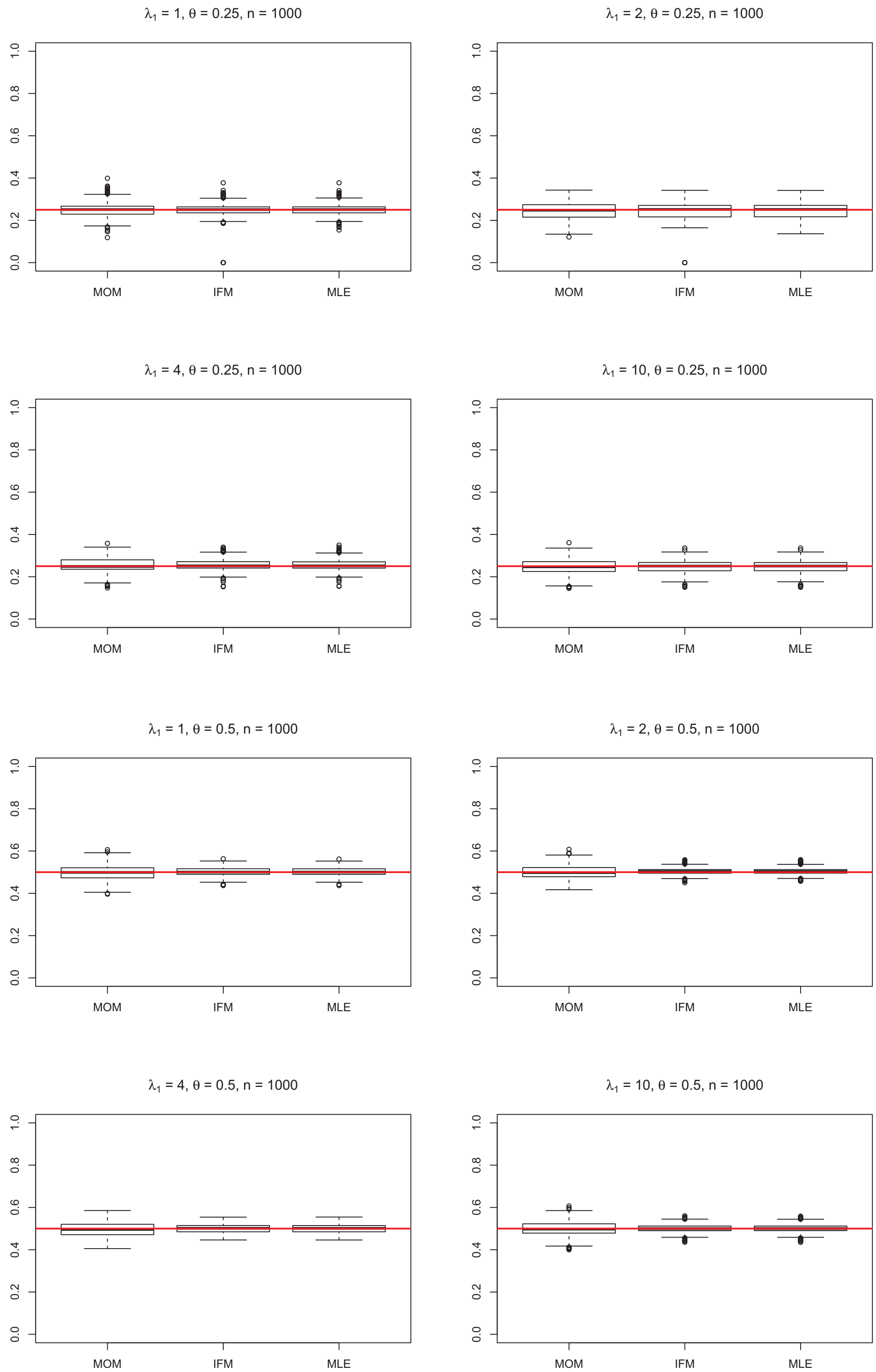

Figure 5.9: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

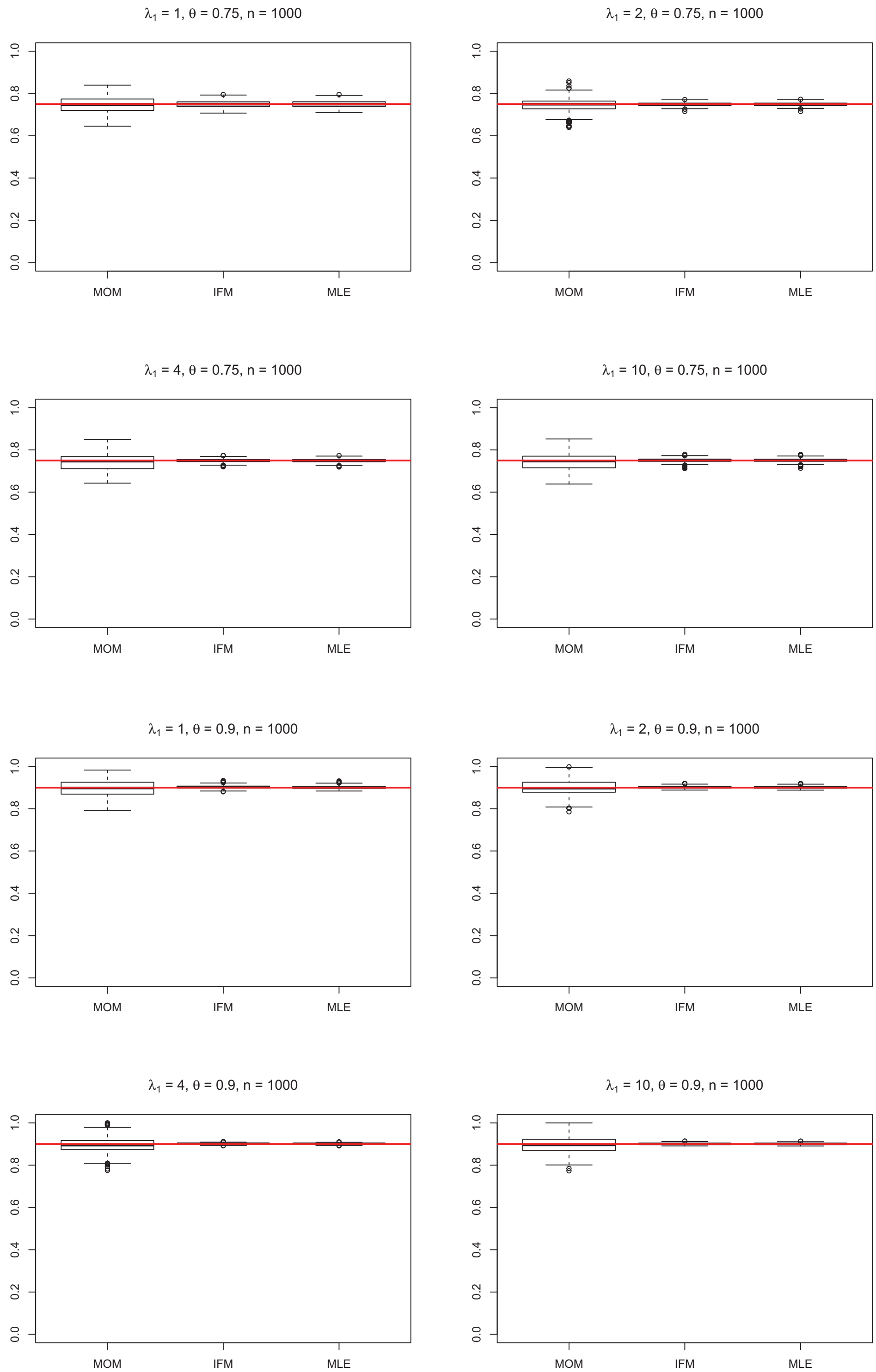

Figure 5.10: Estimation results for θ from the method of moments (left), inference function for margins (middle), and maximum likelihood estimation (right) in indicated scenario.

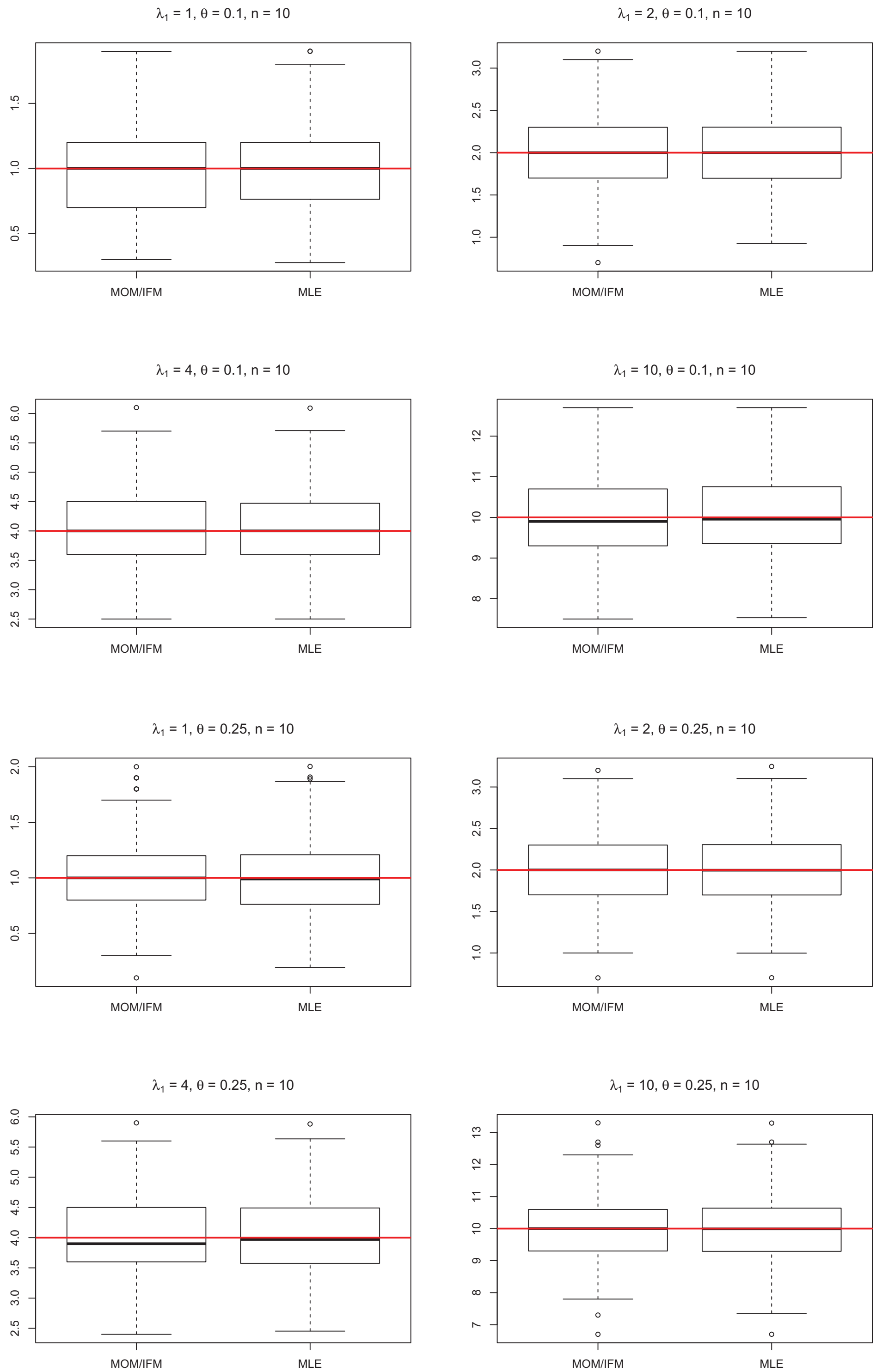

Figure 5.11: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

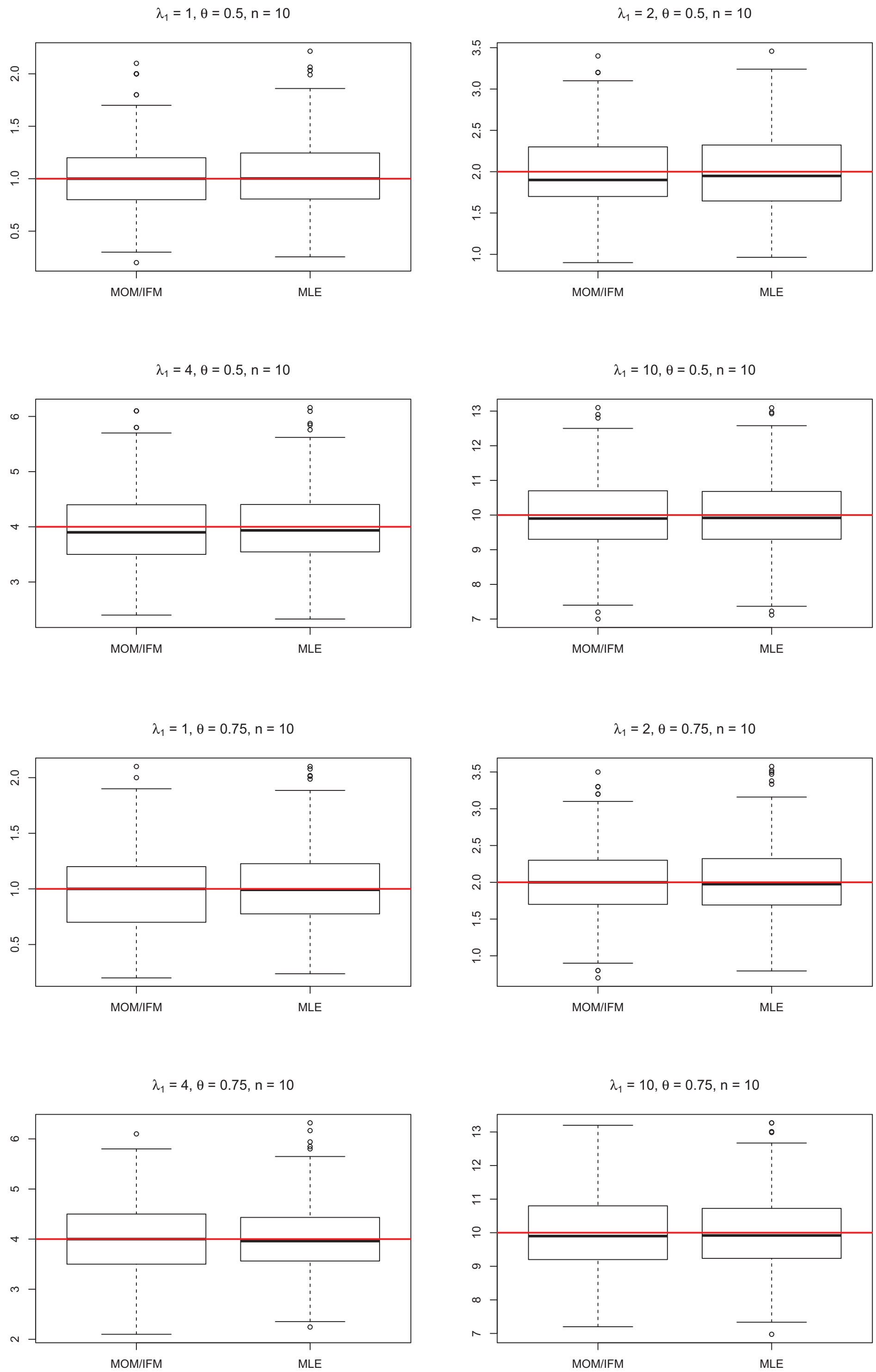

Figure 5.12: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

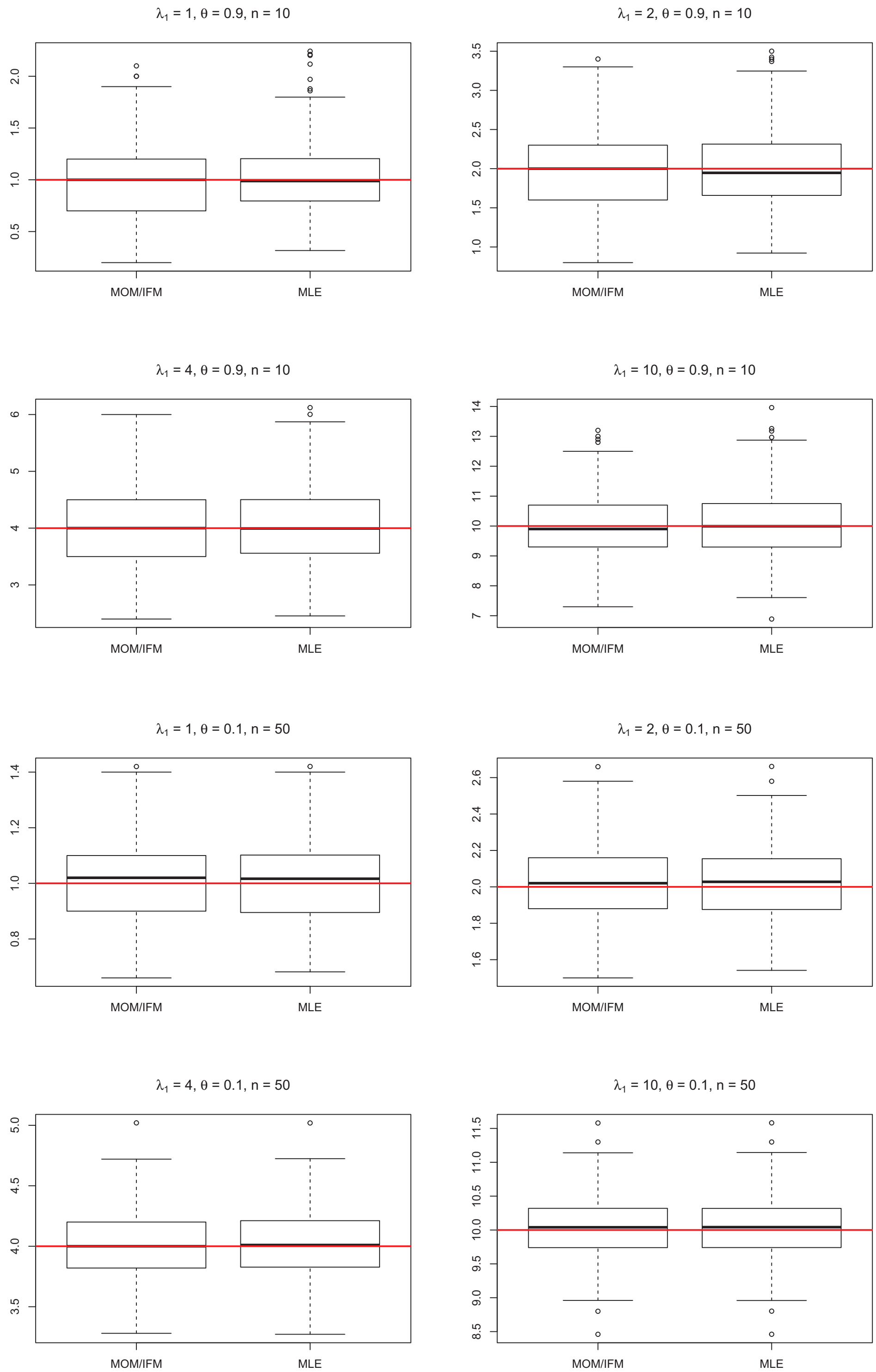

Figure 5.13: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

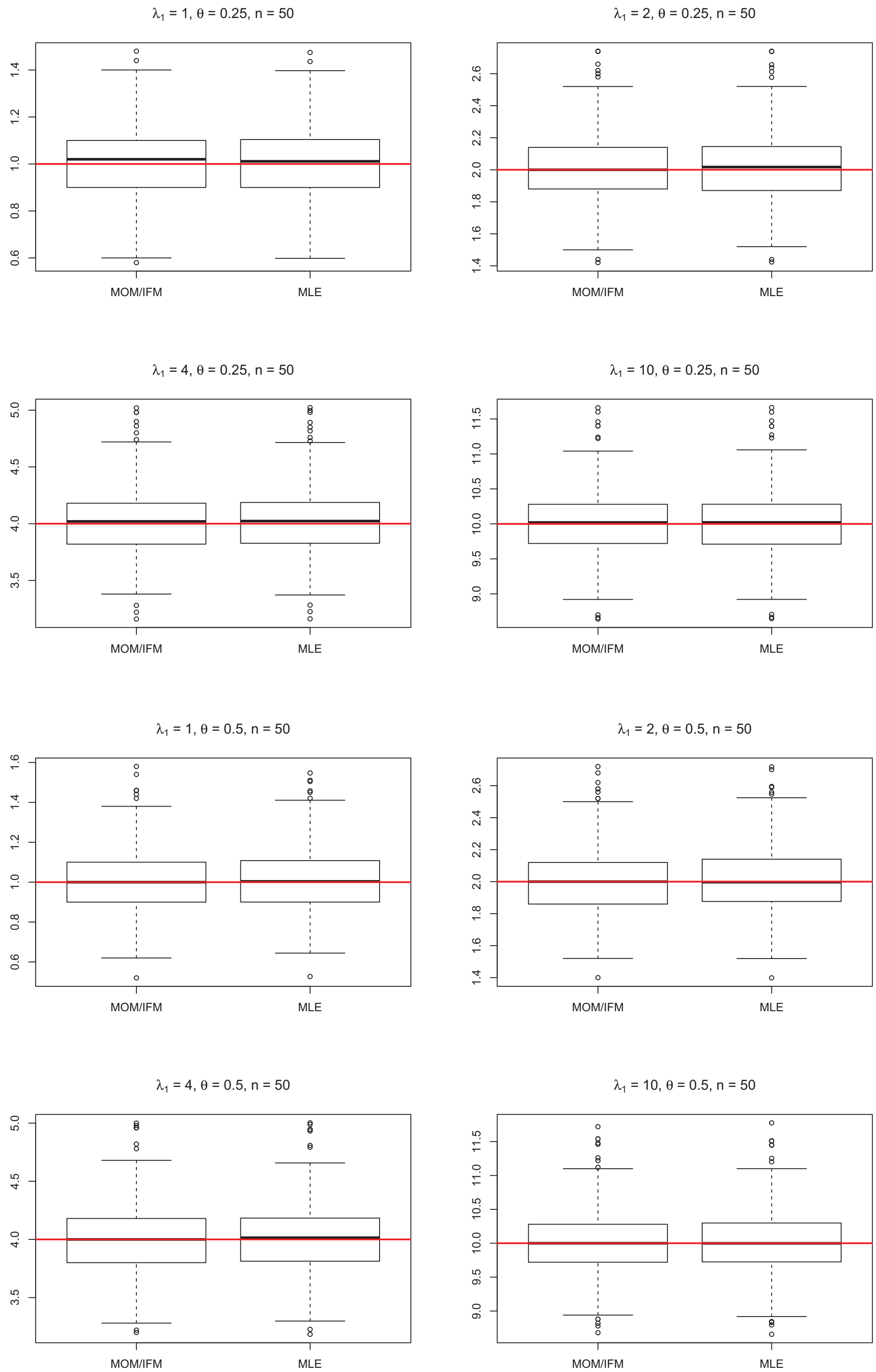

Figure 5.14: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

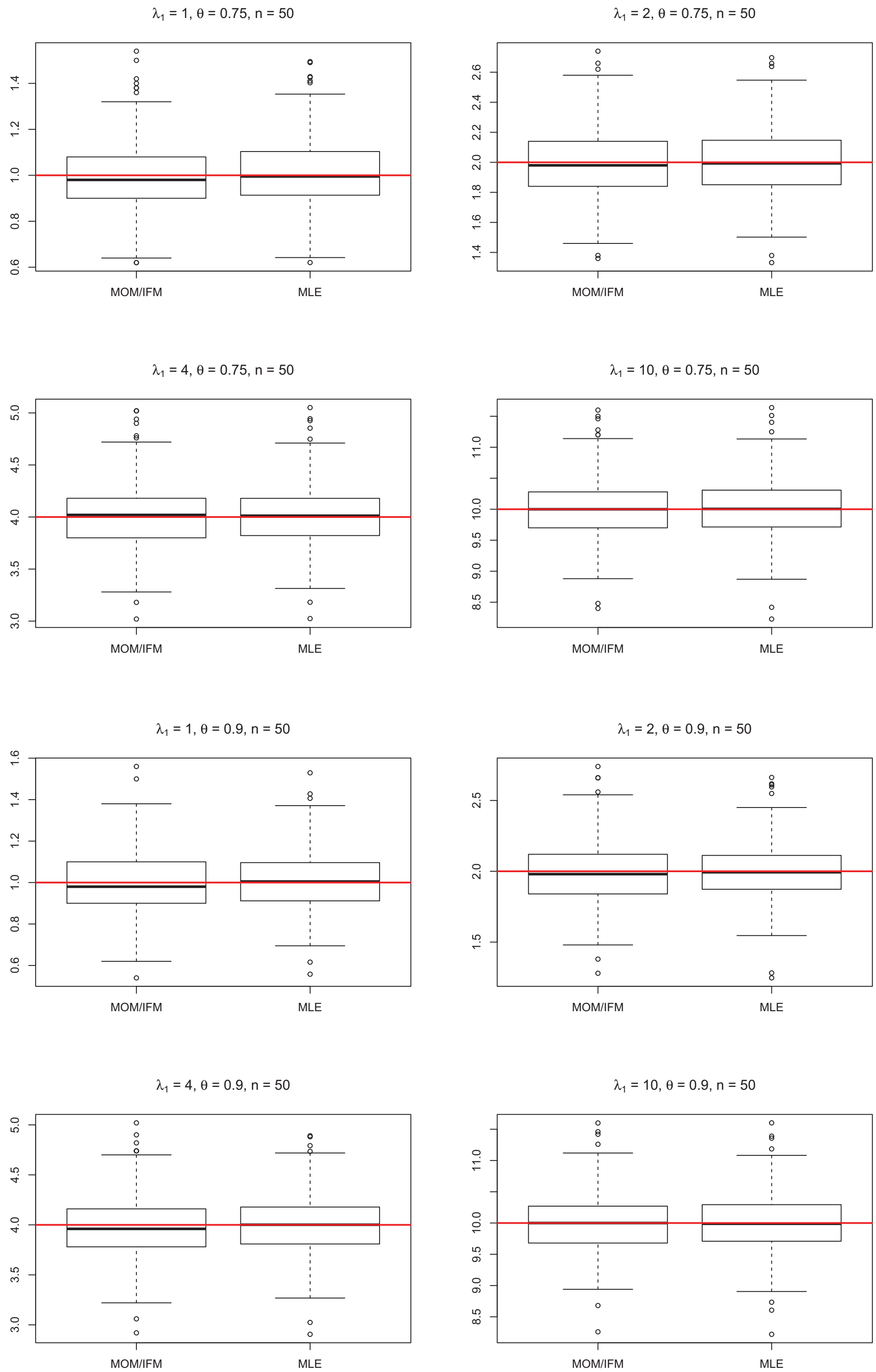

Figure 5.15: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

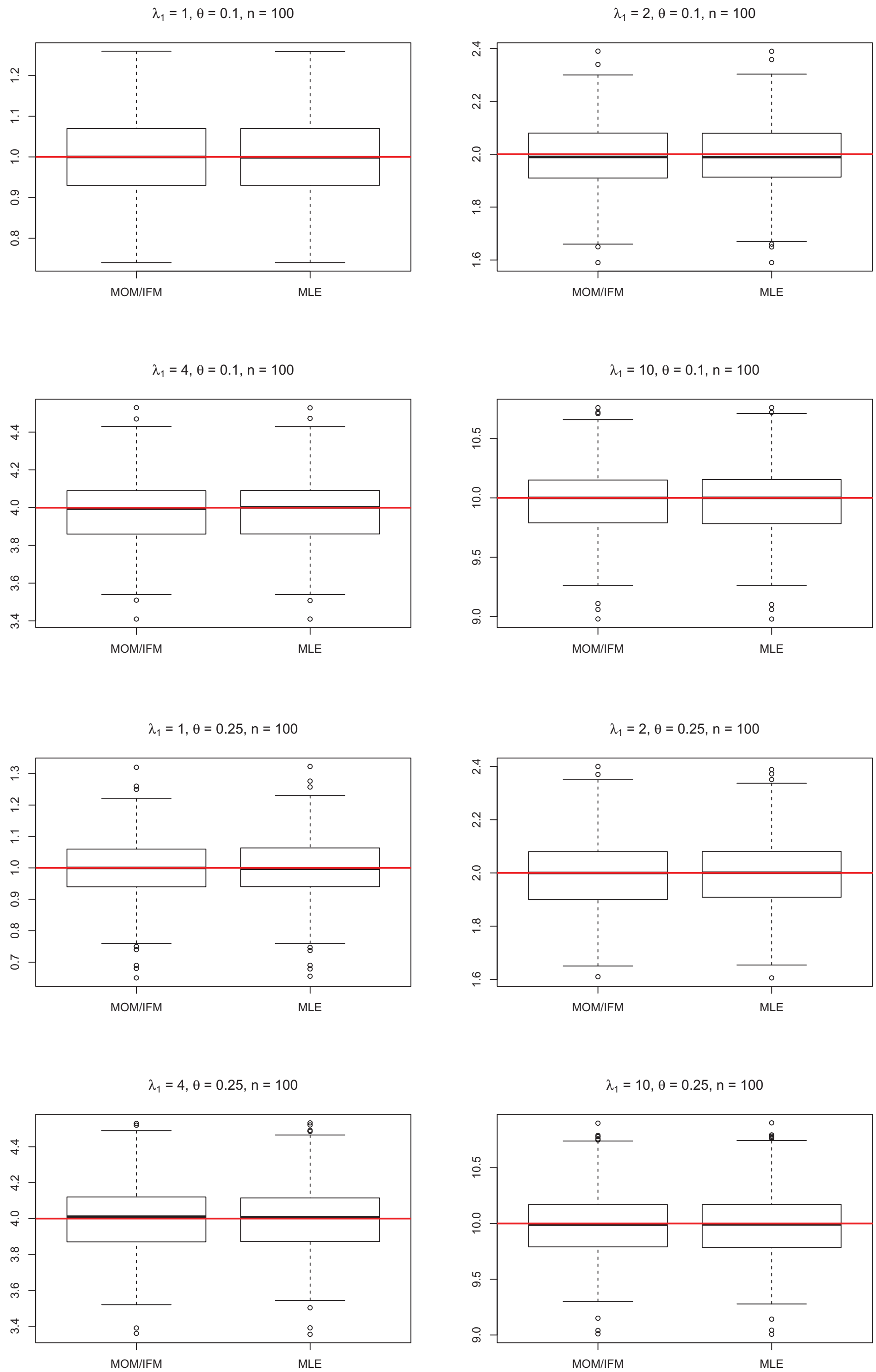

Figure 5.16: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

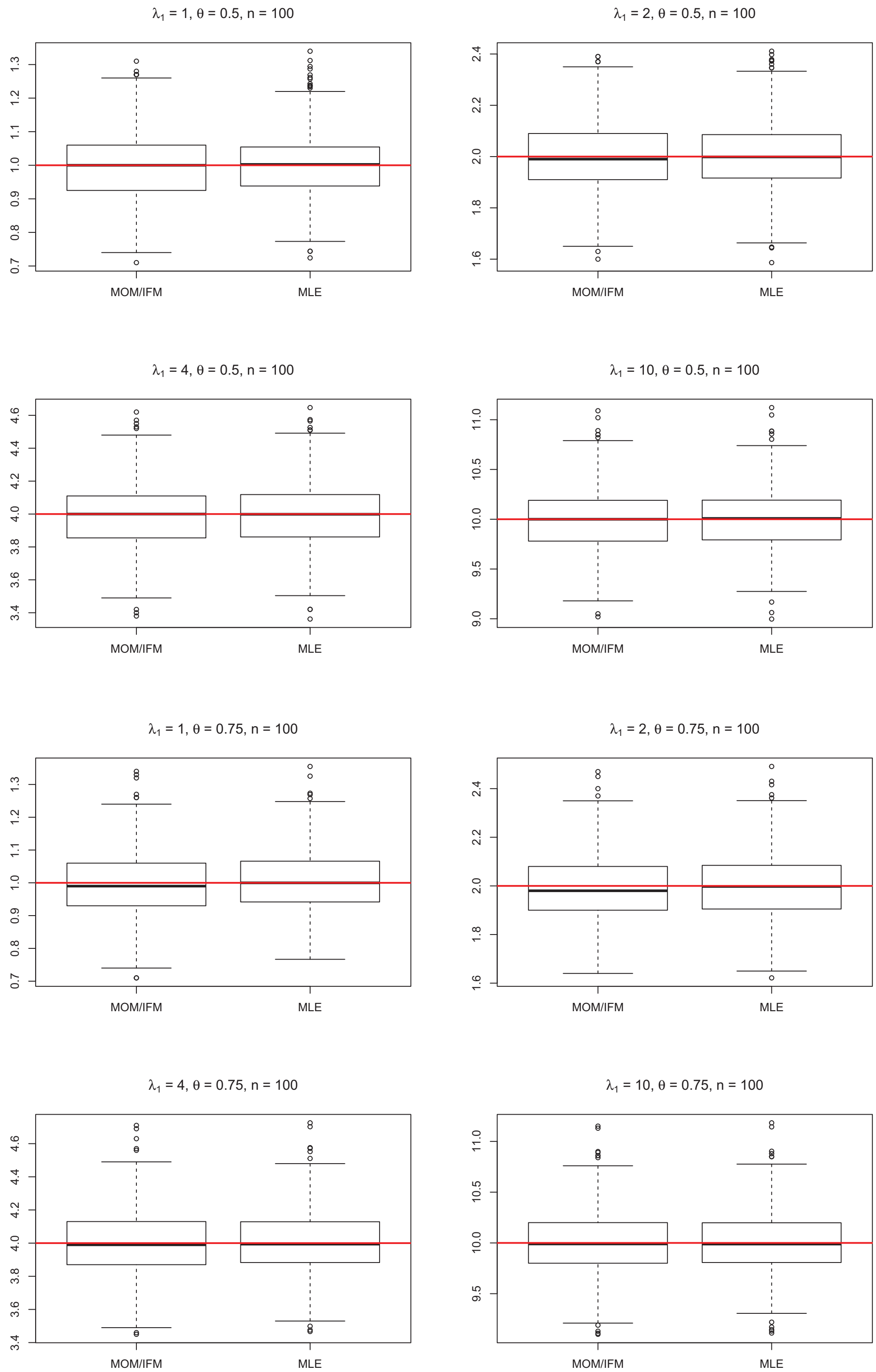

Figure 5.17: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

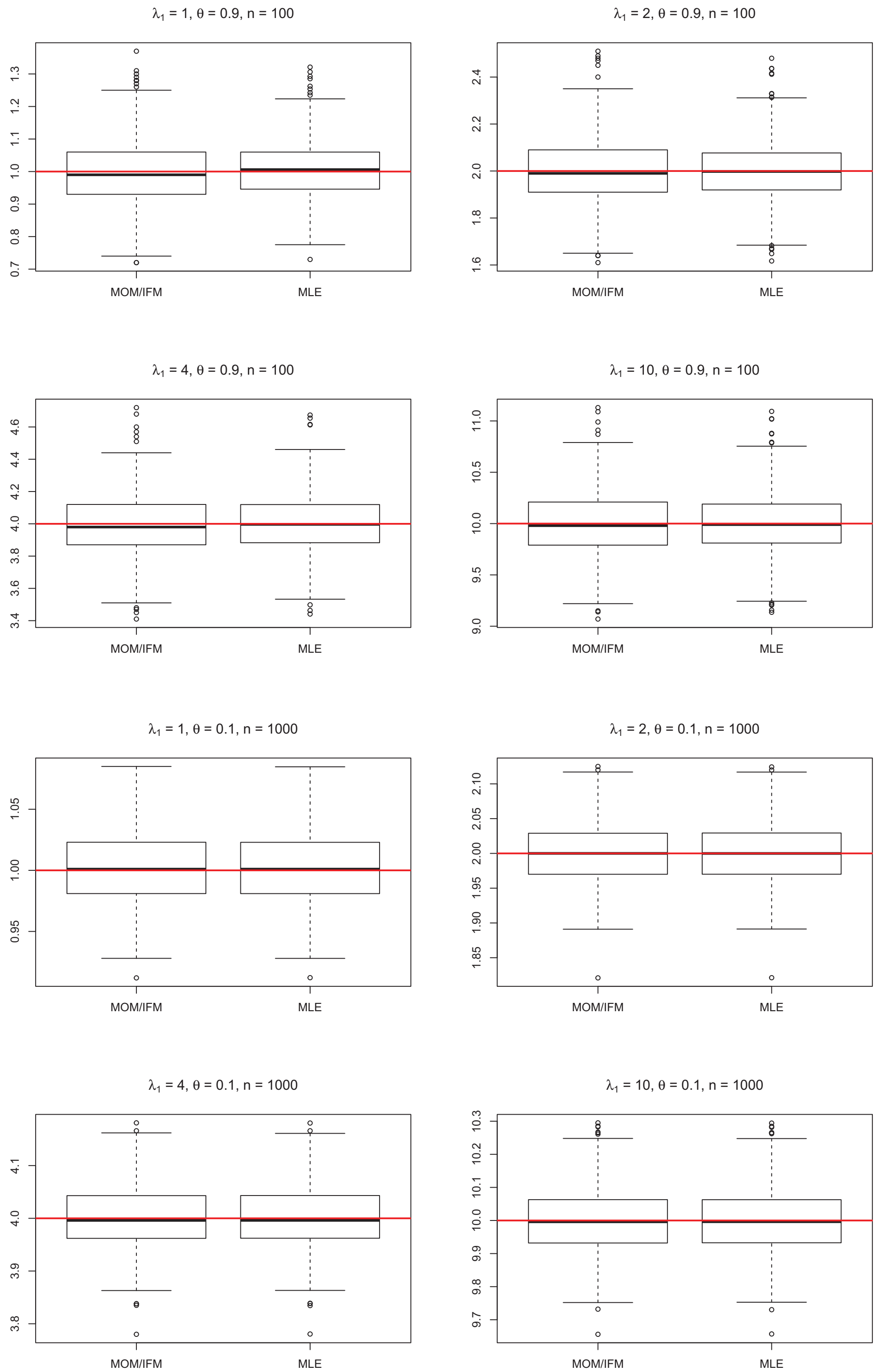

Figure 5.18: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

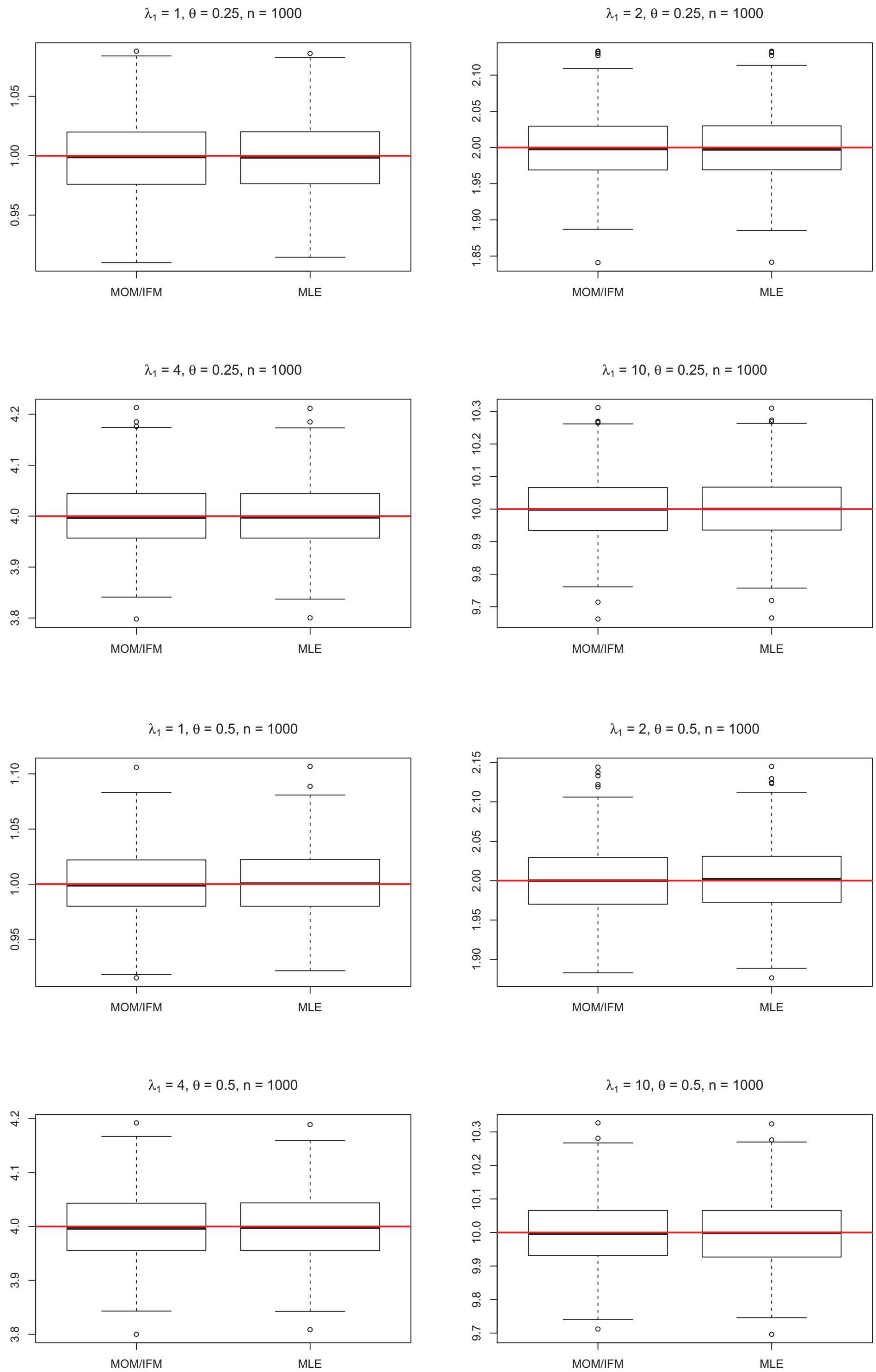

Figure 5.19: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

Table 5.1 shows the error rates (i.e., the percentage of iterations that resulted in errors) for each of the four possible starting values. These results are not surprising based on the behavior of the log-likelihood revealed in Figures 5.22 through 5.26. Indeed, if θ is initialized too far from its true value an error will ensue since the log-likelihood will yield $-\infty$. This explains why using $\theta_4^{(0)}$ always resulted in an error for both likelihood-based methods. Note that in this simulation study errors occur even when using the true value $\theta_3^{(0)} = 0.90$. This is due to the fact that rather than using the true values of Λ to initialize the optimization for the maximum likelihood approach, the IFM estimates $(\bar{X}_1, \bar{X}_2)$ were used as starting values for the marginal parameters.

Similarly to the first set of simulations, the majority of the errors occurred due to the starting value itself in which $\ell^{-}(\check{\Lambda}, \theta^{(0)}) = -\infty$. There were also a few instances where the error was due to "non-finite finite difference value". As was previously discussed, this error likely occurs when the parameter update approaches the boundary of the parameter space and gradient evaluations become infinite. Table 5.2 details the error occurrences by type for both the IFM and MLE approaches. Clearly, in the $\mathcal{BP}^{-}$ model, the estimation is relatively sensitive to the choice of starting value, which is not surprising given the shape of the log-likelihood. Figure 5.21 depicts the estimation results for the IFM and MLE methods, i.e., $\check{\theta}$ and $\hat{\theta}$, respectively, for the various starting values. The boxplots reveal that when the optimization procedures run without error, the estimates produced are indeed reasonable.

Table 5.1: Error rates by starting value.

Method	$\theta_1^{(0)}$	$\theta_2^{(0)}$	$\theta_3^{(0)}$	$\theta_4^{(0)}$
IFM	45.6%	0.6%	15.6%	100.0%
MLE	49.6%	8.8%	19.6%	100.0%

Table 5.2: Error occurrences by start value and type.

IFM	$\theta_1^{(0)}$	$\theta_2^{(0)}$	$\theta_3^{(0)}$	$\theta_4^{(0)}$
Error: initial value not finite	227	0	76	500
Error: non-finite finite difference value	1	3	2	0
MLE	$\theta_1^{(0)}$	$\theta_2^{(0)}$	$\theta_3^{(0)}$	$\theta_4^{(0)}$
Error: initial value not finite	227	0	76	500
Error: non-finite finite difference value	21	44	22	0

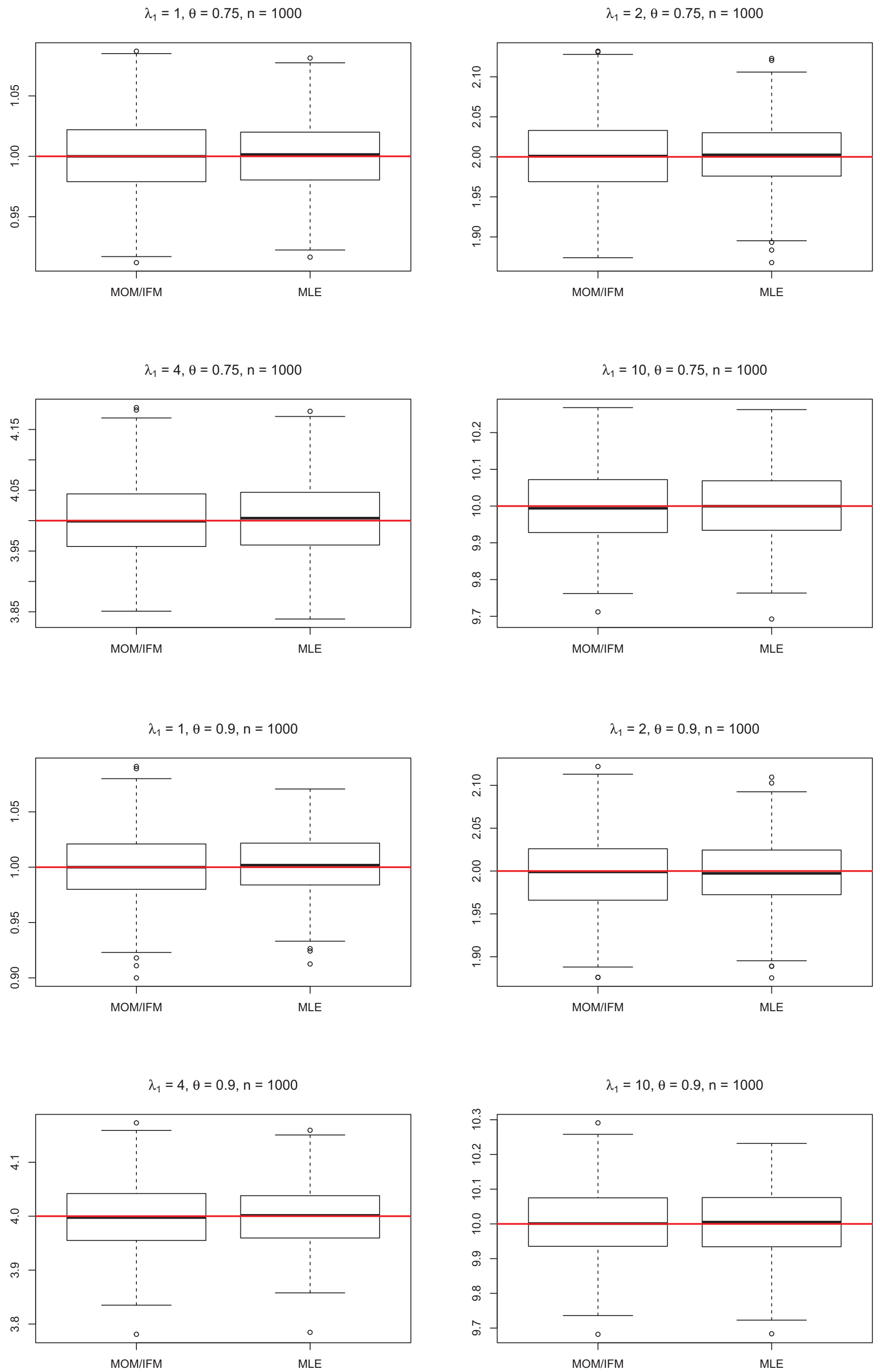

Figure 5.20: Estimation results for λ_1 from the method of moments / inference function for margins (left), and maximum likelihood estimation (right) in indicated scenario.

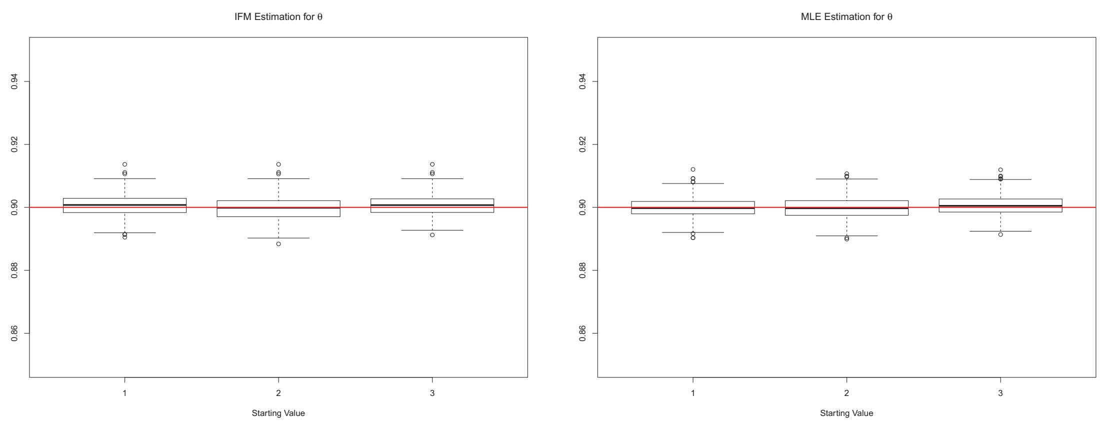

Figure 5.21: Estimation results for θ using different starting values. The IFM results are shown on the left and MLE on the right.

All of the simulations described thus far used the BFGS method in the optimization procedures for the IFM and MLE methods. Note that other optimization algorithms could be used, e.g., Nelder–Mead, Conjugate Gradient, etc. However, given that the majority of errors occurred due to inappropriate starting values, changing the algorithm would not decrease the incidences of errors drastically. Nonetheless, the second simulation study was rerun using the Nelder-Mead method for both the IFM and MLE optimization procedures. The Nelder–Mead method is derivative free and as such eliminated the error of "non-finite finite-difference value" that was previously observed when using the BFGS method. Accordingly, in this rerun of the second simulation study, the MLE method resulted in errors for 803 of the 2000 iterations, effectively rectifying the previously reported 87 errors due to the aforementioned issue. Obviously, all 803 errors in this setting were due to inappropriate starting values wherein $\ell^{-}(\tilde{\Lambda}, \theta^{(0)}) = -\infty$. When implementing the Nelder–Mead method for the IFM approach, however, each iteration resulted in a warning stating "one-dimensional optimization by Nelder–Mead is unreliable".

The log-likelihood plots for the proposed $\mathcal{BP}^{-}$ family revealed an interesting pattern. Similarly to the first set of simulations, the true underlying parameter values were varied; specifically $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and $\lambda_1 \in \{1, 2, 4, 10\}$, while holding λ_2 and n fixed at 5 and 1000, respectively. For each unique scenario $(\theta, \lambda_1, \lambda_2, n)$, a random sample was generated from the counter-monotonic shock model and the log-likelihood was evaluated over a grid of θ and λ_1 values in order to produce the plots shown in Figures 5.22 through 5.26. Specifically, the grid for θ consisted of the interval $(0.01, 0.99)$ with increments of 0.01 while that for λ_1 considered values from 0.1 to 15.0 by increments of 0.1.

Note that for each unique scenario, there were several instances where evaluating the log-likelihood yielded $-\infty$. In order to produce reasonable plots, the infinite values were replaced by the minimum finite value of $\ell^{-}(\Lambda, \theta)$ produced for that particular scenario. Across all combinations of θ and λ_1 values, the log-likelihood plots (shown in two dimensions as the three-dimensional versions were difficult to discern), there is a prominent cut-off were the log-likelihood rapidly tends to $-\infty$, implying implausible values of (θ, λ_1). Even more strikingly, this cut-off begins very close to the true underlying parameter values. Certainly, the particular behavior of the log-likelihood function for the $\mathcal{BP}^{-}$ model is what rendered the simulations previously considered rather challenging. Nonetheless, as shown in Table 5.3, the values of (θ, λ_1) yielding the maximum log-likelihood value (shown in the third column) are relatively close to the true values (θ_0, λ_{10}). Moreover, evaluating the log-likelihood at the true parameter values always yielded a finite value.

The particular pattern exhibited by the log-likelihood plots reflects the stringent behavior of counter-monotonic Poisson pairs. The plots show that as θ increases, a substantial region of the parameter space in terms of (θ, λ_1) is blotted out and rendered implausible as a result of the form that counter-monotonicity imposes. This region where the log-likelihood becomes negatively infinite diminishes as λ_1 increases and the Poisson-distributed margin begins to approach the Gaussian distribution. This phenomena was not observed in the comonotonic bivariate Poisson model. Seemingly, the dependence structure implied by comonotonicity is less restrictive than that resulting from counter-monotonicity.

Note that for $(Z_1, Z_2) \sim \mathcal{M}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\}$, the probability of observing $(z_1, z_2) = (0, 0)$ is always non-zero. This follows since

$$c_{\lambda,\theta}(0,0) = \min\{G_{\theta\lambda_1}(0), G_{\theta\lambda_2}(0)\} = \exp\{-\theta\max(\lambda_1, \lambda_2)\} > 0.$$

Thus, even when θ is large, the fact that there is a non-zero probability that the pair (X_1, X_2) is generated entirely from the independent components (Y_1, Y_2) causes the extremal values of θ to still be plausible. In other words, any observed pair (x_1, x_2) has probability

$$f_{\Lambda,\theta}(x_1, x_2) \geqslant g_{(1-\theta)\lambda_1}(x_1) g_{(1-\theta)\lambda_2}(x_2)$$

so that the rate at which the log-likelihood tends to $-\infty$ is regulated in a sense by the probability $g_{(1-\theta)\lambda_1}(x_1) g_{(1-\theta)\lambda_2}(x_2)$. That is why the log-likelihood plots for the $\mathcal{BP}$ model never exhibited $\ell(\Lambda, \theta) = -\infty$ over the grid of (θ, λ_1) values.

In the case of counter-monotonicity, the above reasoning is no longer valid. For $(Z_1, Z_2) \sim \mathcal{W}\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2)\}$, the probability of observing $(z_1, z_2) = (0, 0)$ is given by

$$c^{-}_{\Lambda,\theta}(0,0) = [\exp(-\theta\lambda_1) + \exp(-\theta\lambda_2) - 1]_{+}$$

which can indeed be zero for certain parameter values. As a result, when θ is large, the form of the log-likelihood is dominated by the behavior of the counter-monotonic components (Z_1, Z_2). The stringent form that $c^{-}_{\Lambda,\theta}$ imposes on the log-likelihood abruptly voids a pronounced region of the parameter space outlining the values of (θ, λ_1) that yield a probability of zero.

Table 5.3: Log-likelihood values evaluated over a grid of θ and λ_1 values for each of the 20 unique scenarios.

Scenario	True Values (θ_0, λ_{10})	ML Values $(\hat{\theta}, \hat{\lambda}_1)$	$\ell^{-}(\lambda_{10}, \lambda_{20}, \theta_0)$	$\ell^{-}(\hat{\lambda}_1, \lambda_{20}, \hat{\theta})$
1	$(0.10, 1)$	$(0.18, 1)$	-3466.202	-3464.772
2	$(0.10, 2)$	$(0.08, 2)$	-3923.468	-3923.385
3	$(0.10, 4)$	$(0.07, 4)$	-4234.624	-4234.397
4	$(0.10, 10)$	$(0.13, 10.1)$	-4729.466	-4728.594
5	$(0.25, 1)$	$(0.24, 1)$	-3511.272	-3511.224
6	$(0.25, 2)$	$(0.21, 2)$	-3932.784	-3932.366
7	$(0.25, 4)$	$(0.26, 3.9)$	-4246.989	-4246.727
8	$(0.25, 10)$	$(0.25, 10.2)$	-4725.729	-4723.912
9	$(0.50, 1)$	$(0.54, 1)$	-3409.696	-3408.501
10	$(0.50, 2)$	$(0.49, 2)$	-3805.287	-3805.263
11	$(0.50, 4)$	$(0.50, 3.9)$	-4188.063	-4187.654
12	$(0.50, 10)$	$(0.51, 9.9)$	-4642.339	-4641.032
13	$(0.75, 1)$	$(0.72, 1)$	-3250.518	-3248.494
14	$(0.75, 2)$	$(0.75, 2)$	-3516.678	-3516.678
15	$(0.75, 4)$	$(0.74, 4.1)$	-3898.940	-3898.134
16	$(0.75, 10)$	$(0.75, 9.9)$	-4350.046	-4348.816
17	$(0.90, 1)$	$(0.90, 1)$	-2956.340	-2956.340
18	$(0.90, 2)$	$(0.91, 2)$	-3153.144	-3151.248
19	$(0.90, 4)$	$(0.90, 4)$	-3468.668	-3468.668
20	$(0.90, 10)$	$(0.90, 10.1)$	-3965.380	-3964.658

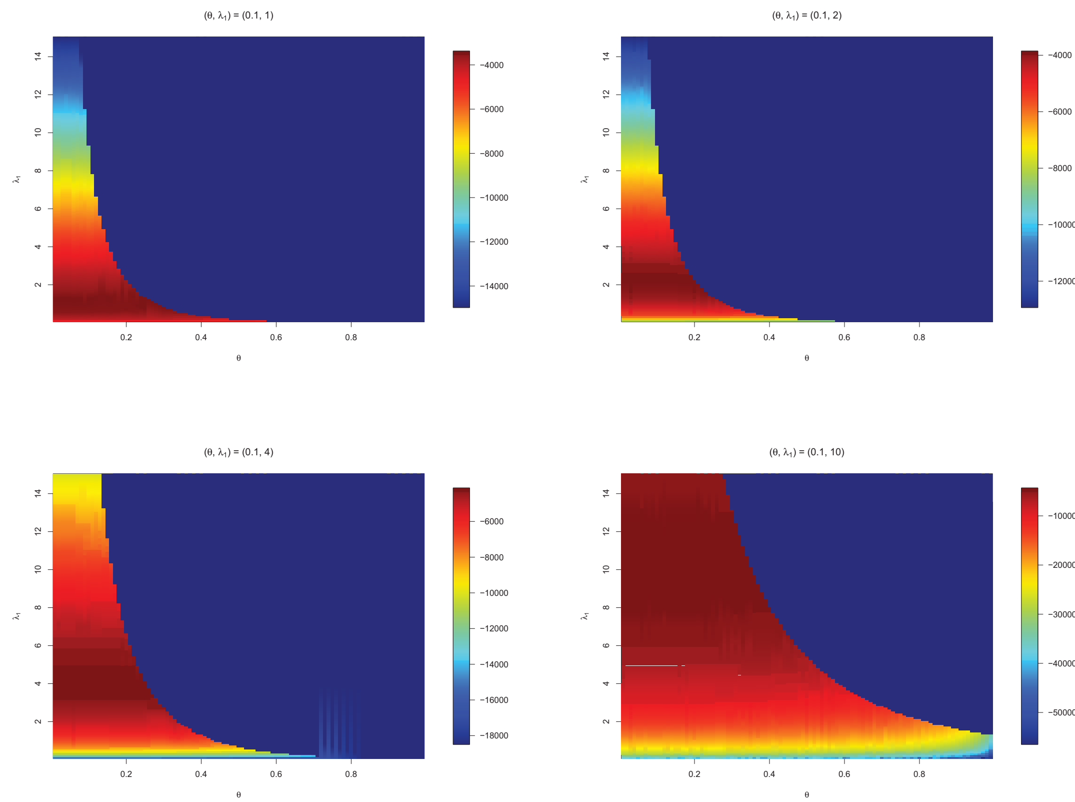

Figure 5.22: Log-likelihood plots for varying values of θ and λ_1, with $\lambda_2 = 5$ and $n = 1000$.

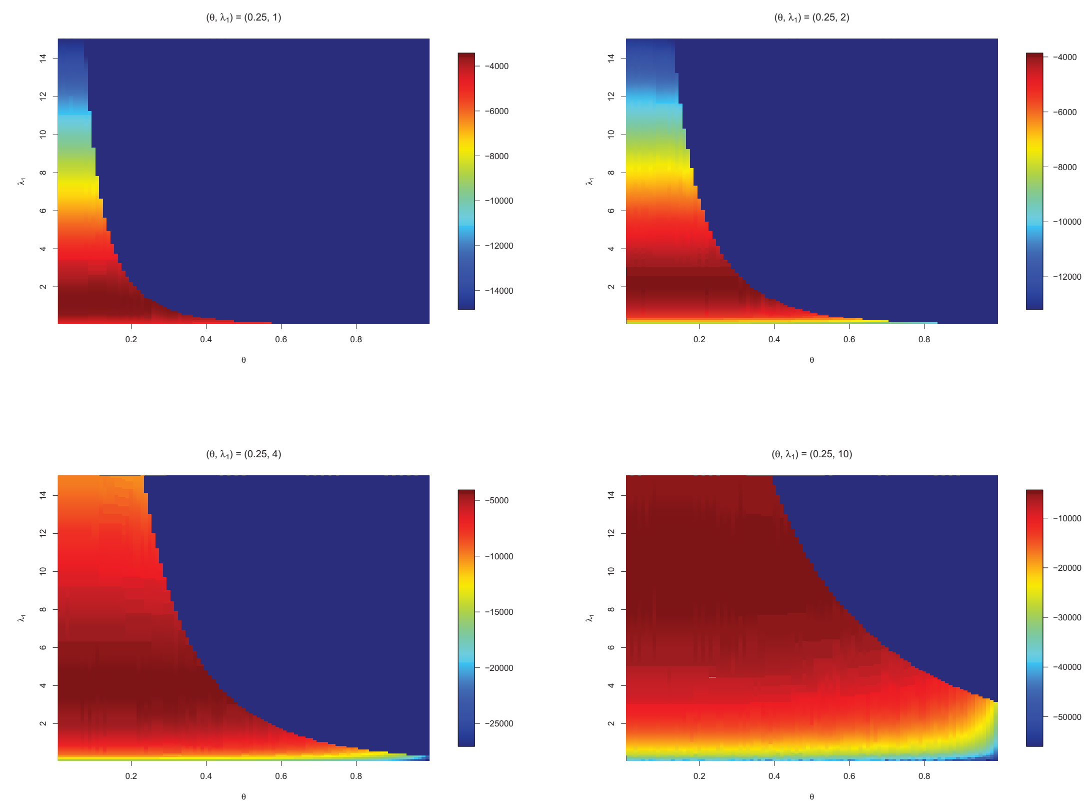

Figure 5.23: Log-likelihood plots for varying values of θ and λ_1, with $\lambda_2 = 5$ and $n = 1000$.

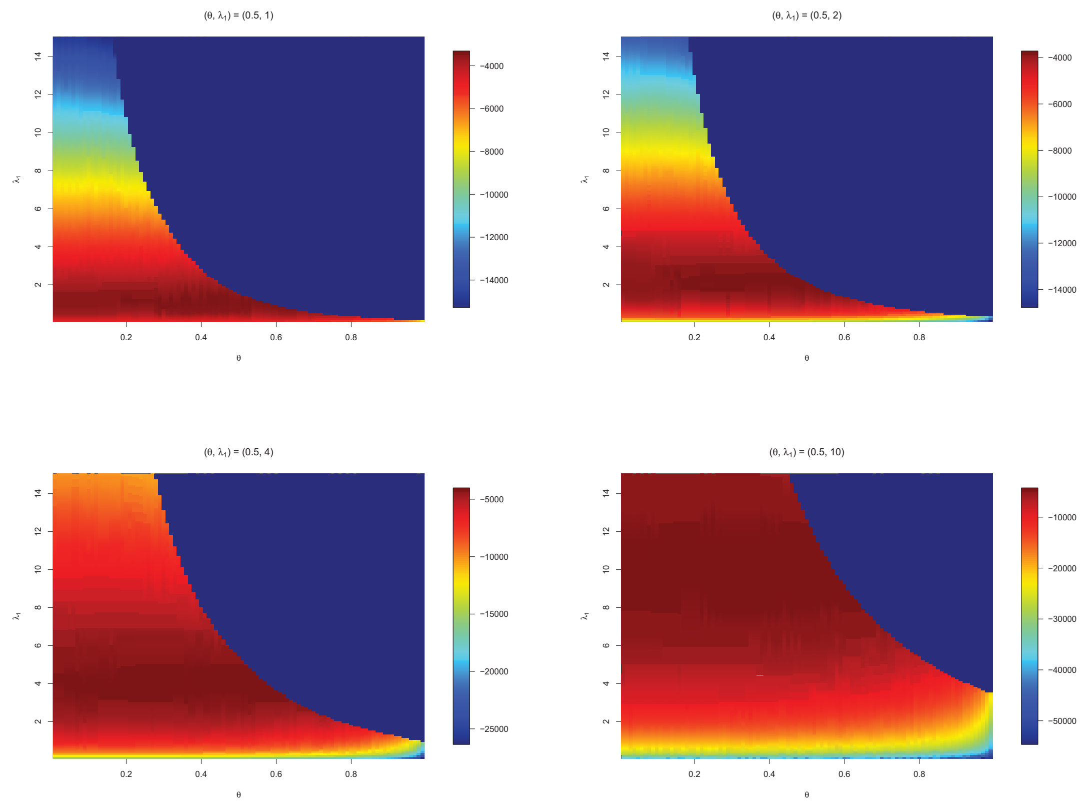

Figure 5.24: Log-likelihood plots for varying values of θ and λ_1, with $\lambda_2 = 5$ and $n = 1000$.

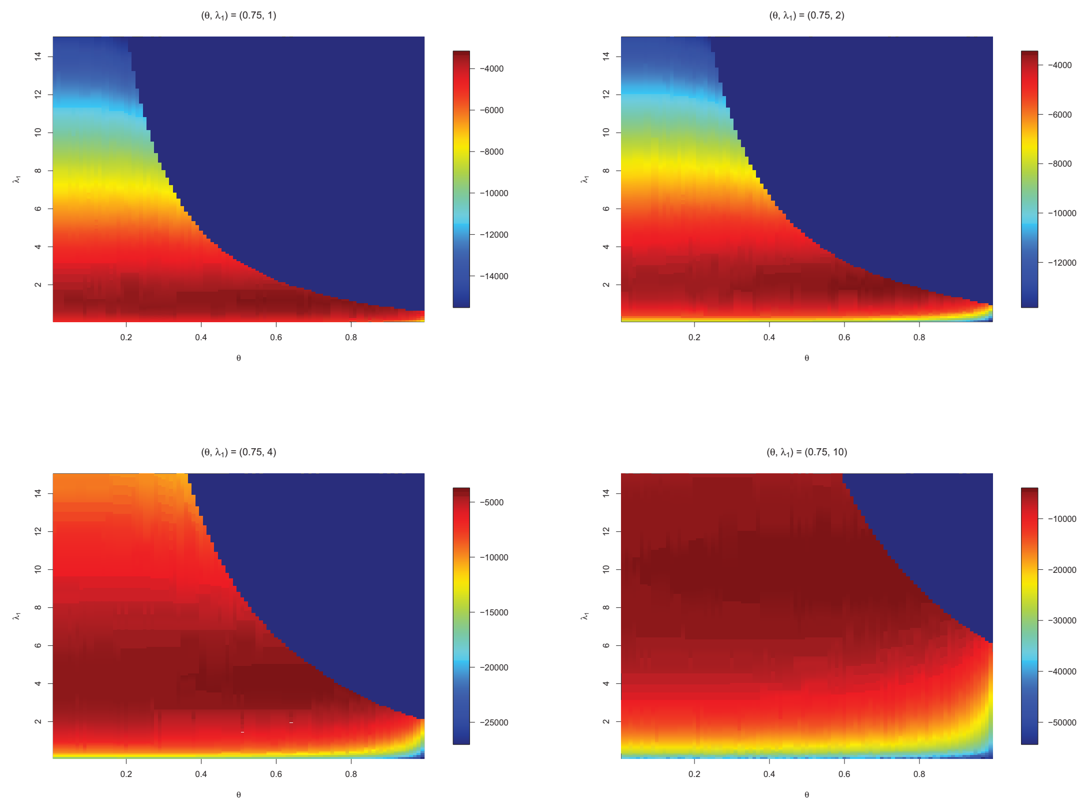

Figure 5.25: Log-likelihood plots for varying values of θ and λ_1, with $\lambda_2 = 5$ and $n = 1000$.

5.5 Data illustration

As an illustration of the proposed $\mathcal{BP}^{-}(\Lambda, \theta)$ class, the counter-monotonic shock model was applied to NHL data freely available online (http://www.nhl.com/stats/). The NHL provides several measures from each hockey game. In this particular application, the bivariate Poisson pairs (X_1, X_2) consist of the goals against and goals for the *Canadiens de Montréal* at each game for the years 2010 to 2017, for a total of 566 observations.

Innately, the number of goals scored for or against a hockey team could follow a Poisson distribution as such variables take on values in $\mathbb{N}$. The assumption of Poisson margins can be readily checked using various statistical tests. In particular, a Chi-squared goodness of fit test yields p-values of 0.11 and 0.61 for the goals against and for, respectively, both indicating evidence of Poisson margins. For both components, the sample mean and variance roughly coincide; specifically $\bar{X}_1 = 2.55$, $S_1^2 = 2.59$, $\bar{X}_2 = 2.63$, $S_2^2 = 2.64$. This is in line with the property of the Poisson distribution wherein the mean and variance are both equal to λ. QQ plots, shown in Figure 5.27,

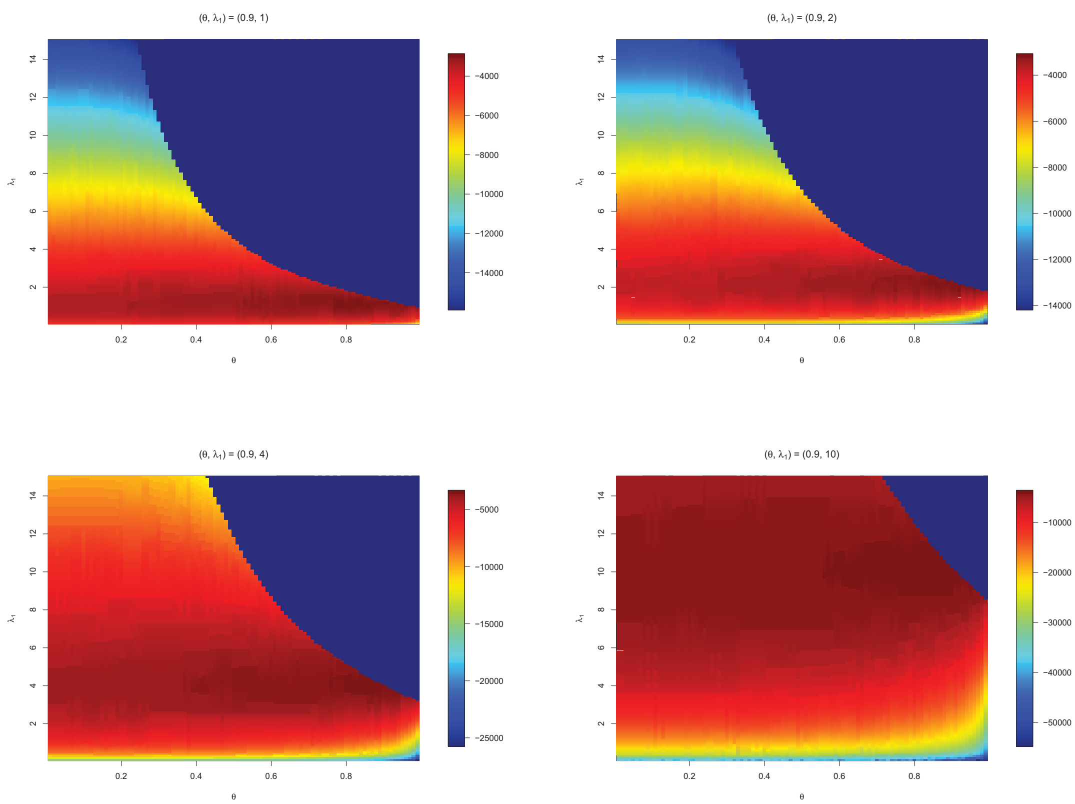

Figure 5.26: Log-likelihood plots for varying values of θ and λ_1, with $\lambda_2 = 5$ and $n = 1000$.

provide further evidence of Poisson margins.

As previously mentioned, the *Canadiens de Montréal* scored on average $\bar{X}_2 = 2.63$ goals per game, while an average of $\bar{X}_1 = 2.55$ goals were scored against the team. Conceivably, the variables X_1 and X_2 should exhibit a negative association: When the team is playing poorly the goals against will tend to be large and the goals for will be lower while the reverse will hold if the team is having a good game. Indeed, this phenomenon is observed in the data: the sample covariance is found to be $S_{12} = -0.39$. If the sample means are used as estimates of the marginal Poisson rates, the implied correlation is given by $R_{12} = S_{12}/\sqrt{\bar{X}_1\bar{X}_2} = -0.15$. Given the sample variance measures, R_{12} is essentially equivalent to the sample correlation measure given by $S_{12}/\sqrt{S_1^2 S_2^2} = -0.15$.

Obviously, the fact that the pair (X_1, X_2) exhibits a negative association renders the classical bivariate Poisson model inappropriate. Of course, the $\mathcal{BP}^-$ model can accommodate the negative dependence. The various estimation frameworks outlined in Section 5.3 were applied to the NHL data; the three sets of estimates are summarized in Table 5.4 and provide the estimate along with 95% bootstrap confidence intervals. The results for the IFM and MLE approaches used the MM estimates as starting values. Note that the IFM values given in Table 5.4 are those resulting from directly optimizing $\ell^-(\check{\Lambda}, \theta)$. Applying the EM algorithm to the pseudo log-likelihood $\ell^-(\check{\Lambda}, \theta)$ yielded a similar estimate of $\check{\theta} = 0.26$ with an applied correlation of $\rho = -0.17$. All three estimation procedures yield very similar results. Moreover, the observed sample correlation $R_{12} = -0.15$ is contained in each of the 95% bootstrap confidence intervals obtained for the three estimation methods.

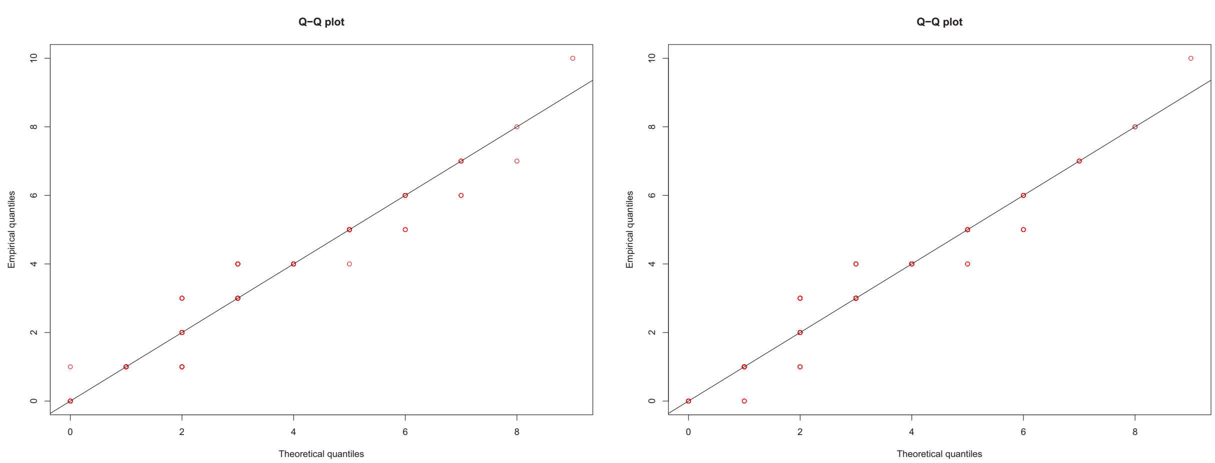

Figure 5.27: QQ-plots assessing the marginal Poisson assumption for variables X_1 (left) and X_2 (right) in the data illustration. Theoretical and empirical quantiles are on the x- and y-axis, respectively.

Note that out of a total of 1000 bootstrap replications, there were 125 instances where the optimization procedures produced errors for both the IFM and MLE approaches. There was also one case where the MM estimate $\tilde{\theta} = NA$ since the bootstrap sample covariance was found to be positive. The bootstrap confidence intervals provided in Table 5.4 omit these 126 problematic replications.

Note that in the instances where the likelihood-based methods failed, the method of moments estimate for the dependence parameter tended to be larger than average, ranging from 0.26 to 0.36. This subset of the MM estimates $\tilde{\theta}$ represent roughly those exceeding the 75th percentile. In light of the difficulties in optimizing the log-likelihood exhibited in the simulations in Section 5.4, it is not surprising that starting the IFM and MLE estimation procedures at $\tilde{\theta}$ in these instances produced errors. Of these 125 problematic replications, most (124) of the errors occurred due to inappropriate starting values while once the error ensued due to "non-finite finite-difference value". The latter error occurred when $\tilde{\theta} = 0.28$.

The effects of the starting value for θ was also tested. The results are shown in Table 5.5. For lower values of $\theta^{(0)} < \tilde{\theta}$, both the IFM and MLE procedures yielded similar results and seemed to be unaffected by the choice of starting value. However, when the starting value becomes much larger than $\tilde{\theta}$, both likelihood-based estimation methods fail. This is not surprising based on the results from the simulation studies detailed in Section 5.4.

Table 5.4: Estimation results for the counter-monotonic model. Estimates are given with 95% bootstrap confidence intervals in parenthesis.

	λ_1	λ_2	θ	ρ
MM	$2.55\ (2.42, 2.67)$	$2.63\ (2.50, 2.76)$	$0.24\ (0.16, 0.33)$	$-0.15\ (-0.22, -0.06)$
IFM	$2.55\ (2.42, 2.67)$	$2.63\ (2.50, 2.76)$	$0.25\ (0.19, 0.33)$	$-0.16\ (-0.22, -0.10)$
MLE	$2.55\ (2.42, 2.67)$	$2.63\ (2.50, 2.75)$	$0.25\ (0.19, 0.33)$	$-0.16\ (-0.22, -0.10)$

Table 5.5: Estimation results using different starting values for θ.

Starting Value	IFM	MLE
0.10	0.25	0.25
0.25	0.25	0.25
0.50	NA	NA
0.75	NA	NA
0.90	NA	NA

6

Extension to Higher Dimensions

6.1 Introduction

The formulation of the proposed bivariate Poisson model for positive dependence introduced in Chapter 4 is based on the notion of a comonotonic shock. This construction extends readily to higher dimensions, leading to a multivariate comonotonic shock model. This proposed multivariate Poisson model is similar to the bivariate model in many regards. Notably, the pairwise correlations implied by the multivariate comonotonic shock model span the full spectrum of dependence from 0 in the case of independence to $\rho_{\max}$ when the model attains the upper Fréchet–Hoeffding bound. In this sense, the proposed multivariate Poisson model allows for greater flexibility in the implied correlation structure than the classical model described in, e.g., Karlis (2003). The distributional properties and estimation methods in the proposed multivariate Poisson model extend directly from the bivariate setting. Accordingly, much of the content developed here is somewhat repetitive of Chapter 4 and will thus be kept relatively brief.

6.2 The proposed model construction

Let $(X_1, \ldots, X_d)$ be a random vector with Poisson margins where each component can be written as a sum of Poisson random variables viz.

$$X_1 = Y_1 + Z_1, \quad \ldots, \quad X_d = Y_d + Z_d. \tag{6.1}$$

In this formulation, $Y_1, \ldots, Y_d$ are mutually independent Poisson random variables with respective rates $(1-\theta)\lambda_1, \ldots, (1-\theta)\lambda_d$, which are also independent of the comonotonic shock vector $(Z_1, \ldots, Z_d) \sim \mathcal{M}\{\mathcal{P}(\theta\lambda_1), \ldots, \mathcal{P}(\theta\lambda_d)\}$. This comonotonic shock multivariate Poisson model will be denoted as

$$(X_1, \ldots, X_d) \sim \mathcal{MP}_d(\Lambda, \theta),$$

where $\Lambda = (\lambda_1, \ldots, \lambda_d) \in (0, \infty)^d$ and $\theta \in (0, 1)$. As before, the limiting cases $\theta = 0$ and $\theta = 1$ can also be encompassed in the definition with the convention that $W \sim \mathcal{P}(0) \Leftrightarrow W \equiv 0$. It then follows that the components, $X_1, \ldots, X_d$ are mutually independent when $\theta = 0$ and comonotonic, i.e., perfectly positively dependent, when $\theta = 1$. In terms of independent standard uniform random variables $V_1, \ldots, V_d, U$, the formulation in (6.1) is equivalent to

$$
\begin{aligned}
X_1 &= G^{-1}_{(1-\theta)\lambda_1}(V_1) + G^{-1}_{\theta\lambda_1}(U), \\
&\vdots \\
X_d &= G^{-1}_{(1-\theta)\lambda_d}(V_d) + G^{-1}_{\theta\lambda_d}(U).
\end{aligned}
$$

As mentioned above, if $\theta = 0$, the random vector $(X_1, \ldots, X_d)$ will have independent components so that

$$
\Pr(X_1 \leqslant x_1, \ldots, X_d \leqslant x_d) = \prod_{i=1}^{d} G_{\lambda_i}(x_i).
$$

In contrast, when $\theta = 1$, the distribution of $(X_1, \ldots, X_d)$ is the upper Fréchet–Hoeffding bound, i.e., for all $x_1, \ldots, x_d \in \mathbb{N}$,

$$
\Pr(X_1 \leqslant x_1, \ldots, X_d \leqslant x_d) = \min\left\{G_{\lambda_1}(x_1), \ldots, G_{\lambda_d}(x_d)\right\}.
$$

For any $\theta \in (0, 1)$, the joint probability distribution of the random vector $(X_1, \ldots, X_d) \sim \mathcal{MP}_d(\Lambda, \theta)$, denoted by $F_{\Lambda,\theta}$, is given by

$$
\begin{aligned}
F_{\Lambda,\theta}(x_1, \ldots, x_d) &= \Pr(X_1 \leqslant x_1, \ldots, X_d \leqslant x_d) \\
&= \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \Pr(X_1 \leqslant x_1, \ldots, X_d \leqslant x_d \mid Y_1 = x_1 - z_1, \ldots, Y_d = x_d - z_d) \\
&\qquad \times \Pr(Y_1 = x_1 - z_1, \ldots, Y_d = x_d - z_d) \\
&= \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \left[\Pr(Z_1 \leqslant z_1, \ldots, Z_d \leqslant z_d) \prod_{j=1}^{d} \Pr(Y_j = x_j - z_j) \right].
\end{aligned}
$$

The above then further simplifies to

$$
F_{\Lambda,\theta}(x_1, \ldots, x_d) = \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \left[\min\left\{G_{\theta\lambda_1}(z_1), \ldots, G_{\theta\lambda_d}(z_d)\right\} \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) \right] \tag{6.2}
$$

Note that the joint probability distribution can also be derived by conditioning on the underlying uniform random variable U that generates the comonotonic shock random vector. This approach leads to the following expression:

$$F_{\Lambda,\theta}(x_1,\ldots,x_d) = \int_0^1 \Pr(X_1 \leqslant x_1,\ldots,X_d \leqslant x_d \mid U=u) \times f_U(u)du$$
$$= \int_0^1 \Pr\{Y_1 \leqslant x_1 - G^{-1}_{\theta\lambda_1}(u),\ldots,Y_d \leqslant x_d - G^{-1}_{\theta\lambda_d}(u)\}du.$$

Thus, an equivalent representation of (6.2) is given by

$$\Pr(X_1 \leqslant x_1,\ldots,X_d \leqslant x_d) = \int_0^1 \prod_{j=1}^d G_{(1-\theta)\lambda_j}\{x_j - G^{-1}_{\theta\lambda_j}(u)\}du.$$

The joint probability mass function for $(X_1,\ldots,X_d) \sim \mathcal{MP}_d(\Lambda,\theta)$, denoted as $f_{\Lambda,\theta}$, can be derived by differencing the cumulative distribution function as

$$f_{\Lambda,\theta}(x_1,\ldots,x_d) = \Pr(X_1 = x_1,\ldots,X_d = x_d) = \sum (-1)^{v(c)} F_{\Lambda,\theta}(c_1,\ldots,c_d),$$

where the sum is taken over all $c_j \in \{x_j - 1, x_j\}$, with $x_j \in \mathbb{N}, j \in \{1,\ldots,d\}$, and $v(c) = \#\{j : c_j = x_j - 1\}$. An equivalent expression for $f_{\Lambda,\theta}$ can be obtained by conditioning:

$$f_{\Lambda,\theta}(x_1,\ldots,x_d) = \sum_{z_1=0}^{x_1}\cdots\sum_{z_d=0}^{x_d} \Pr(X_1 = x_1,\ldots,X_d = x_d \mid Z_1 = z_1,\ldots,Z_d = z_d)$$
$$\times \Pr(Z_1 = z_1,\ldots,Z_d = z_d)$$
$$= \sum_{z_1=0}^{x_1}\cdots\sum_{z_d=0}^{x_d} \Pr(Y_1 = x_1 - z_1,\ldots,Y_d = x_d - z_d)\Pr(Z_1 = z_1,\ldots,Z_d = z_d)$$
$$= \sum_{z_1=0}^{x_1}\cdots\sum_{z_d=0}^{x_d} \left\{\prod_{j=1}^d g_{(1-\theta)\lambda_j}(x_j - z_j)\right\}\Pr(Z_1 = z_1,\ldots,Z_d = z_d).$$

By definition, the comonotonic random vector $(Z_1,\ldots,Z_d)$ has joint probability mass function given by

$$\Pr(Z_1 = z_1,\ldots,Z_d = z_d) = \Pr\{G^{-1}_{\theta\lambda_1}(U) = z_1,\ldots,G^{-1}_{\theta\lambda_d}(U) = z_d\}$$
$$= \Pr\{G_{\theta\lambda_1}(z_1 - 1) < U \leqslant G_{\theta\lambda_1}(z_1),\ldots,G_{\theta\lambda_d}(z_d - 1) < U \leqslant G_{\theta\lambda_d}(z_d)\}$$
$$= \Pr\left[\max\{G_{\theta\lambda_1}(z_1 - 1),\ldots,G_{\theta\lambda_d}(z_d - 1)\} < U \leqslant \min\{G_{\theta\lambda_1}(z_1),\ldots,G_{\theta\lambda_d}(z_d)\}\right].$$

Using $c_{\Lambda,\theta}$ to denote the probability mass function of the comonotonic random vector $(Z_1,\ldots,Z_d)$, we can write the above as

$$c_{\Lambda,\theta}(z_1,\ldots,z_d) = \left[\min_{j\in\{1,\ldots,d\}} G_{\theta\lambda_j}(z_j) - \max_{j\in\{1,\ldots,d\}} G_{\theta\lambda_j}(z_j - 1)\right]_+ \tag{6.3}$$

where, as introduced in Chapter 4, $[x]_+ = x\mathbf{1}(x > 0)$ and $\mathbf{1}$ is the indicator function. Then, putting everything together gives the following joint probability mass function for the $\mathcal{MP}_d(\Lambda, \theta)$ family

$$f_{\Lambda,\theta}(x_1, \ldots, x_d) = \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \left\{ \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) \right\} \times c_{\Lambda,\theta}(z_1, \ldots, z_d), \tag{6.4}$$

Again, an equivalent expression for the joint probability mass function can be obtained by conditioning on the underlying uniform random variable generating the comonotonic shock, viz.

$$\int_0^1 \prod_{j=1}^{d} g_{(1-\theta)\lambda_j} \left\{ x_j - G_{\theta\lambda_j}^{-1}(u) \right\} du.$$

6.2.1 PQD ordering

As discussed in Chapter 3, any d-variate distribution function F is bounded by the Fréchet–Hoeffding bounds. Specifically, for a joint distribution function F with fixed margins $F_1, \ldots, F_d$, it can be shown that

$$F_L(x_1, \ldots, x_d) \leqslant F(x_1, \ldots, x_d) \leqslant F_U(x_1, \ldots, x_d), \tag{6.5}$$

for all $(x_1, \ldots, x_d) \in \mathbb{R}^d$, where

$$F_U(x_1, \ldots, x_d) = \min\{F_1(x_1), \ldots, F_d(x_d)\}$$

and

$$F_L(x_1, \ldots, x_d) = \max\left\{0, \sum_{j=1}^{d} F_j(x_j) - (d-1)\right\}.$$

In the above, F_U and F_L respectively denote the upper and lower Fréchet–Hoeffding bounds. As noted in Chapter 3, the Fréchet–Hoeffding lower bound is a proper CDF when $d = 2$. In this case, the latter represents the distribution function of a counter-monotonic pair such that

$$(W_1, W_2) =_d (F_1^{-1}(U), F_2^{-1}(1-U))$$

for $U \sim \mathcal{U}(0, 1)$.

In general, for dimensions $d > 2$, the lower bound F_L is not a proper CDF, although Theorem 3.7 of Joe (1997) specifies criteria under which F_L is a proper CDF. On the other hand, the upper Fréchet–Hoeffding bound F_U is always a proper CDF and represents the distribution function of a comonotonic random vector

$$(W_1, \ldots, W_d) =_d \left(F_1^{-1}(U), \ldots, F_d^{-1}(U)\right),$$

where U is again a standard uniform random variable on the interval $(0, 1)$.

Analogously, Theorem 3.5 of Joe (1997) provides bounds for a multivariate survival function $\bar{F}(x_1, \ldots, x_d) = \Pr(X_1 > x_1, \ldots, X_d > x_d)$ with given marginal survival functions $\bar{F}_1, \ldots, \bar{F}_d$. Specifically, the theorem asserts that

$$\max\left\{0, \sum_{j=1}^{d} \bar{F}_j(x_j) - (d-1)\right\} \leqslant \bar{F}(x_1, \ldots, x_d) \leqslant \min_{j \in \{1,\ldots,d\}} \bar{F}_j(x_j). \tag{6.6}$$

The upper bound on the multivariate survival function is the corresponding survival function of F_U. In dimension 2, when F_L represents a proper CDF, the lower bound in (6.6) is the corresponding survival function of F_L.

In the multivariate case, the definition of positive quandrant dependence entails an ordering involving both the distribution function as well as the survival function. As outlined in Shaked and Shanthikumar (2007), the following definition extends the concept of PQD ordering to the multivariate setting.

Definition 6.1 *Suppose the random vector $\mathbf{X} = (X_1, \ldots, X_d)$ has distribution function F with corresponding survival function $\bar{F}$ and that $\mathbf{X}' = (X_1', \ldots, X_d')$ has distribution function F' with survival function $\bar{F}'$. Further suppose that both $\mathbf{X}$ and $\mathbf{X}'$ have the same marginal distributions. If, for all $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$,*

$$F(\mathbf{x}) \leqslant F'(\mathbf{x}) \quad \textit{and} \quad \bar{F}(\mathbf{x}) \leqslant \bar{F}'(\mathbf{x})$$

then the random vector $\mathbf{X}$ is smaller than $\mathbf{X}'$ in the PQD order. This will be denoted as $\mathbf{X} \prec_{PQD} \mathbf{X}'$.

Note that in the bivariate setting, the above definition can be slightly simplified as $F(x_1, x_2) \leqslant F'(x_1, x_2)$ is equivalent to $\bar{F}(x_1, x_2) \leqslant \bar{F}'(x_1, x_2)$ when both F and F' have the same margins. See Chapter 4 for details.

Remark 6.1 From (6.5) and (6.6), it follows that for any random vector $\mathbf{X} = (X_1, \ldots, X_d)$ with fixed marginal distributions $F_1, \ldots, F_d$, its comonotonic counterpart $\mathbf{Z} = (Z_1, \ldots, Z_d) =_d (F_1^{-1}(U), \ldots, F_d^{-1}(U))$, $U \sim \mathcal{U}(0, 1)$, is always greater in the sense of the PQD ordering, i.e., $\mathbf{X} \prec_{PQD} \mathbf{Z}$.

As was the case for the bivariate model, the proposed $\mathcal{MP}_d(\Lambda, \theta)$ family is ordered in the PQD ordering as regulated by the dependence parameter θ. For the proposed multivariate Poisson model, the PQD ordering in terms of θ then implies that for all $x_1, \ldots, x_d \in \mathbb{R}$, $\theta, \theta' \in [0, 1]$ and fixed $\Lambda \in (0, \infty)^d$, if $\theta < \theta'$ then $F_{\Lambda,\theta}(x_1, \ldots, x_d) \leqslant F_{\Lambda,\theta'}(x_1, \ldots, x_d)$ and $\bar{F}_{\Lambda,\theta}(x_1, \ldots, x_d) \leqslant \bar{F}_{\Lambda,\theta'}(x_1, \ldots, x_d)$. This will be formalized in Lemma 6.1. Establishing the PQD ordering in the

$\mathcal{MP}$ family relies on Theorem 9.A.4 given in Shaked and Shanthikumar (2007), which is detailed in the following theorem.

Theorem 6.1 *Suppose that the four random vectors* $\mathbf{X} = (X_1, \ldots, X_d)$, $\mathbf{Y} = (Y_1, \ldots, Y_d)$, $\mathbf{U} = (U_1, \ldots, U_d)$ *and* $\mathbf{V} = (V_1, \ldots, V_d)$ *satisfy*

$$\mathbf{X} \prec_{PQD} \mathbf{Y} \quad \text{and} \quad \mathbf{U} \prec_{PQD} \mathbf{V},$$

and suppose that $\mathbf{X}$ *and* $\mathbf{U}$ *are independent, and also that* $\mathbf{Y}$ *and* $\mathbf{V}$ *are independent. Then*

$$(\phi_1(X_1, U_1), \ldots, \phi_d(X_d, U_d)) \prec_{PQD} (\phi_1(Y_1, V_1), \ldots, \phi_d(Y_d, V_d)),$$

for all increasing functions $\phi_1, \ldots, \phi_d$.

In the multivariate setting, the notion of positive quadrant dependence also has the useful property of being closed under conjunctions and marginalization, as listed in parts (a) and (b) of Theorem 9.A.5 of Shaked and Shanthikumar (2007) and stated in the following theorem for completeness.

Theorem 6.2 *(a) Let* $\mathbf{X}_1, \ldots, \mathbf{X}_m$ *be a set of independent random vectors such that for each* $i \in \{1, \ldots, m\}$, *the random vector* $\mathbf{X}_i$ *has dimension* d_i. *Let* $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ *denote another set of independent random vectors where for each* $i \in \{1, \ldots, m\}$, *the random vector* $\mathbf{Y}_i$ *has dimension* d_i. *If* $\mathbf{X}_i \prec_{PQD} \mathbf{Y}_i$ *for all* $i \in \{1, \ldots, m\}$, *then*

$$(\mathbf{X}_1, \ldots, \mathbf{X}_m) \prec_{PQD} (\mathbf{Y}_1, \ldots, \mathbf{Y}_m).$$

(b) Let $\mathbf{X} = (X_1, \ldots, X_d)$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_d)$ *be random vectors of dimension* d. *If* $\mathbf{X} \prec_{PQD} \mathbf{Y}$, *then* $\mathbf{X}_I \prec_{PQD} \mathbf{Y}_I$ *for any subset* $I \subseteq \{1, \ldots, d\}$.

Establishing the PQD ordering for the proposed $\mathcal{MP}_d(\Lambda, \theta)$ family follows from Theorem 6.1. This is made formal in the following lemma.

Lemma 6.1 ***PQD ordering in the*** $\mathcal{MP}_d(\Lambda, \theta)$ ***family.***
Let $\mathbf{X} = (X_1, \ldots, X_d) \sim \mathcal{MP}_d(\Lambda, \theta)$ *and* $\mathbf{X}' = (X'_1, \ldots, X'_d) \sim \mathcal{MP}_d(\Lambda, \theta')$. *Then* $\theta < \theta' \Rightarrow \mathbf{X} \prec_{PQD} \mathbf{X}'$.

Proof. Fix the marginal parameters $\Lambda = (\lambda_1, \ldots, \lambda_d) \in (0, \infty)^d$ and fix $\theta < \theta'$, with both θ and θ' in the interval $[0, 1]$. By definition (6.1), for $\mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$ the components can be written as

$$\mathbf{X} = \mathbf{Y} + \mathbf{Z}$$

for random vectors $\mathbf{Y} = (Y_1, \ldots, Y_d)$ and $\mathbf{Z} = (Z_1, \ldots, Z_d)$, where $\mathbf{Y}$ has mutually independent components $Y_1 \sim \mathcal{P}\{(1-\theta)\lambda_1\}, \ldots, Y_d \sim \mathcal{P}\{(1-\theta)\lambda_d\}$ and is independent of the comonotonic shock $\mathbf{Z} \sim \mathcal{M}\{\mathcal{P}(\theta\lambda_1), \ldots, \mathcal{P}(\theta\lambda_d)\}$. Analogously, $\mathbf{X}' \sim \mathcal{MP}_d(\Lambda, \theta')$ can be expressed as

$$\mathbf{X}' = \mathbf{Y}' + \mathbf{Z}',$$

where again $\mathbf{Y}'$ has independent components $Y_1' \sim \mathcal{P}\{(1-\theta')\lambda_1\}, \ldots, Y_d' \sim \mathcal{P}\{(1-\theta')\lambda_d\}$ and is independent of the comonotonic shock $\mathbf{Z}' = (Z_1', \ldots, Z_d') \sim \mathcal{M}\{\mathcal{P}(\theta'\lambda_1), \ldots, \mathcal{P}(\theta'\lambda_d)\}$. Since the univariate Poisson distribution is infinitely divisible, the random vector $\mathbf{Y}$ can be rewritten as

$$\mathbf{Y} = \mathbf{T} + \mathbf{S},$$

where $\mathbf{T}$ is comprised of independent components $T_1 \sim \mathcal{P}\{(1-\theta')\lambda_1\}, \ldots, T_d \sim \mathcal{P}\{(1-\theta')\lambda_d\}$ and is independent of $\mathbf{S}$ with components $S_1 \sim \mathcal{P}\{(\theta'-\theta)\lambda_1\}, \ldots, S_d \sim \mathcal{P}\{(\theta'-\theta)\lambda_d\}$, where $S_1, \ldots, S_d$ are also mutually independent. Then, both $\mathbf{T}$ and $\mathbf{Y}'$ have the same marginal distributions and it follows trivially that $\mathbf{T} \prec_{PQD} \mathbf{Y}'$. Moreover, the random vectors $\mathbf{S} + \mathbf{Z}$ and $\mathbf{Z}'$ both have marginal distributions $\mathcal{P}(\theta'\lambda_1), \ldots, \mathcal{P}(\theta'\lambda_d)$. By definition, $\mathbf{Z}'$ is comonotonic and thus $\mathbf{S} + \mathbf{Z} \prec_{PQD} \mathbf{Z}'$ as noted in Remark 6.1. Then, by Theorem 6.1, it follows that

$$\mathbf{T} + \mathbf{S} + \mathbf{Z} \prec_{PQD} \mathbf{Y}' + \mathbf{Z}'.$$

In the above equation, $\mathbf{T} + \mathbf{S} + \mathbf{Z} = \mathbf{X}$ and the right-hand side is equivalent to $\mathbf{X}'$. Hence, the desired result has been established. □

Remark 6.2 Lemma 6.1 together with Theorem 6.2 imply that for any $\theta, \theta' \in [0, 1]$ such that $\theta < \theta'$, if $\mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$ and $\mathbf{X}' \sim \mathcal{MP}_d(\Lambda, \theta')$ then for any $j \neq k \in \{1, \ldots, d\}$, $(X_j, X_k) \prec_{PQD} (X_j', X_k')$ and consequently

$$\text{cov}(X_j, X_k) \leqslant \text{cov}(X_j', X_k'), \quad \text{corr}(X_j, X_k) \leqslant \text{corr}(X_j', X_k'),$$

where the latter inequality holds since the marginal parameters are fixed at $\Lambda = (\lambda_1, \ldots, \lambda_d)$. Accordingly, the pairwise covariance (and correlation) between any two components of $\mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$ are monotone increasing functions of the dependence parameter θ.

6.2.2 Covariance structure

The construction of the proposed multivariate Poisson family outlined in (6.1) is such that each component of the random vector $\mathbf{X} = (X_1, \ldots, X_d) \sim \mathcal{MP}_d(\Lambda, \theta)$ follows a Poisson distribution, with $X_j \sim \mathcal{P}(\lambda_j)$ for all $j \in \{1, \ldots, d\}$. As such, the vector of parameters $\Lambda = (\lambda_1, \ldots, \lambda_d)$ can be regarded as a mean vector representing the marginal Poisson rates, i.e.,

$$\mathrm{E}(\mathbf{X}^\top) = (\lambda_1, \ldots, \lambda_d)^\top. \tag{6.7}$$

In the proposed $\mathcal{MP}_d(\Lambda, \theta)$ formulation, there is a single dependence parameter θ that will regulate the strength of the dependence in the model. This implies a specific correlation structure. Indeed, for each pair (X_j, X_k), $j \neq k \in \{1, \ldots, d\}$, one has

$$\mathrm{cov}(X_j, X_k) = \mathrm{cov}(Z_j, Z_k) = m_{\lambda_j, \lambda_k}(\theta)$$

where, recall from Chapter 4,

$$m_{\lambda_j, \lambda_k}(\theta) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \min\{\bar{G}_{\theta\lambda_j}(m), \bar{G}_{\theta\lambda_k}(n)\} - \theta^2 \lambda_j \lambda_k.$$

As detailed in Chapter 4, $m_{\lambda_j, \lambda_k}(\theta)$ represents the largest possible covariance between two Poisson random variables Z_j, Z_k, with respective marginal rates $\theta\lambda_j$ and $\theta\lambda_k$. As noted in Remark 6.2, for fixed $\Lambda \in (0, \infty)^d$, the pairwise correlations given, for all $j \neq k \in \{1, \ldots, d\}$, by

$$\rho_\theta(\lambda_j, \lambda_k) = \frac{m_{\lambda_j, \lambda_k}(\theta)}{\sqrt{\lambda_j \lambda_k}}$$

are each an increasing function of θ. In particular, when the margins are independent, i.e., $\theta = 0$, the resulting correlation is 0 while the correlation attains the upper bound $\rho_{\max}$ when $\theta = 1$, i.e., the case of comonotonicity. Recall that the upper bound for the correlation is given by

$$\rho_{\max}(\lambda_j, \lambda_k) = \frac{1}{\sqrt{\lambda_j \lambda_k}} \left[-\lambda_j \lambda_k + \sum_{m \in \mathbb{N}} \sum_{n \in \mathbb{N}} \min\{\bar{G}_{\lambda_j}(m), \bar{G}_{\lambda_k}(n)\} \right].$$

Thus, in the proposed $\mathcal{MP}_d(\Lambda, \theta)$ model the pairwise correlations can attain all values within the interval $[0, \rho_{\max}]$, which represents the full range of possible correlation for an arbitrary pair of positively dependent Poisson random variables. Moreover, it is clear that any pair $(X_j, X_k) \subset (X_1, \ldots, X_d) \sim \mathcal{MP}_d(\Lambda, \theta)$ follows the proposed bivariate Poisson model introduced in Chapter 4, i.e., $(X_j, X_k) \sim \mathcal{BP}(\lambda_j, \lambda_k, \theta)$.

Marginally, each component of $\mathbf{X}$ is the sum of independent Poisson random variables. Specifically, for $j \in \{1, \ldots, d\}$, the jth component can be written as $X_j = Y_j + Z_j$ for $Y_j \sim \mathcal{P}\{(1-\theta)\lambda_j\}$ independent of $Z_j \sim \mathcal{P}(\theta\lambda_j)$. In this sense, the parameter θ can be regarded as a weighting factor that dictates what portion of the marginal mean is due to an independent component Y_j and the comonotonic shock Z_j.

The proposed $\mathcal{MP}(\Lambda, \theta)$ model, as given in 6.1, certainly provides a more flexible definition of a multivariate Poisson distribution than that obtained in the classical setting detailed in Chapter 2. The latter model is derived from a single common shock variable whereby each component is the convolution of an independent Poisson random variable with the same common shock. In dimension d, the classical common shock model has the form

$$X_1 = Y_1 + Z, \quad \ldots, \quad X_d = Y_d + Z, \tag{6.8}$$

for independent Poisson random variables $Y_1 \sim \mathcal{P}(\lambda_1 - \xi), \ldots, Y_d \sim \mathcal{P}(\lambda_d - \xi)$, $Z \sim \mathcal{P}(\xi)$, subject to the constraint $0 \leqslant \xi \leqslant \min(\lambda_1, \ldots, \lambda_d)$.

In the classical multivariate Poisson model, the common shock, Z, leads to $\text{cov}(X_j, X_k) = \xi$ for all $j \neq k \in \{1, \ldots, d\}$ so that the correlation between (X_j, X_k) is given by $\xi/\sqrt{\lambda_j \lambda_k}$. Accordingly, the classical multivariate model is characterized by a heterogeneous exchangeable correlation structure, with covariance matrix given by

$$\Sigma = \begin{bmatrix} \lambda_1 & \xi & \ldots & \xi \\ \xi & \lambda_2 & \ldots & \xi \\ \vdots & \vdots & \ddots & \vdots \\ \xi & \xi & \ldots & \lambda_d \end{bmatrix}.$$

Similarly to what was noted in the bivariate case, the pairwise correlations are restricted to fall in the interval $[0, \min(\lambda_j, \lambda_k)/\sqrt{\lambda_j \lambda_k}]$, which, unless $\lambda_j = \lambda_k = \lambda$, will fall short of the full range $[0, \rho_{\max}]$. Note that when the marginal parameters coincide, i.e., $\lambda_1 = \cdots = \lambda_d = \lambda$, both the classical common shock model and the proposed $\mathcal{MP}_d(\Lambda, \theta)$ model will coincide and the resulting correlation will fall in the interval $[0, 1]$.

Clearly, although the proposed $\mathcal{MP}_d(\Lambda, \theta)$ model is defined in terms of a single dependence parameter, this does not impose the constraint that all pairs exhibit a common covariance. In the proposed multivariate Poisson model, the covariance is a function of both the marginal parameters and the dependence parameter so that the pairwise covariances are not restricted to coincide. Thus, the $\mathcal{MP}_d(\Lambda, \theta)$ model can be regarded as having a "marginally-regulated" heterogeneous exchangeable correlation structure. In the proposed model, the covariance matrix is given by

$$\Sigma = \begin{bmatrix} \lambda_1 & \sigma_{12} & \ldots & \sigma_{1d} \\ \sigma_{21} & \lambda_2 & \ldots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \ldots & \lambda_d \end{bmatrix},$$

where $\sigma_{jk} = m_{\lambda_j,\lambda_k}(\theta) = m_{\lambda_k,\lambda_j}(\theta) = \sigma_{kj}$ for any $j \neq k \in \{1, \ldots, d\}$. Accordingly, the dependence parameter θ dictates the strength of the dependence, but the magnitude of this strength is regulated by the size of the marginal rates.

Nonetheless, the formulation of the proposed multivariate Poisson model in terms of a single dependence parameter does impose certain limitations in the covariance structure. This follows since θ simultaneously governs the dependence between all components. To see this, let $\mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$. For any $j \neq k \in \{1, \ldots, d\}$, if the pair (X_j, X_k) is comonotonic, i.e., $\rho_\theta = \rho_{\max}$, then $\theta = 1$ and $\mathbf{X}$ itself must also be comonotonic. On the other hand, if $\rho_\theta = 0$, i.e., the pair (X_j, X_k) consists of independent components, then $\theta = 0$ and all components of $\mathbf{X}$ are independent. This drawback of the proposed multivariate comonotonic shock model will be revisited in Chapter 7.

6.2.3 Recurrence relations

The recurrence relations in the proposed multivariate Poisson model are obtained as a straightforward extension of the approach used in the bivariate setting. Recall from Section 4.2.3, the recurrence results followed from the univariate Poisson relation, viz.

$$g_\lambda(x) = \frac{\lambda}{x} g_\lambda(x-1),$$

valid for all $x \in \{1, 2, \ldots\}$. Working with (6.4), it then follows that

$$\begin{aligned} &f_{\Lambda,\theta}(x_1 - 1, x_2, \ldots, x_d) = \\ &\sum_{z_1=0}^{x_1-1} \sum_{z_2=0}^{x_2} \cdots \sum_{z_d=0}^{x_d} g_{(1-\theta)\lambda_1}(x_1 - 1 - z_1) g_{(1-\theta)\lambda_2}(x_2 - z_2) \cdots g_{(1-\theta)\lambda_d}(x_d - z_d) c_{\Lambda,\theta}(z_1, \ldots, z_d). \end{aligned}$$

Through various algebraic manipulations, the right-hand side can be successively written as

$$\begin{aligned} &\{(1-\theta)\lambda_1\}^{-1} \sum_{z_1=0}^{x_1-1} \sum_{z_2=0}^{x_2} \cdots \sum_{z_d=0}^{x_d} (x_1 - z_1) g_{(1-\theta)\lambda_1}(x_1 - z_1) \\ &\qquad\qquad \times g_{(1-\theta)\lambda_2}(x_2 - z_2) \cdots g_{(1-\theta)\lambda_d}(x_d - z_d) c_{\Lambda,\theta}(z_1, \ldots, z_d), \end{aligned}$$

as

$$\begin{aligned} &\{(1-\theta)\lambda_1\}^{-1} f_{\Lambda,\theta}(x_1, \ldots, x_d) \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} (x_1 - z_1) \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) \\ &\qquad\qquad \times c_{\Lambda,\theta}(z_1, \ldots, z_d) / f_{\Lambda,\theta}(x_1, \ldots, x_d), \end{aligned}$$

and as

$$\{(1-\theta)\lambda_1\}^{-1} f_{\Lambda,\theta}(x_1, \ldots, x_d) \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} (x_1 - z_1) p_{\Lambda,\theta}(z_1, \ldots, z_d | x_1, \ldots, x_d).$$

Therefore,

$$f_{\Lambda,\theta}(x_1-1,x_2,\ldots,x_d) = \{(1-\theta)\lambda_1\}^{-1} f_{\Lambda,\theta}(x_1,\ldots,x_d)\,\{x_1 - \mathrm{E}(Z_1|x_1,\ldots,x_d)\}\,.$$

In the above, $p_{\Lambda,\theta}(z_1,\ldots,z_d|x_1,\ldots,x_d)$ is used to denote the conditional probability mass function of $\mathbf{z} = (z_1,\ldots,z_d)$ given $\mathbf{x} = (x_1,\ldots,x_d)$, specifically given by

$$p_{\Lambda,\theta}(\mathbf{z}|\mathbf{x}) = \frac{p_{\Lambda,\theta}(\mathbf{z},\mathbf{x})}{f_{\Lambda,\theta}(\mathbf{x})} = \frac{\prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) c_{\Lambda,\theta}(z_1,\ldots,z_d)}{f_{\Lambda,\theta}(x_1,\ldots,x_d)}.$$

Analogously, for any $j \in \{1,\ldots,d\}$, the following recurrence relation holds:

$$f_{\Lambda,\theta}(x_1,\ldots,x_{j-1},x_j-1,x_{j+1},\ldots,x_d) = \{(1-\theta)\lambda_j\}^{-1}\{x_j - \mathrm{E}(Z_j|x_1,\ldots,x_d)\} \\ \times f_{\Lambda,\theta}(x_1,\ldots,x_{j-1},x_j,x_{j+1},\ldots,x_d). \quad (6.9)$$

6.2.4 Convolutions in the $\mathcal{MP}_d$ family

The proposed bivariate Poisson family discussed in Chapter 4 is a special case of the $\mathcal{MP}$ model where $d = 2$. In Section 4.2.4, it was shown that the $\mathcal{BP}(\Lambda,\theta)$ distribution is not closed under convolutions. This then translates immediately to the multivariate model, so that the $\mathcal{MP}_d$ family is also not closed under convolutions.

6.3 Estimation

Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ denote a random sample of size n from the proposed $\mathcal{MP}_d(\Lambda,\theta)$ model and for each $i \in \{1,\ldots,n\}$, write $\mathbf{X}_i = \{(X_{i1},\ldots,X_{id})\}$. It is then of interest to estimate the marginal parameters $\Lambda \in (0,\infty)^d$ as well as the dependence parameter $\theta \in [0,1]$. To this end, there are several approaches that will yield consistent estimates for $\Psi = (\Lambda,\theta)$. Similarly to what was explored in Chapter 4, three methods will be detailed in this section, specifically, moment-based estimation, maximum likelihood estimation and the two-stage inference function for the margins approach.

6.3.1 Method of moments

Similarly to what was done in the case where $d = 2$, the method of moments approach can be adapted to estimate the parameters $\Psi = (\Lambda,\theta)$ in the $\mathcal{MP}_d(\Lambda,\theta)$ model. Since the parameter vector $\Lambda = (\lambda_1,\ldots,\lambda_d)$ is comprised of the marginal Poisson rates, consistent estimates are obtained by setting, for each $j \in \{1,\ldots,d\}$,

$$\tilde{\lambda}_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij} = \bar{X}_j. \quad (6.10)$$

As noted in Chapter 4, for univariate observations arising from a Poisson distribution, the method of moments estimator coincides with the marginal maximum likelihood estimators. Standard maximum likelihood theory then implies that, for each $j \in \{1, \ldots, d\}$,

$$\sqrt{n}\,(\tilde{\lambda}_j - \lambda_j) \rightsquigarrow \mathcal{N}(0, I_j^{-1}),$$

where I_j represents the Fisher Information associated with the jth margin. Recall that in the case of univariate $\mathcal{P}(\lambda_j)$ observations, I_j is given by λ_j^{-1} so that the asymptotic variance of $\tilde{\lambda}_j$ can be consistently estimated by $\bar{X}_j$ for each $j \in \{1, \ldots, d\}$.

Estimation of the dependence parameter requires mixed moments. As previously mentioned, for any bivariate subset of $\mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$, say (X_j, X_k), $j \neq k \in \{1, \ldots, d\}$, the covariance takes the form

$$m_{\lambda_j,\lambda_k}(\theta) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \min\{\bar{G}_{\theta\lambda_j}(m), \bar{G}_{\theta\lambda_k}(n)\} - \theta^2 \lambda_j \lambda_k.$$

Let the sample covariance between the components X_j and X_k be denoted by S_{jk}, where

$$S_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k).$$

Then, holding the marginal parameters λ_j and λ_k fixed at their respective MM estimates, an estimate for θ is obtained by solving

$$m_{\tilde{\lambda}_j,\tilde{\lambda}_k}(\theta) = S_{jk}.$$

As previously noted, for fixed (λ_j, λ_k), the covariance $m_{\lambda_j,\lambda_k}(\theta)$ is an increasing function of θ. Thus, the above equation will have a unique solution, say $\tilde{\theta}_{jk}$, provided that S_{jk} falls in the interval of permissible covariance values given by

$$[0, \rho_{\max}(\bar{X}_j, \bar{X}_k)\sqrt{\bar{X}_j, \bar{X}_k}],$$

where $\rho_{\max}(\bar{X}_j, \bar{X}_k)\sqrt{\bar{X}_j, \bar{X}_k} = m_{\bar{X}_j,\bar{X}_k}(1)$. Similarly to what was discussed in the bivariate case in Chapter 4, in small samples it could happen that S_{jk} falls outside the range of possible covariance values implied by the $\mathcal{MP}_d(\Lambda, \theta)$ family. In the case that S_{jk} is less than 0, a model accommodating negative dependence should be used in lieu of the $\mathcal{MP}_d(\Lambda, \theta)$ model. When S_{jk} falls above the upper bound, $\rho_{\max}(\bar{X}_j, \bar{X}_k)\sqrt{\bar{X}_j, \bar{X}_k}$, the convention will be to set $\tilde{\theta}_{jk} = 1$.

It can be shown, as in, e.g., Theorem 8 of Ferguson (1996), that as $n \to \infty$,

$$\sqrt{n}\,\{S_{jk} - m_{\lambda_j,\lambda_k}(\theta)\} \rightsquigarrow \mathcal{N}[0, \sigma^2(\theta, \lambda_j, \lambda_k)],$$

with $\sigma^2(\theta, \lambda_j, \lambda_k) = \text{var}\{(X_j - \lambda_j)(X_k - \lambda_k)\}$. Using the notation introduced in Chapter 4, define the function γ_{jk} to be the inverse of covariance function (for fixed Λ), i.e., the mapping $\gamma_{jk} : x \mapsto$

$m^{-1}_{\lambda_j,\lambda_k}(x)$, so that $\tilde{\theta}_{jk} = \gamma_{jk}(S_{jk})$. The Delta Method then implies that, as $n \to \infty$,

$$\sqrt{n}\,(\tilde{\theta}_{jk} - \theta) \rightsquigarrow \mathcal{N}[0, \{\gamma'_{jk}(\sigma_{jk})\}^2 \sigma^2(\theta, \lambda_j, \lambda_k)],$$

where γ'_{jk} denotes the derivative of γ_{jk}.

It then follows that in the multivariate setting, one can define several consistent and asymptotically Gaussian estimators for the dependence parameter θ using the aforementioned procedure. Specifically, there are $\binom{d}{2}$ distinct estimators of the form $\tilde{\theta}_{jk}$ derived from the pairwise covariance obtained for all possible pairs $(X_j, X_k) \subseteq (X_1, \ldots, X_d)$, $j < k \in \{1, \ldots, d\}$. Slutsky's theorem then implies that $\tilde{\theta}$ defined by

$$\tilde{\theta} = \binom{d}{2}^{-1} \sum_{j<k \in \{1,\ldots,d\}} \tilde{\theta}_{jk}$$

is a consistent estimator for the dependence parameter θ. This moment-based estimator is also asymptotically Normal, as will be established in Theorem 6.6. The proof of this theorem relies on several standard statistical results, which will be stated here for completeness.

Remark 6.3 Note that we could consider all $d(d-1)$ possible pairs (X_j, X_k) with $j \neq k \in \{1, \ldots, d\}$ to obtain a set of estimators $\tilde{\theta}_{jk}$, but this is redundant since $S_{jk} = S_{kj}$ and also $m_{\lambda_j,\lambda_k}(\theta) = m_{\lambda_k,\lambda_j}(\theta)$ so that $\tilde{\theta}_{jk} = \tilde{\theta}_{kj}$. Furthermore, if one considers the pair (X_j, X_j), $j \in \{1, 2, \ldots, d\}$, $m_{\lambda_j,\lambda_j}(\theta)$ reduces to var$(X_j) = \lambda_j$ so that any $\theta \in [0, 1]$ will yield $m_{\lambda_j,\lambda_j} = S_j^2$, where

$$S_{jj} = S_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2.$$

Thus, it is also unnecessary to consider all d^2 pairs (X_j, X_k), $j, k \in \{1, \ldots, d\}$.

The following theorem, which is Corollary 1.2.18 of Muirhead (1982), establishes the asymptotic distribution of the sample covariance matrix S, where the results are stated in terms of a vectorization of S. Following the notation of Muirhead (1982), the vectorization of a $p \times q$ matrix T, denoted vec(T), is the $pq \times 1$ vector formed by stacking the q columns of the matrix T. That is, writing T in terms of its columns such that $T = [\mathbf{t}_1 \cdots \mathbf{t}_q]$, vec$(T)$ is the vector given by

$$\text{vec}(T) = \begin{bmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_q \end{bmatrix}.$$

Theorem 6.3 *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be a sequence of iid m-variate random vectors with finite fourth moments such that* $\mathrm{E}(\mathbf{X}_i) = \boldsymbol{\mu}$ *and cov*$(\mathbf{X}_i) = \Sigma$ *for each $i \in \{1, \ldots, n\}$. Define the sample mean* *as* $\bar{\mathbf{X}} = (\mathbf{X}_1 + \cdots + \mathbf{X}_n)/n$ *and let* $\mathbf{S}$ *denote the sample covariance matrix given by*

$$\frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^\top.$$

Then the vectorized sample covariance matrix is asymptotically Gaussian. Specifically, as $n \to \infty$,

$$\sqrt{n}\,\{vec(\mathbf{S} - \Sigma)\} \rightsquigarrow \mathcal{N}_{m^2}(0, V),$$

where the asymptotic variance V *is given by*

$$V = cov\left[vec\{(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^\top\}\right].$$

Note that, as mentioned in Muirhead (1982), the asymptotic m^2-variate Gaussian distribution of the vectorized sample covariance matrix is singular. This follows since the sample covariance S is symmetric so that its vectorized counterpart has repeat entries as $S_{jk} = S_{kj}$ for each $j, k \in \{1, \ldots, m\}$.

The Cramér–Wold Theorem, stated next, is a necessary tool for establishing the asymptotic normality of the moment-based estimator $\tilde{\theta}$. For completeness, this is detailed in Theorem 6.4, as taken from Billingsley (1995).

Theorem 6.4 *For random vectors* $\mathbf{X}_n = (X_{n1}, \ldots, X_{nk})$ *and* $\mathbf{Y} = (Y_1, \ldots, Y_k)$, *a necessary and sufficient condition for* $\mathbf{X}_n \rightsquigarrow \mathbf{Y}$ *is that, for each* $(t_1, \ldots, t_k) \in \mathbb{R}^k$, *one has, as* $n \to \infty$,

$$\sum_{u=1}^{k} t_u X_{nu} \rightsquigarrow \sum_{u=1}^{k} t_u Y_u.$$

Recall that in the bivariate Poisson model (i.e., when $d = 2$), proving the asymptotic normality of the MM estimator relies on the Delta Method. Accordingly, in higher dimensions the multivariate analogue of the Delta Method is necessary. The latter is formally stated in the following Theorem, as given on p. 674 of Wakefield (2013) .

Theorem 6.5 *Suppose* $\sqrt{n}\,(\mathbf{Y}_n - \boldsymbol{\mu}) \rightsquigarrow \mathbf{Z}$ *and suppose that* $g : \mathbb{R}^p \to \mathbb{R}^k$ *has a derivative* g' *at* $\boldsymbol{\mu}$, *where* g' *is a* $k \times p$ *matrix of derivatives. Then, as* $n \to \infty$,

$$\sqrt{n}\,\{g(\mathbf{Y}) - g(\boldsymbol{\mu})\} \rightsquigarrow g'(\boldsymbol{\mu})\mathbf{Z}.$$

If $\mathbf{Z} \sim \mathcal{N}_p(0, \Sigma)$, *then, as* $n \to \infty$,

$$\sqrt{n}\,\{g(\mathbf{Y}) - g(\boldsymbol{\mu})\} \rightsquigarrow \mathcal{N}_k(0, \{g'(\boldsymbol{\mu})\}\Sigma\{g'(\boldsymbol{\mu})\}^\top).$$

The asymptotic normality of the proposed method of moments estimator $\tilde{\theta}$ can now be established. This is formally stated in the following theorem.

Theorem 6.6 ***Asymptotic normality of the method of moments estimator***
Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent random vectors, with $\mathbf{X}_i \sim \mathcal{MP}_d(\Lambda, \theta)$ for all $i \in \{1, \ldots, n\}$. For each $j < k \in \{1, \ldots, d\}$, let $\tilde{\theta}_{jk}$ denote the (unique) solution to the equation

$$m_{\bar{X}_j, \bar{X}_k}(\theta) = S_{jk},$$

where the function $m_{\lambda_j, \lambda_k}(\theta)$ denotes the theoretical covariance for the pair $(X_j, X_k) \subseteq \mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$ and S_{jk} denotes the sample covariance given by

$$S_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k).$$

Define $\tilde{\theta}$ to be the average of the set of estimators $\tilde{\theta}_{jk}$, $j < k \in \{1, \ldots, d\}$, i.e.,

$$\tilde{\theta} = \left\{ \binom{d}{2} \right\}^{-1} \sum_{j<k \in \{1,\ldots,d\}} \tilde{\theta}_{jk}.$$

Then $\tilde{\theta}$ is a consistent estimator of θ and, as $n \to \infty$,

$$\sqrt{n}\,(\tilde{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, \tilde{V}),$$

where the asymptotic variance is given by

$$\begin{aligned} \tilde{V} &= \frac{1}{(d^*)^2} \{\Gamma'(\Sigma^*)\} V^* \{\Gamma'(\Sigma^*)\}^\top \\ &= \frac{1}{(d^*)^2} \sum_{j<k \in \{1,\ldots,d\}} \sum_{r<s \in \{1,\ldots,d\}} \{\gamma'_{jk}(\sigma_{jk})\}\{\gamma'_{rs}(\sigma_{rs})\} \\ &\qquad \times \mathit{cov}\,\{(X_{ij} - \lambda_j)(X_{ik} - \lambda_k), (X_{ir} - \lambda_r)(X_{is} - \lambda_s)\} \end{aligned}$$

Proof. According to Theorem 6.3, the vectorized sample covariance matrix,

$$vec(\mathbf{S}) = (S_1^2, S_{12}, \ldots, S_{1d}, S_{12}, S_2^2, \ldots, S_{2d}, \ldots, S_{12}, S_{2d}, \ldots, S_d^2)^\top,$$

is asymptotically normally distributed with asymptotic variance given by

$$V = \text{cov}\left[\text{vec}\{(\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^\top\}\right].$$

By the Cramér–Wold theorem, the dimensionality of $\text{vec}(\mathbf{S})$ can be reduced from d^2 to $d^* = \binom{d}{2}$, thereby removing duplicate terms as well as the sample variance terms. That is, the Cramér–Wold theorem ensures that the vector vec($\mathbf{S}^*$) with components S_{jk}, $j < k \in \{1, \ldots, d\}$, i.e.

$$\text{vec}(\mathbf{S}^*) = (S_{12}, S_{13}, \ldots, S_{1d}, S_{23}, \ldots, S_{2d}, \ldots, S_{d-1d}),$$

is also asymptotically Gaussian with mean given by

$$\text{vec}(\Sigma^*) = (\sigma_{12}, \sigma_{13}, \ldots, \sigma_{1d}, \sigma_{23}, \ldots, \sigma_{d-1d}),$$

where $\sigma_{jk} = \text{cov}(X_j, X_k) = m_{\lambda_j,\lambda_k}(\theta)$. Properties of the Gaussian distribution imply that the asymptotic variance of vec($\mathbf{S}^*$) is

$$V^* = \text{cov}\left[\text{vec}\{(\mathbf{X}_i^* - \boldsymbol{\mu}^*)(\mathbf{X}_i^* - \boldsymbol{\mu}^*)^\top\}\right],$$

where vec$\{(\mathbf{X}_i^* - \boldsymbol{\mu}^*)(\mathbf{X}_i^* - \boldsymbol{\mu}^*)^\top\}$ is the vector given by

$$\begin{bmatrix} (X_{i1} - \lambda_1)(X_{i2} - \lambda_2) \\ (X_{i1} - \lambda_1)(X_{i3} - \lambda_3) \\ \vdots \\ (X_{i1} - \lambda_1)(X_{id} - \lambda_d) \\ (X_{i2} - \lambda_2)(X_{i3} - \lambda_3) \\ \vdots \\ (X_{i2} - \lambda_2)(X_{id} - \lambda_d) \\ \vdots \\ (X_{id-1} - \lambda_{d-1})(X_{id} - \lambda_d) \end{bmatrix}.$$

As previously introduced, let the function γ_{jk} be the inverse of the covariance function such that $\gamma_{jk}(S_{jk}) = m^{-1}_{\lambda_j,\lambda_k}(S_{jk}) = \tilde{\theta}_{jk}$ for $j < k \in \{1, \ldots, d\}$. Now define the function $\Gamma : \mathbb{R}^{d^*} \mapsto [0,1]^{d^*}$ such that

$$\begin{aligned} \Gamma(\mathbf{S}^*) &= ((\gamma_{12}(S_{12}), \gamma_{13}(S_{13}), \ldots, \gamma_{1d}(S_{1d}), \gamma_{23}(S_{23}), \ldots, \gamma_{d-1d}(S_{d-1d})) \\ &= (\tilde{\theta}_{12}, \tilde{\theta}_{13}, \ldots, \tilde{\theta}_{1d}, \tilde{\theta}_{23}, \ldots, \tilde{\theta}_{d-1d}). \end{aligned}$$

Applying the multivariate Delta Method, one has that

$$\sqrt{n}\,\text{vec}\{\Gamma(S^*) - \Gamma(\Sigma^*)\} \rightsquigarrow \mathcal{N}_{d^*}(0, \{\Gamma'(\Sigma^*)\}V^*\{\Gamma'(\Sigma^*)\}^\top).$$

Note that in the above, $\Gamma(\Sigma^*)$ yields a vector of length d^* wherein each component is identically θ. By the Cramér–Wold theorem, it follows that any linear combination of the components $\Gamma(S^*)$

is also asymptotically normally distributed. Thus, the linear combination $\tilde{\theta}$ given by

$$\tilde{\theta} = \left\{ \binom{d}{2} \right\}^{-1} \sum_{j<k \in \{1,\ldots,d\}} \tilde{\theta}_{jk}$$

is asymptotically Gaussian. Properties of the Gaussian distribution then yield the desired result, i.e.,

$$\sqrt{n}\,(\tilde{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, \tilde{V}),$$

with

$$\tilde{V} = \frac{1}{(d^*)^2} \sum_{j<k \in \{1,\ldots,d\}} \sum_{r<s \in \{1,\ldots,d\}} \{\gamma'_{jk}(\sigma_{jk})\}\{\gamma'_{rs}(\sigma_{rs})\} \times \text{cov}\left\{(X_{ij} - \lambda_j)(X_{ik} - \lambda_k), (X_{ir} - \lambda_r)(X_{is} - \lambda_s)\right\}.$$

Thus we can conclude. □

As was the case in the bivariate setting, some simplifications ensue when the marginal parameters coincide, i.e., when $\lambda_1 = \cdots = \lambda_d = \lambda$. In this case, the comonotonic shock reduces to a common shock, i.e., $(Z_1, \ldots, Z_d) =_d (Z, \ldots, Z)$ where $Z \sim \mathcal{P}(\theta\lambda)$.

Similarly to the notation introduced in Chapter 4, for $\ell \in \{1, \ldots, d\}$, write

$$X_\ell - \lambda_\ell = \mathring{Y}_\ell + \mathring{Z}_\ell,$$

where $\mathring{Y}_\ell$ and $\mathring{Z}_\ell$ represent the centred random variables given by

$$\mathring{Y}_\ell = Y_\ell - (1-\theta)\lambda_\ell, \quad \mathring{Z}_\ell = Z_\ell - \theta\lambda_\ell.$$

It then follows that

$$\begin{aligned}
\text{cov}\{(X_j - \lambda_j)(X_k - \lambda_k),&(X_r - \lambda_r)(X_s - \lambda_s)\} \\
&= \text{cov}\{(\mathring{Y}_j + \mathring{Z}_j)(\mathring{Y}_k + \mathring{Z}_k), (\mathring{Y}_r + \mathring{Z}_r)(\mathring{Y}_s + \mathring{Z}_s)\} \\
&= \text{cov}\{\mathring{Y}_j\mathring{Y}_k + \mathring{Y}_j\mathring{Z}_k + \mathring{Y}_k\mathring{Z}_j + \mathring{Z}_j\mathring{Z}_k,\ \mathring{Y}_r\mathring{Y}_s + \mathring{Y}_r\mathring{Z}_s + \mathring{Y}_s\mathring{Z}_r + \mathring{Z}_r\mathring{Z}_s\}.
\end{aligned}$$

Simplifications of the above covariance expression depend on the overlap of the indices j, k, r, s. For example, for $j < k$ and $r < s$ with $j \neq k \neq r \neq s \in \{1, \ldots, d\}$, the random variables $\mathring{Y}_j$, $\mathring{Y}_k$, $\mathring{Y}_r$ and $\mathring{Y}_s$ are mutually independent with mean zero. Moreover, $\mathring{Y}_\ell$ is independent of $\mathring{Z}_t$ for each $\ell, t = j, k, r, s$. Thus, in this particular setting the covariance simplifies to cov$(\mathring{Z}_j\mathring{Z}_k, \mathring{Z}_r\mathring{Z}_s)$. In the special case where the marginal parameters coincide, i.e., $\lambda_1 = \cdots = \lambda_d = \lambda$, this further simplifies:

$$\text{cov}(\mathring{Z}_j\mathring{Z}_k, \mathring{Z}_r\mathring{Z}_s) = \text{var}(\mathring{Z}^2) = \theta\lambda(1+3\theta\lambda) - \theta^2\lambda^2 = \theta\lambda(1+2\theta\lambda).$$

As another example, when $j < k = r < s$, the covariance expression reduces to

$$\text{cov}(\mathring{Z}_j\mathring{Y}_k, \mathring{Y}_k\mathring{Z}_s) + \text{cov}(\mathring{Z}_j\mathring{Z}_k, \mathring{Z}_k\mathring{Z}_s) \;=\; \text{E}(\mathring{Y}_k^2)\text{E}(\mathring{Z}_j\mathring{Z}_s) + \text{E}(\mathring{Z}_j\mathring{Z}_k^2\mathring{Z}_s) - \text{E}(\mathring{Z}_j\mathring{Z}_k)\text{E}(\mathring{Z}_k\mathring{Z}_s).$$

Then, when $\lambda_1 = \cdots = \lambda_d = \lambda$, the above further simplifies to

$$\begin{aligned} \text{var}(Y_k)\text{var}(Z) + \text{E}\{(Z-\theta\lambda)^4\} - \text{var}(Z)^2 &= (1-\theta)\theta\lambda^2 + \theta\lambda(1+3\theta\lambda) - \theta^2\lambda^2 \\ &= \theta\lambda(1+\lambda+\theta\lambda). \end{aligned}$$

In general, simplifying the asymptotic variance V is tedious even when $\lambda_1 = \cdots = \lambda_d$ since, as already mentioned, simplifications of the covariance term

$$\text{cov}\left\{(X_{ij}-\lambda_j)(X_{ik}-\lambda_k), (X_{ir}-\lambda_r)(X_{is}-\lambda_s)\right\}$$

depends on the overlap of the indices j, k, r, s. When $d = 2$, it is clear that the asymptotic results coincide with those detailed for the $\mathcal{BP}$ model in Chapter 4.

6.3.2 Maximum likelihood estimation

Likelihood-based estimation is also feasible in the proposed $\mathcal{MP}_d(\Lambda, \theta)$ model. Under the usual regularity conditions, which are here satisfied, maximum likelihood estimation yields consistent and efficient estimates, say $(\hat{\lambda}_1, \ldots, \hat{\lambda}_d, \hat{\theta})$. Working with the formulation of the probability mass function given in (6.4), the likelihood function is given by

$$L(\Lambda, \theta) = \prod_{i=1}^{n} \left[\sum_{z_1=0}^{x_{i1}} \cdots \sum_{z_d=0}^{x_{id}} \left\{ \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_{ij}-z_j) \right\} c_{\Lambda,\theta}(z_1, \ldots, z_d) \right],$$

with corresponding log-likelihood

$$\ell(\Lambda, \theta) = \sum_{i=1}^{n} \ln \left\{ \sum_{z_1=0}^{x_{i1}} \cdots \sum_{z_d=0}^{x_{id}} \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_{ij}-z_j) c_{\Lambda,\theta}(z_1, \ldots, z_d) \right\}. \tag{6.11}$$

The maximum likelihood estimates, $\hat{\Psi} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_d, \hat{\theta})$, are then established as

$$\hat{\Psi} = \underset{\lambda_1,\ldots,\lambda_d,\theta}{\text{argmax}}\; \ell(\Lambda, \theta),$$

subject to the constraints that $\lambda_1, \ldots, \lambda_d \in (0, \infty)$ and $\theta \in [0, 1]$. This is equivalent to solving

$$\left(\frac{\partial}{\partial \lambda_1} \, \ell(\Lambda, \theta), \ldots, \frac{\partial}{\partial \lambda_d} \, \ell(\Lambda, \theta), \frac{\partial}{\partial \theta} \, \ell(\Lambda, \theta) \right) = 0^\top.$$

Similarly to what was shown in Section 4.3.2, it is straightforward to show that for any $k \in \{1, \ldots, d\}$,

$$\frac{\partial}{\partial \lambda_k} \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j)$$
$$= -(1-\theta) \left\{ g_{(1-\theta)\lambda_k}(x_k - z_k) - g_{(1-\theta)\lambda_k}(x_k - 1 - z_k) \right\} \prod_{j=1, j \neq k}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j).$$

Additionally, it can be shown that

$$\frac{\partial}{\partial \theta} \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) = (\lambda_1 + \cdots + \lambda_d) \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j)$$
$$- \lambda_1 g_{(1-\theta)\lambda_1}(x_1 - 1 - z_1) \prod_{j=2}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) - \cdots$$
$$- \lambda_k g_{(1-\theta)\lambda_k}(x_k - 1 - z_k) \prod_{j=1, j \neq k}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) - \cdots$$
$$- \lambda_d g_{(1-\theta)\lambda_d}(x_d - 1 - z_d) \prod_{j=1}^{d-1} g_{(1-\theta)\lambda_j}(x_j - z_j).$$

It then follows that, for any $k \in \{1, \ldots, d\}$,

$$\frac{\partial}{\partial \lambda_k} \, f_{\Lambda,\theta}(x_1, \ldots, x_d) = \frac{\partial}{\partial \lambda_k} \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) c_{\Lambda,\theta}(z_1, \ldots, z_d)$$

with the right-hand side reducing to

$$\sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \Big[-(1-\theta) \prod_{j=1, j\neq k}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) \\ \times \left\{ g_{(1-\theta)\lambda_k}(x_k - z_k) - g_{(1-\theta)\lambda_k}(x_k - 1 - z_k) \right\} c_{\Lambda,\theta}(z_1,\ldots,z_d) \\ + \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) \left\{ \frac{\partial}{\partial \lambda_k} c_{\Lambda,\theta}(z_1,\ldots,z_d) \right\} \Big] \\ = -(1-\theta) \left\{ f_{\Lambda,\theta}(x_1,\ldots,x_d) - f_{\Lambda,\theta}(x_1,\ldots,x_{k-1},x_k - 1, x_{k+1},\ldots,x_d) \right\} \\ + f_{\Lambda,\theta}(x_1,\ldots,x_d) \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \Big[\left\{ \frac{\partial}{\partial \lambda_k} \ln c_{\Lambda,\theta}(z_1,\ldots,z_d) \right\} \\ \times \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) c_{\Lambda,\theta}(z_1,\ldots,z_d) / f_{\Lambda,\theta}(x_1,\ldots,x_d) \Big].$$

Using the recurrence relation (6.9), this further simplifies as

$$\frac{\partial}{\partial \lambda_k} f_{\Lambda,\theta}(x_1,\ldots,x_d) = f_{\Lambda,\theta}(x_1,\ldots,x_d) \Big[-(1-\theta) + \frac{1}{\lambda_k} \left\{ x_k - \mathrm{E}(Z_k | x_1,\ldots,x_d) \right\} \\ + \mathrm{E} \left\{ \frac{\partial}{\partial \lambda_k} \ln c_{\Lambda,\theta}(Z_1,\ldots,Z_d) | x_1,\ldots,x_d \right\} \Big].$$

In a similar manner, it can be shown that

$$\frac{\partial}{\partial \theta} f_{\Lambda,\theta}(x_1,\ldots,x_d) = f_{\Lambda,\theta}(x_1,\ldots,x_d) \left[\sum_{j=1}^{d} \lambda_j - \left(\frac{1}{1-\theta} \right) \sum_{j=1}^{d} \left\{ x_j - \mathrm{E}(Z_j | x_1,\ldots,x_d) \right\} \right. \\ \left. + \mathrm{E} \left\{ \frac{\partial}{\partial \theta} \ln c_{\Lambda,\theta}(Z_1,\ldots,Z_d) | x_1,\ldots,x_d \right\} \right].$$

It is then straightforward to derive the score equations in the $\mathcal{MP}_d(\Lambda,\theta)$ model. In particular,

$$\frac{\partial}{\partial \lambda_k} \ell(\Lambda,\theta) = \frac{\partial}{\partial \lambda_k} \sum_{i=1}^{n} \ln f_{\Lambda,\theta}(x_{i1},\ldots,x_{id}) \\ = -n(1-\theta) + \frac{n}{\lambda_k} \left\{ \bar{x}_k - \bar{q}_k(\Lambda,\theta) \right\} + \sum_{i=1}^{n} \mathrm{E} \left\{ \frac{\partial}{\partial \lambda_k} \ln c_{\Lambda,\theta}(Z_1,\ldots,Z_d) | x_{i1},\ldots,x_{id} \right\},$$

where, following the same notation introduced in Chapter 4,

$$\bar{q}_k(\Lambda,\theta) = \frac{1}{n} \sum_{i=1}^{n} q_{ik}(\Lambda,\theta) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}(Z_k | x_{i1},\ldots,x_{id}).$$

Note that the score equation for λ_k can also be rewritten as

$$\frac{\partial}{\partial \lambda_k}\,\ell(\Lambda,\theta) = \frac{n}{\lambda_k}(\bar{x}_k - \lambda_k) - \frac{n}{\lambda_k}\,\{\bar{q}_k(\Lambda,\theta) - \theta\lambda_k\} + \sum_{i=1}^{n} \mathrm{E}\left\{\frac{\partial}{\partial \lambda_k}\,\ln c_{\Lambda,\theta}(Z_1,\ldots,Z_d)|x_{i1},\ldots,x_{id}\right\}. \quad (6.12)$$

Notice the similarity between (6.12) and (4.15). In addition, it can be shown that

$$\frac{\partial}{\partial \theta}\,\ell(\Lambda,\theta) = n\sum_{j=1}^{d}\lambda_j - \frac{n}{(1-\theta)}\sum_{j=1}^{d}\{\bar{x}_j - \bar{q}_j(\Lambda,\theta)\} + \sum_{i=1}^{n} \mathrm{E}\left\{\frac{\partial}{\partial \theta}\,\ln c_{\Lambda,\theta}(Z_1,\ldots,Z_d)|x_{i1},\ldots,x_{id}\right\},$$

which can be equivalently expressed as

$$\frac{\partial}{\partial \theta}\,\ell(\Lambda,\theta) = -\frac{n}{(1-\theta)}\sum_{j=1}^{d}(\bar{x}_j - \lambda_j) + \frac{n}{(1-\theta)}\sum_{j=1}^{d}\{\bar{q}_j(\Lambda,\theta) - \theta\lambda_j\} + \sum_{i=1}^{n} \mathrm{E}\left\{\frac{\partial}{\partial \theta}\,\ln c_{\Lambda,\theta}(Z_1,\ldots,Z_d)|x_{i1},\ldots,x_{id}\right\}. \quad (6.13)$$

Again, note the similarity between (6.13) and (4.18).

As was the case in the bivariate model, the form of the log-likelihood and corresponding score equations does not allow for an analytical maximization of the log-likelihood. Rather, numerical optimization procedures are necessary to find the maximum likelihood estimates $\hat{\Lambda}, \hat{\theta}$. In order to simplify the optimization, a reparametrization of $\ell(\Lambda,\theta)$ should be considered so as to remove the parameter constraints. In the case of the $\mathcal{MP}_d(\Lambda,\theta)$ model, this is accomplished by setting, for all $j \in \{1,\ldots,d\}$,

$$\zeta_j = \ln(\lambda_j) \Leftrightarrow \lambda_j = \exp(\zeta_j), \quad \eta = \ln\left(\frac{\theta}{1-\theta}\right) \Leftrightarrow \theta = \exp(\eta)/\{1+\exp(\eta)\}.$$

The optimization can then be carried out on $\ell(\zeta_1,\ldots,\zeta_d,\eta)$ to yield maximum likelihood estimates $\hat{\zeta}_1,\ldots,\hat{\zeta}_d,\hat{\eta}$. By the invariance properties of MLEs, one then has $\hat{\lambda}_j = \exp(\hat{\zeta}_j)$ for all $j \in \{1,\ldots,d\}$ and $\hat{\theta} = \exp(\hat{\eta})/\{1+\exp(\hat{\eta})\}$.

Maximum likelihood theory ensures that, under certain regularity conditions, the MLEs $\hat{\Psi} = (\hat{\lambda}_1,\ldots,\hat{\lambda}_d,\hat{\theta})$ are asymptotically Gaussian, with mean zero and asymptotic variance given by the inverse of the Fisher information matrix, i.e., as $n \to \infty$,

$$\sqrt{n}\,(\hat{\Psi} - \Psi) \rightsquigarrow \mathcal{N}(0,\mathcal{I}^{-1})$$

In the practical implementation of ML estimation, an estimate of the asymptotic variance can be calculated as the sample variance of a set of ML estimates resulting from a large number of bootstrap replications.

Remark 6.4 Note that in the case where all marginal parameters are equal, the components of the random vector $(X_1, \ldots, X_d)$ are in fact identically distributed. Setting $\lambda_1 = \cdots = \lambda_d = \lambda$, one has that $X_j = Y_j + Z$ for all $j \in \{1, \ldots, d\}$, where Z is a common shock variable. Specifically, $Z \sim \mathcal{P}(\theta\lambda)$ and the components $(Y_1, \ldots, Y_d)$ are independent and identically distributed $\mathcal{P}\{(1-\theta)\lambda\}$ random variables. In this setting, the $\mathcal{MP}_d(\Lambda, \theta)$ model corresponds to the classical common shock multivariate Poisson model with $\lambda_1 = \cdots = \lambda_d$. The latter was discussed in detail in Chapter 2.

The EM algorithm

Analogously to the bivariate model, the construction of the proposed $\mathcal{MP}_d(\Lambda, \theta)$ model, as in (6.1), is based on the convolution of unobserved random vectors $\mathbf{Y}$ and $\mathbf{Z}$. As was noted in Chapter 4, this formulation lends itself naturally to the use of the Expectation-Maximization (EM) algorithm for finding the maximum likelihood estimates $\hat{\Psi} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_d, \hat{\theta})$.

In the classical multivariate Poisson model, the construction is based on a single common shock variable. In this setting, the application of the EM algorithm is straightforward and leads to an easily-implemented algorithm, as shown in Chapter 2. For the proposed $\mathcal{MP}_d(\Lambda, \theta)$ family, however, this is not the case, similarly to what was seen in the bivariate model in Chapter 4. Naturally, the implementation of the EM algorithm in the $\mathcal{MP}$ family leads to an analogous form for the E-step, $Q(\Psi|\Psi^{(k)})$ as was obtained for the bivariate model. Nonetheless, for completeness, the steps are outlined in what follows.

Consider a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathcal{MP}_d(\Lambda, \theta)$. Let the missing or latent variables consist of the underlying uniform random variables $U_1, \ldots, U_n$ generating the comonotonic shock vectors, viz.

$$(Z_{i1}, \ldots, Z_{id}) = \left\{G^{-1}_{\theta\lambda_1}(U_i), \ldots, G^{-1}_{\theta\lambda_d}(U_i)\right\}$$

for all $i \in \{1, \ldots, n\}$. The joint distribution of the observed data $\mathbf{x}$ and the missing data u is given by

$$p_{\Lambda,\theta}(x_1, \ldots, x_d, u) = p_{\Lambda,\theta}(x_1, \ldots, x_d|u) f_U(u) = \prod_{j=1}^{d} g_{(1-\theta)\lambda_j} \left\{x_j - G^{-1}_{\theta\lambda_j}(u)\right\}.$$

Analogously to what was shown in Chapter 4, the complete data log-likelihood is then

$$\ell_C(\Lambda, \theta) = \sum_{i=1}^{n}\sum_{j=1}^{d} \left[-(1-\theta)\lambda_j + \{x_{ij} - G^{-1}_{\theta\lambda_j}(u_i)\} \ln\{(1-\theta)\lambda_j\} - \ln\left[\{x_{ij} - G^{-1}_{\theta\lambda_j}(u_i)\}!\right]\right].$$

With some manipulations, the above can be rewritten as

$$\ell_C(\Lambda,\theta) = -n(1-\theta)\sum_{j=1}^{d}\lambda_j + \ln\left\{(1-\theta)\sum_{j=1}^{d}\lambda_j\right\}\sum_{i=1}^{n}\sum_{j=1}^{d}\{x_{ij} - G^{-1}_{\theta\lambda_j}(u_i)\}$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{d}\ln\left(\frac{\lambda_j}{\sum_{k=1}^{d}\lambda_k}\right)\{x_{ij} - G^{-1}_{\theta\lambda_j}(u_i)\} - \sum_{i=1}^{n}\sum_{j=1}^{d}\ln\left[\{x_{ij} - G^{-1}_{\theta\lambda_j}(u_i)\}!\right].$$

Taking the conditional expectation of $\ell_C(\Lambda,\theta)$, given the observed data $\mathbf{x}$ and current parameter estimates $\Psi^{(k)} = (\Lambda^{(k)}, \theta^{(k)})$, one gets

$$Q(\Psi|\Psi^{(k)}) = -n(1-\theta)\sum_{j=1}^{d}\lambda_j + n\ln\left\{(1-\theta)\sum_{j=1}^{d}\lambda_j\right\}\sum_{j=1}^{d}\left\{\bar{x}_j - \bar{q}_j(\Lambda^{(k)},\theta^{(k)})\right\}$$
$$+ n\sum_{j=1}^{d}\ln\left(\frac{\lambda_j}{\sum_{k=1}^{d}\lambda_k}\right)\left\{\bar{x}_j - \bar{q}_j(\Lambda^{(k)},\theta^{(k)})\right\} - R(\mathbf{X},\Psi^{(k)}),$$

where $R(\mathbf{X},\Psi^{(k)})$ is the remainder term, which depends only on the observed data and current parameter estimates, is equal to

$$\sum_{i=1}^{n}\sum_{j=1}^{d}\mathrm{E}\left[\ln[\{x_{ij} - G^{-1}_{\theta\lambda_j}(u_i)\}!]|\mathbf{x},\Psi^{(k)}\right].$$

As was done in Chapter 4, consider a reparametrization of the above where $\alpha_j = \lambda_j/(\lambda_1+\cdots+\lambda_d)$ for all $j \in \{1,\ldots,d\}$ and $\beta = (1-\theta)(\lambda_1+\cdots+\lambda_d)$. Note that $\alpha_1+\cdots+\alpha_d = 1$. In terms of $\alpha_1,\ldots,\alpha_d,\beta$, the E-step can be rewritten as

$$Q(\Psi|\Psi^{(k)}) = -n\beta + n\ln(\beta)\sum_{j=1}^{d}\left\{\bar{x}_j - \bar{q}_j(\Lambda^{(k)},\theta^{(k)})\right\}$$
$$+ n\sum_{j=1}^{d}\ln(\alpha_j)\sum_{j=1}^{d}\left\{\bar{x}_j - \bar{q}_j(\Lambda^{(k)},\theta^{(k)})\right\} - R(\mathbf{x},\Psi^{(k)}).$$

As was shown in the $\mathcal{BP}$ model, this leads to identifiability issues as only $(\alpha_1,\ldots,\alpha_{d-1},\beta)$ can be identified from the above. Thus, the E-step does not allow to identify the parameter updates $(\Lambda^{(k+1)},\theta^{(k+1)})$.

6.3.3 Inference for margins

Let $p_{\Lambda,\theta}(\mathbf{z},\mathbf{x})$ denote the joint probability mass function of the comonotonic shock $\mathbf{Z} = (Z_1,\ldots,Z_d)$ and the observed data $\mathbf{X} = (X_1,\ldots,X_d)$, where

$$p_{\lambda,\theta}(\mathbf{z},\mathbf{x}) = \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j)c_{\Lambda,\theta}(z_1,\ldots,z_d).$$

The first part of the above equations consists of the marginal contributions due to each independent univariate Poisson component $Y_1,\ldots,Y_d$. Suppressing the subscript j, it is straightforward to show that

$$g_{(1-\theta)\lambda}(x-z) = g_\lambda(x)\{g_{\theta\lambda}(z)\}^{-1}b_{x,\theta}(z)$$

where, as introduced in Chapter 4, $b_{x,\theta}(z)$ represents a Binomial probability mass function with size x and probability θ, evaluated at z. Using this formulation, the joint probability mass function in (6.4) can be rewritten as

$$\begin{aligned} f_{\Lambda,\theta}(\mathbf{x}) &= \sum_{z_1=0}^{x_1}\cdots\sum_{z_d=0}^{x_d}\left\{\prod_{j=1}^{d} g_{\lambda_j}(x_j)b_{x_j,\theta}(z_j)/g_{\theta\lambda_j}(z_j)\right\}c_{\Lambda,\theta}(z_1,\ldots,z_d) \\ &= \left\{\prod_{j=1}^{d} g_{\lambda_j}(x_j)\right\}\sum_{z_1=0}^{x_1}\cdots\sum_{z_d=0}^{x_d}\prod_{j=1}^{d} b_{x_j,\theta}(z_j)c_{\Lambda,\theta}(z_1,\ldots,z_d)/g_{\theta\lambda_j}(z_j). \end{aligned}$$

Following the same notation as in Chapter 4, let $\omega_{\Lambda,\theta}$ denote weights given by

$$\omega_{\Lambda,\theta}(\mathbf{z};\mathbf{x}) = \prod_{j=1}^{d}\left\{b_{x_j,\theta}(z_j)/g_{\theta\lambda_j}(z_j)\right\}.$$

Then, an alternative expression for the joint probability mass function in the $\mathcal{MP}_d(\Lambda,\theta)$ family is

$$f_{\Lambda,\theta}(x_1,\ldots,x_d) = \left\{\prod_{j=1}^{d} g_{\lambda_j}(x_j)\right\}\sum_{z_1=0}^{x_1}\cdots\sum_{z_d=0}^{x_d}\omega_{\Lambda,\theta}(\mathbf{z};\mathbf{x})c_{\Lambda,\theta}(z_1,\ldots,z_d). \tag{6.14}$$

The above is the multivariate extension of (4.22).

Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ denote a random sample from the proposed $\mathcal{MP}_d(\Lambda,\theta)$ distribution. Using the expression for the probability mass function given in (6.14), the resulting log-likelihood has the form

$$\ell(\Lambda,\theta) = \sum_{i=1}^{n}\left[\sum_{j=1}^{d}\ln\{g_{\lambda_j}(x_{ij})\} + \ln\left\{\sum_{z_1=0}^{x_{i1}}\cdots\sum_{z_d=0}^{x_{id}}\omega_{\Lambda,\theta}(\mathbf{z};\mathbf{x}_i)c_{\Lambda,\theta}(z_1,\ldots,z_d)\right\}\right]$$

For each $j \in \{1, \ldots, d\}$, let

$$\ell_j(\lambda_j) = \sum_{i=1}^{n} \ln\{g_{\lambda_j}(x_{ij})\}$$

denote the marginal log-likelihoods and define $\ell_D(\Lambda, \theta)$ as

$$\ell_D(\Lambda, \theta) = \ln\left\{\sum_{z_1=0}^{x_{i1}} \cdots \sum_{z_d=0}^{x_{id}} \omega_{\Lambda,\theta}(\mathbf{z}; \mathbf{x}_i) c_{\Lambda,\theta}(z_1, \ldots, z_d)\right\}. \quad (6.15)$$

Then the log-likelihood can be expressed as

$$\ell(\Lambda, \theta) = \sum_{j=1}^{d} \ell_j(\lambda_j) + \ell_D(\Lambda, \theta), \quad (6.16)$$

yielding an analogous multivariate version of the log-likelihood obtained in the $\mathcal{BP}$ model; see (4.23). The first part of (6.16) consists of the marginal contributions to the log-likelihood of each component $X_1, \ldots, X_d$ and is equivalent to the full log-likelihood in the case of independence, i.e., when $\theta = 0$. The remaining term, $\ell_D(\Lambda, \theta)$, thus encompasses the dependence inherent in the multivariate model.

Based on the form of the log-likelihood given in (6.16), the maximum likelihood estimators $\hat{\Psi} = (\hat{\Lambda}, \hat{\theta})$ are the solutions to the set of score equations

$$\begin{aligned}
\frac{\partial}{\partial \lambda_1} \ell(\Lambda, \theta) &= \frac{\partial}{\partial \lambda_1} \ell_1(\lambda_1) + \frac{\partial}{\partial \lambda_1} \ell_D(\Lambda, \theta) = 0, \\
&\vdots \\
\frac{\partial}{\partial \lambda_d} \ell(\Lambda, \theta) &= \frac{\partial}{\partial \lambda_d} \ell_d(\lambda_d) + \frac{\partial}{\partial \lambda_d} \ell_D(\Lambda, \theta) = 0, \\
\frac{\partial}{\partial \theta} \ell(\Lambda, \theta) &= \frac{\partial}{\partial \theta} \ell_D(\Lambda, \theta) = 0.
\end{aligned}$$

As previously discussed, the MLEs do not have a closed form expression, except in trivial cases. Moreover, as pointed out in Joe (2005), maximum likelihood estimation can be computationally difficult, or even infeasible, as d increases. As discussed in Chapter 4, the inference function for margins (IFM) method provides an alternative approach which reduces the computationally complexity by using a two-stage estimation method. Recall that in the IFM approach, the marginal parameters are first estimated via their respective univariate log-likelihoods and then an estimate for the dependence parameter is computed using the full log-likelihood, holding the marginal parameters fixed at their respective marginal MLEs.

The application of the IFM approach in the proposed multivariate Poisson model is analogous to what was shown in the $\mathcal{BP}$ model in Section 4.3.3. Specifically, the IFM estimates $(\check{\lambda}_1, \ldots, \check{\lambda}_d, \check{\theta})$

are obtained as the solution to

$$\left(\frac{\partial}{\partial \lambda_1}\,\ell_1(\lambda_1),\ldots,\frac{\partial}{\partial \lambda_d}\,\ell_d(\lambda_d),\frac{\partial}{\partial \theta}\,\ell(\check{\Lambda},\theta)\right)=\mathbf{0}^\top.$$

The implementation of this approach results in a two-stage estimation procedure. At the first step, the marginal likelihoods of each component is maximized separately to obtain $\check{\lambda}_1,\ldots,\check{\lambda}_d$. The second stage entails maximizing the full likelihood with respect to θ, holding the marginal parameters fixed at $\check{\Lambda}=(\check{\lambda}_1,\ldots,\check{\lambda}_d)$.

In the $\mathcal{MP}(\Lambda,\theta)$ model, $X_j\sim\mathcal{P}(\lambda_j)$ for all $j\in\{1,\ldots,d\}$ so that the marginal log-likelihoods have the form

$$\ell_j(\lambda_j)=-n\lambda_j+n\bar{x}_j\ln(\lambda_j)-\sum_{i=1}^{n}\ln(x_{ij}!),$$

with corresponding score equation given by

$$\frac{\partial}{\partial \lambda_j}\,\ell_j(\lambda_j)=-n\left(1-\frac{\bar{x}_j}{\lambda_j}\right)=0$$

thus yielding $\check{\lambda}_j=\bar{X}_j$. The IFM estimate of the dependence parameter, $\check{\theta}$, is then determined as

$$\check{\theta}=\arg\max_{\theta}\ell(\check{\Lambda},\theta)=\arg\max_{\theta}\ell_D(\check{\Lambda},\theta).$$

The above amounts to solving $\partial\ell(\check{\Lambda},\theta)/\partial\theta=\partial\ell_D(\check{\Lambda},\theta)/\partial\theta=0$. Using the score equation derived in (6.13), the IFM estimating equation for the dependence parameter then simplifies to

$$\begin{aligned}\frac{\partial}{\partial \theta}\,\ell(\check{\Lambda},\theta)=\frac{\partial}{\partial \theta}\,\ell_D(\check{\Lambda},\theta)=\frac{n}{(1-\theta)}\sum_{j=1}^{d}\left\{\bar{q}_j(\check{\Lambda},\theta)-\theta\check{\lambda}_j\right\}&\\+\sum_{i=1}^{n}\mathrm{E}\left\{\frac{\partial}{\partial \theta}\,\ln c_{\check{\Lambda},\theta}(Z_1,\ldots,Z_d)|x_{i1},\ldots,x_{id}\right\}&.\end{aligned}$$

To reiterate, this is analogous to what was shown for the $\mathcal{BP}$ model in Chapter 4.

EM algorithm

Similarly to the two-dimensional setting, in the proposed multivariate Poisson model the EM algorithm can be applied within the IFM framework to determine the estimate for the dependence parameter $\check{\theta}$. In particular, this is accomplished by applying the EM algorithm to the pseudo likelihood $L(\check{\Lambda},\theta)$. In a similar manner to the bivariate case, the latent variables will be taken as the underlying $\mathcal{U}(0,1)$ random variables generating the comotononic shock viz.

$$(Z_1,\ldots,Z_d)=(G^{-1}_{\theta\lambda_1}(U),\ldots,G^{-1}_{\theta\lambda_d}(U)).$$

Analogously to (4.25), the complete data pseudo log-likelihood within the IFM framework is given by

$$\ell_C(\check{\Lambda}, \theta; \mathbf{x}, \mathbf{u}) = -n(1-\theta)\sum_{j=1}^{d}\check{\lambda}_j + \sum_{i=1}^{n}\sum_{j=1}^{d}\ln\left\{(1-\theta)\check{\lambda}_j\right\}\left\{x_{ij} - G^{-1}_{\theta\check{\lambda}_j}(u_i)\right\} - \sum_{i=1}^{n}\sum_{j=1}^{d}\ln\left[\{x_{ij} - G^{-1}_{\theta\check{\lambda}_j}(u_i)\}!\right].$$

This leads to the E-step

$$Q(\theta|\theta^{(k)}) = -n(1-\theta)\sum_{j=1}^{d}\check{\lambda}_j + n\ln(1-\theta)\sum_{j=1}^{d}\left\{x_{ij} - \bar{q}_j(\check{\Lambda}, \theta^{(k)})\right\} + n\sum_{j=1}^{d}\ln(\check{\lambda}_j)\left\{x_{ij} - \bar{q}_j(\check{\Lambda}, \theta^{(k)})\right\} - R(\mathbf{x}, \theta^{(k)}),$$

where $R(\mathbf{x}, \theta^{(k)})$ consists of a remainder term which does not depend of the unknown parameter θ, specifically

$$R(\mathbf{x}, \theta^{(k)}) = \sum_{i=1}^{n}\sum_{j=1}^{d}\mathrm{E}\left[\ln[\{x_{ij} - G^{-1}_{\theta\check{\lambda}_j}(u_i)\}!]|\mathbf{x}, \theta^{(k)}\right].$$

Analogously to what was found in the bivariate setting, the parameter update is given by

$$\theta^{(k+1)} = \arg\max_{\theta} Q(\theta|\theta^{(k)}) = \frac{\sum_{j=1}^{d}\bar{q}_j(\check{\Lambda}, \theta)}{\sum_{j=1}^{d}\check{\lambda}_j} = \frac{\sum_{j=1}^{d}\bar{q}_j(\check{\Lambda}, \theta)}{\sum_{j=1}^{d}\bar{x}_j}.$$

The form of the parameter update $\theta^{(k+1)}$ is intuitive as it represents the proportion of the overall average, $(\bar{x}_1 + \cdots + \bar{x}_d)/d$, that is due to the comonotonic shock. In addition, each iterate $\{\theta^{(k)}, k = 1, 2, \ldots\}$ will always fall within the interval $[0, 1]$ since by definition $0 \leqslant \bar{q}_j(\check{\Lambda}, \theta) \leqslant \bar{x}_j$. Under certain conditions (see Chapter 4), the sequence of EM updates $\{\theta^{(k)}, k = 1, 2, \ldots\}$ will converge to the IFM estimate $\check{\theta}$.

IFM theory, as outlined in Chapter 3, ensures that, as $n \to \infty$,

$$\sqrt{n}\,(\check{\Psi} - \Psi) \rightsquigarrow \mathcal{N}(0, V).$$

The elements of the asymptotic variance V can be derived in a similar manner to what was done for the $\mathcal{BP}$ model in Section 4.3.3. In particular, let p denote the dimension of the parameter vector $\Psi = (\lambda_1, \ldots, \lambda_d, \theta)$, so that $p = d + 1$ and let V_{jk} denote the (j, k)th element of V. Then, for each

$j, k \in \{1, \ldots, d\}$, it can be shown that

$$V_{jj} = \lambda_j, \quad V_{jk} = m_{\lambda_j,\lambda_k}(\theta), \quad V_{jp} = -\mathcal{I}_{pp}^{-1} \sum_{k=1}^{d} m_{\lambda_j,\lambda_k}(\theta)\mathcal{I}_{pk}$$

$$V_{pp} = \mathcal{I}_{pp}^{-1} + \left(\mathcal{I}_{pp}^{-1}\right)^2 \sum_{j=1}^{d}\sum_{k=1}^{d} \mathcal{I}_{pj}\mathcal{I}_{pk} m_{\lambda_j,\lambda_k}(\theta),$$

where, recall from Section 4.3.3,

$$\mathcal{I} = -\mathrm{E}\left\{ \frac{\partial^2}{\partial\Psi\partial\Psi^\top} \ln f_{\Lambda,\theta}(X_1, \ldots, X_d) \right\},$$

and $\mathcal{I}_{jk}$ denotes the (j, k)th element of the Fisher Information matrix $\mathcal{I}$.

As was noted in Chapter 4, there is a loss in efficiency when the IFM approach is taken rather than basing the estimation on the full likelihood. This will be further investigated in the various simulations summarized in the following section.

6.4 Simulations

A set of simulations were carried out to assess the performance of the three estimation techniques outlined in Section 6.3. In total, 30 scenarios were tested, each of which resulted from the unique combination of values for the dimension $d \in \{2, 3, 4\}$, the dependence parameter $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and sample size $n \in \{50, 500\}$. The marginal parameters were fixed throughout, specifically with $\Lambda = (1, \ldots, d)$, and the focus was placed on the estimation of the dependence parameter.

For every scenario resulting from specific values of (d, θ, n), a random sample of size n was generated by setting, for each $i \in \{1, \ldots, n\}$,

$$X_{i1} = G^{-1}_{(1-\theta)\lambda_1}(V_{i1}) + G^{-1}_{\theta\lambda_1}(U_i), \ldots, X_{id} = G^{-1}_{(1-\theta)\lambda_d}(V_{id}) + G^{-1}_{\theta\lambda_d}(U_i),$$

where $V_{i1}, \ldots, V_{id}, U_i$ denote independent $\mathcal{U}(0, 1)$ random variables. The method of moments, inference functions for the margins and maximum likelihood estimation approaches were then each applied to the specific sample to obtain a set of three estimates respectively denoted as $\tilde{\theta}$, $\check{\theta}$ and $\hat{\theta}$. Each of the 30 unique scenarios was replicated 500 times, yielding a total of 15,000 iterations.

At each iteration, estimation was first performed using the method of moments. As was previously detailed, the latter is obtained by averaging the unique estimates $\tilde{\theta}_{jk}$, $j < k \in \{1, \ldots, d\}$, that result from solving the equation that matches the sample covariance to the theoretical covariance, viz.

$$S_{jk} = m_{\bar{X}_j,\bar{X}_k}(\theta).$$

The solution to the above was obtained using the `uniroot` function in R. Similarly to what was noted in the bivariate setting in Chapter 4, it sometimes occurred that the pairwise sample covariances were found to be negative. Any instances where $S_{jk} < 0$, thus resulting in $\tilde{\theta}_{jk} = NA$, were excluded from the average $\tilde{\theta}$. In the majority (13,621) of the $15{,}000$ iterations, there were no occurrences where $\tilde{\theta}_{jk} = NA$. Moreover, overall there were only 228 instances where each pairwise sample covariance was negative leading to the MM estimate $\tilde{\theta}$ being set to NA. Not surprisingly, this tended to occur more frequently for the smaller sample size, lower levels of dependence and lower dimension d.

For both likelihood-based estimation methods, the optimization procedures were initialized using the MM estimates $(\tilde{\Lambda}, \tilde{\theta})$. As was done for the estimation in the $\mathcal{BP}$ model, if the MM estimate was found to either be NA or 0, a starting value of $\theta^{(0)} = 0.01$ was used instead while when $\tilde{\theta} = 1$ the starting value was taken to be 0.99. This was done so as to avoid issues at the boundary of the parameter space for θ. In the simulations, there were no errors with the implementation of both the IFM and MLE methods. There were however 36 iterations where the IFM optimization algorithm did not converge within the maximum number of iterations, which are set to 100 by default in R. This error only occurred in small sample sizes.

As was alluded to in Chapter 4, the run times of the three estimation methods becomes increasingly disparate as the dimension increased. Figure 6.1 graphs the running times for the three estimation approaches in dimensions 2, 3 and 4. It is clear from the plots that for all dimensions, the MM method tends to run the fastest, followed by the IFM approach and then the MLE method. More striking, however, is the drastic increase in the running time for the full maximum likelihood approach as the dimension increases in comparison to both the MM and IFM approaches. Notably, in dimension 4, the average run time of the MLE method is more than 6 times longer than the average IFM run time and almost 17 times longer than the average run time for the method of moments. Conceivably, a full maximum likelihood approach would eventually become infeasible for higher dimensions. Clearly, the IFM method is an appealing alternative for likelihood-based estimation in high dimensional problems.

The estimation results for the dependence parameter are summarized in Figures 6.2 through 6.6. The patterns are similar to what was observed in the $\mathcal{BP}$ model in Chapter 4. The estimation results improve as the sample size increases, as one would expect. The MM performs comparatively better for weaker levels of dependence for all dimensions and sample sizes. Both the IFM and MLE produce similar results. Moreover, the variability in the two likelihood-based methods decreases as θ increases. This pattern was seen in all dimensions and sample sizes. As explained in Chapter 4, this is intuitive since when θ increases, the random vector $(X_1, \ldots, X_d)$ is predominately generated by the comonotonic shock $(Z_1, \ldots, Z_d)$ which induces less uncertainty as the components of the comonotonic shock vector are generated from a common underlying $\mathcal{U}(0, 1)$ random variable.

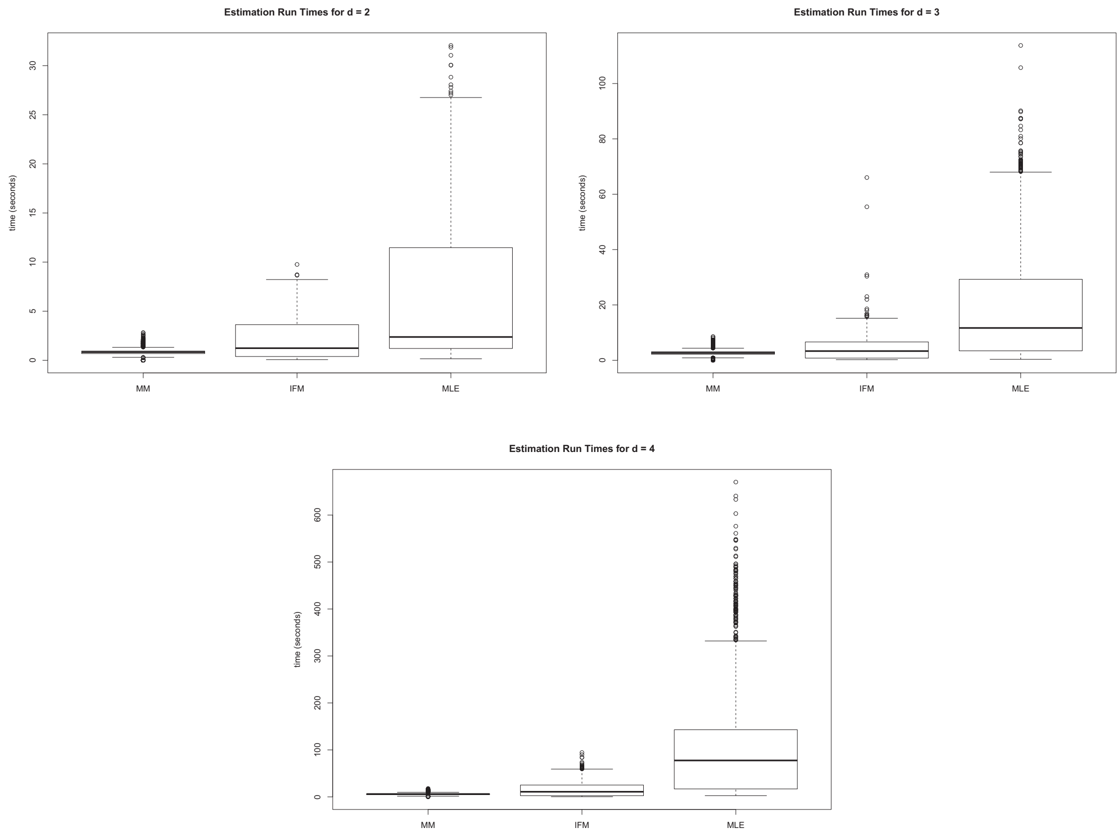

Figure 6.1: Run times for MM (left), IFM (middle) and MLE (right) estimation methods in dimensions 2 (top left plot), 3 (top right plot) and 4 (bottom plot).

As was observed in the bivariate setting in Chapter 4, there were some instances where seemingly outlier estimates were obtained for the likelihood-based methods. For example, this is seen in the graphs for the scenario where $(\theta, d, n) = (0.75, 2, 50)$. As was investigated in the $\mathcal{BP}$ model, this tended to happen when the starting value $\tilde{\theta}$ was much larger than the true parameter value. In the particular scenario where $(\theta, d, n) = (0.75, 2, 50)$, the outlier estimates occurred when $\tilde{\theta} = 1$ so that the optimization procedures for the IFM and MLE approaches were initialized at $\theta^{(0)} = 0.99$. As commented in Chapter 4, it seems that starting the algorithm close to the boundary can sometimes lead to poor estimates in the likelihood-based methods.

Figures 6.7 through 6.11 show the estimation results for Λ in the case of the trivariate model. Similarly to what was observed in the bivariate setting, the MM/IFM estimates for the marginal parameters are very similar to the full maximum likelihood estimates. This held true for all levels of dependence θ. Again, this suggests that there is very little information concerning the marginal parameters Λ in the dependence portion of the log-likelihood $\ell_D(\Lambda, \theta)$.

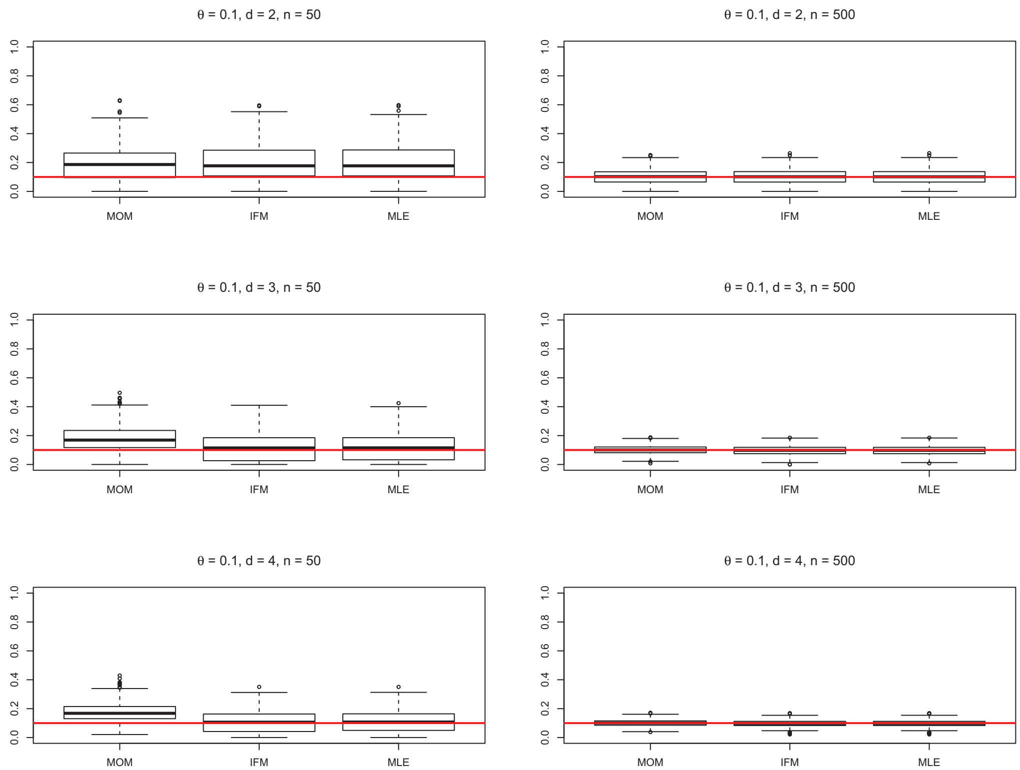

Figure 6.2: Estimation results for θ from the method of moments (left), inference function for margins (middle) and maximum likelihood estimation (right) in indicated scenario.

6.5 Data illustration

As an application of the proposed $\mathcal{MP}$ class, the trivariate model was applied to data from BIXI, Montréal's bike sharing system. BIXI provides open data pertaining to bike usage on their website (https://montreal.bixi.com/en/open-data). The data used in this illustration consist of bike usage information from the 2017 season, spanning from April 15, 2017 to November 15, 2017, inclusively. The raw data details each bike rental occurrence, specifying the exact date and time of departure, the departure station, the arrival station, the arrival date and time, the total time the bike was in use and a binary variable indicating whether the passenger was a BIXI member. For this particular application, only weekday rentals were considered for BIXI members at three specific departure stations located on Montréal's south shore town of Longueuil. The stations are located at the intersection of St-Charles / Charlotte street (Station 1), the intersection of St-Charles / St-Sylvestre street (Station 2) and at Collège Édouard-Montpetit (Station 3). The random vector (X_1, X_2, X_3) then consists of the total number of daily bike rentals taken from Stations 1, 2 and 3, respectively. This resulted in a total of 153 observations.

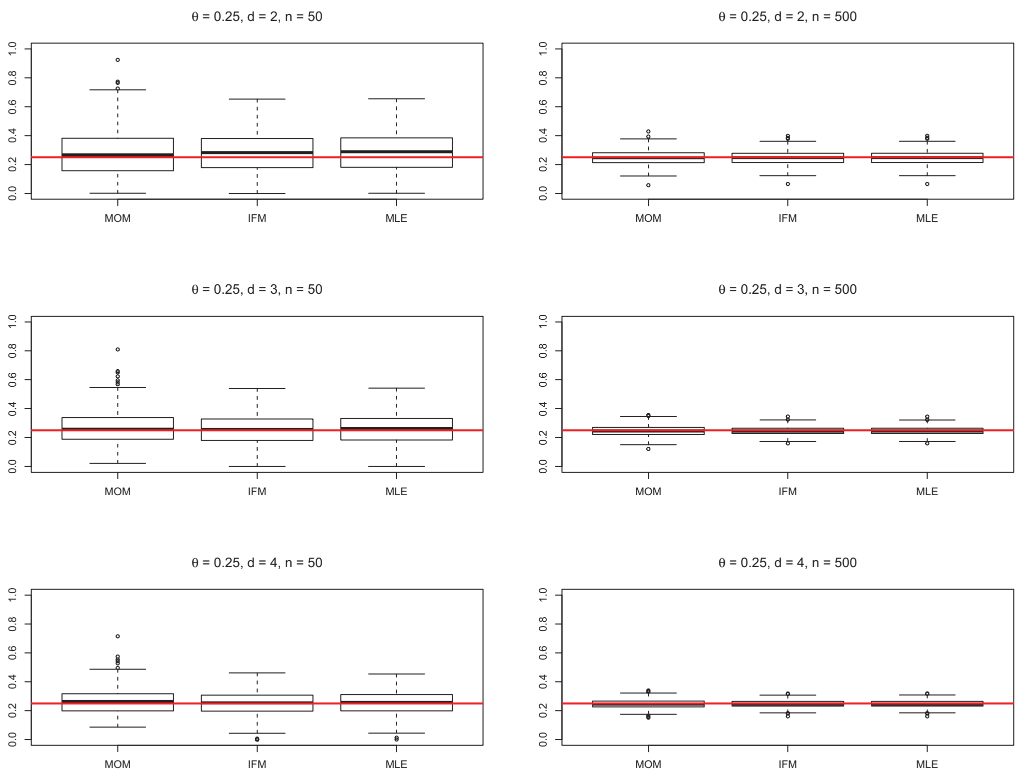

Figure 6.3: Estimation results for θ from the method of moments (left), inference function for margins (middle) and maximum likelihood estimation (right) in indicated scenario.

The three stations exhibit moderate dependence. In particular, defining the sample correlation as

$$R_{jk} = S_{jk}/\sqrt{\bar{X}_j \bar{X}_k},$$

for $j, k \in \{1, 2, 3\}$, the sample statistics are found to be $R_{12} = 0.26$, $R_{13} = 0.25$ and $R_{23} = 0.34$. The data provides evidence of Poisson margins. Indeed, the marginal Poisson assumption can be readily checked using a chi-squared goodness of fit test. Using the `gofstat` built in function in R, the latter tests yields a p-value of 0.91 for X_1, 0.40 for X_2 and 0.34 for X_3. The QQ-plots for the three margins also further support the assumption of Poisson-distributed margins, as shown in Figure 6.12.

The estimation techniques outlined in Section 6.3 were implemented for the BIXI data. The results are summarized in Table 6.1; the estimates are provided along with 95% bootstrap confidence intervals. For comparison, the classical trivariate Poisson model based on a single common shock was also implemented. The estimation results in the latter model are provided in Table 6.2. Note that the confidence intervals are based on 1000 bootstrap replications, although one iteration was omitted due to the IFM method not reaching convergence within the maximum number of

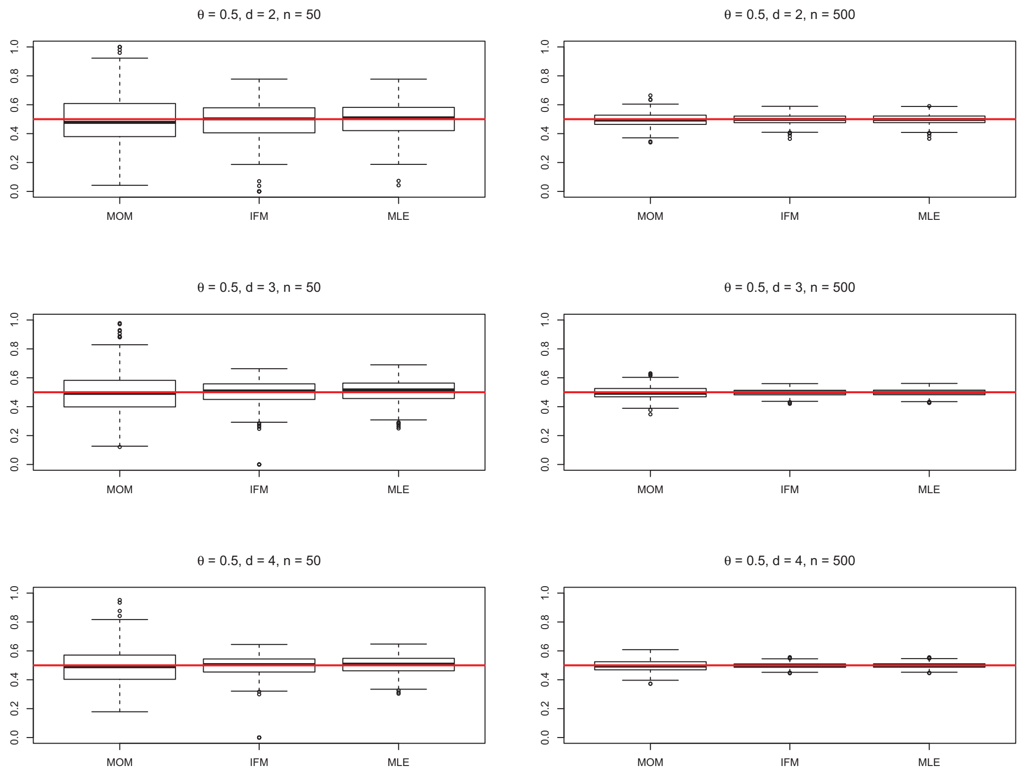

Figure 6.4: Estimation results for θ from the method of moments (left), inference function for margins (middle) and maximum likelihood estimation (right) in indicated scenario.

iterations. As such, the bootstrap confidence intervals presented are based on 999 iterations. Similarly to what was done in Chapter 4, the MM estimate $\tilde{\theta}$ was used as the starting value for the dependence parameter in the implementation of the likelihood-based methods. Additionally, the MM estimates, $(\bar{X}_1, \bar{X}_2, \bar{X}_3)$, were used as the starting values for the marginal parameters for the maximum likelihood estimation in the proposed $\mathcal{MP}$ model.

The estimation results for the proposed trivariate Poisson model are summarized in Table 6.1. The results reveal that the moment-based estimation implies a stronger dependence structure in comparison to the two likelihood-based methods. Indeed, the pairwise correlations resulting from the MM estimates are all higher than the corresponding correlations computed using the IFM and ML estimates. This was also observed in the data application for the bivariate model in Chapter 4. Note that the similarities between the implied correlations ρ_{13} and ρ_{23} are due to the fact that the marginal parameter estimates of λ_1 and λ_2 are relatively similar across all methods. The implementation of the classical trivariate Poisson model revealed similar patterns, as shown in Table 6.2.

Recall that the observed sample correlations were found to be $R_{12} = 0.26$, $R_{13} = 0.25$ and $R_{23} = 0.34$. Based on the results from the proposed $\mathcal{MP}$ model summarized in Table 6.1, it can

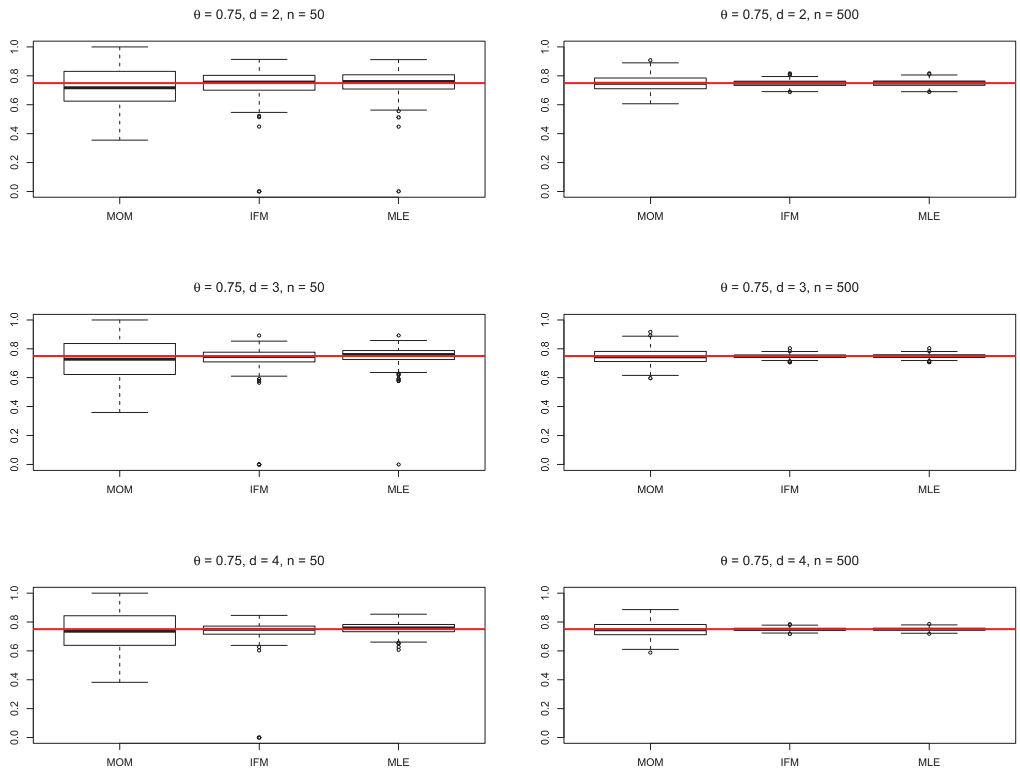

Figure 6.5: Estimation results for θ from the method of moments (left), inference function for margins (middle) and maximum likelihood estimation (right) in indicated scenario.

be seen that the sample correlations are contained in each of their respective bootstrap confidence intervals under the MM. However, the bootstrap confidence intervals produced by both likelihood-based methods do not encompass R_{23}. Note that R_{23} does not even fall within the 99% bootstrap confidence intervals for ρ_{23} based on the IFM and ML estimates, which are both approximately $(0.08, 0.33)$. The same phenomena is observed with the results from the classical trivariate model. However, in the proposed model R_{23} is closer to the upper bound of the confidence intervals resulting from both likelihood-based methods as compared to that in the classical model.

It is not surprising that both models present some difficulties in quantifying the dependence between X_2 and X_3. The data suggest that λ_1 and λ_2 are of similar magnitude with $\bar{X}_1 = 1.45$ and $\bar{X}_2 = 1.61$. As a result, both the classical and proposed trivariate Poisson models imply that ρ_{13} and ρ_{23} should be relatively similar. However, the data suggest that X_2 and X_3 have a stronger association in comparison to that exhibited by X_1 and X_3. Seemingly, both models are unable to capture this distinct correlation structure. Nonetheless, the method of moments is able to better quantify the dependence structure since this estimation approach is based on the observed pairwise sample covariances.

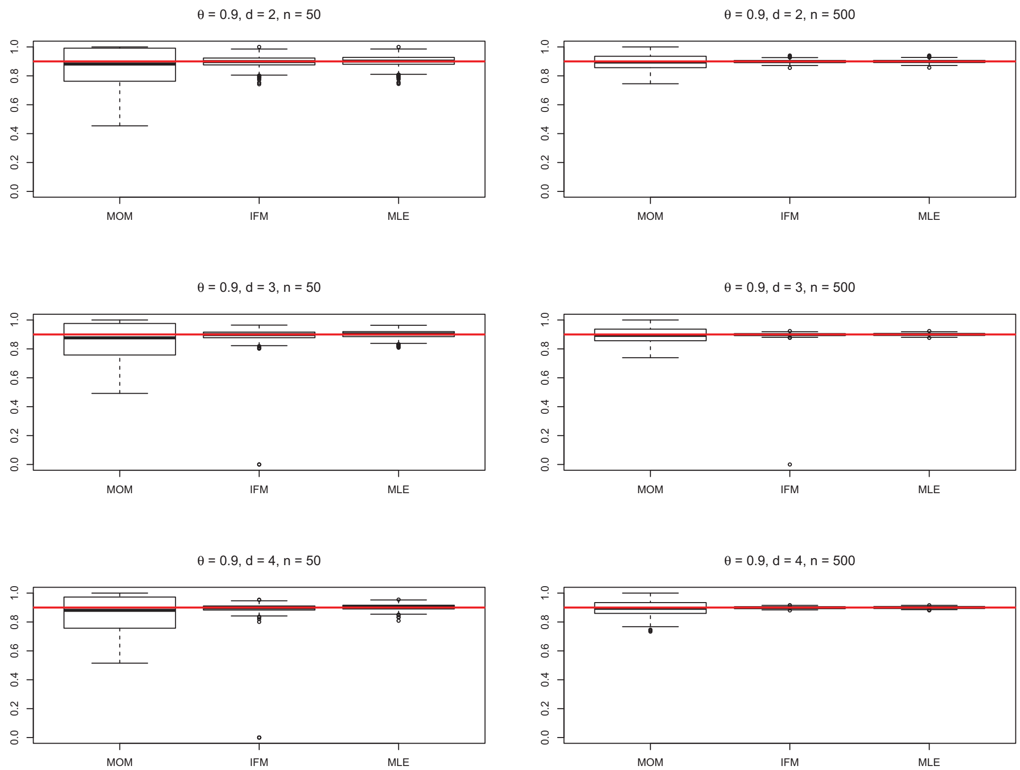

Figure 6.6: Estimation results for θ from the method of moments (left), inference function for margins (middle) and maximum likelihood estimation (right) in indicated scenario.

Note that if the sample correlation is calculated as $R^*_{jk} = S_{jk}/\sqrt{S_j^2 S_k^2}$, where S_j^2 denotes the sample variance given by

$$S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2,$$

slightly different values are obtained. In particular, using this formulation it is found that $R^*_{12} = 0.22$, $R^*_{13} = 0.25$ and $R^*_{23} = 0.29$. Each of these sample correlations are contained within their respective bootstrap confidence intervals under the proposed $\mathcal{MP}$ model; this holds for the intervals obtained from all estimation methods. This is not the case in the classical model.

Arguably, both the classical and the proposed trivariate Poisson models are not entirely satisfactory in capturing the particular correlation structure inherent in the data. As was alluded to earlier in this chapter, the proposed $\mathcal{MP}$ model construction relies on a single dependence parameter θ and consequently restricts the implied correlation structure to a certain extent. This is a shortcoming of both the classical and proposed model formulations. Certainly, it is possible to define a multivariate Poisson model with a more flexible dependence structure while retaining the comonotonic shock construction. To this end, a more flexible trivariate model is briefly introduced

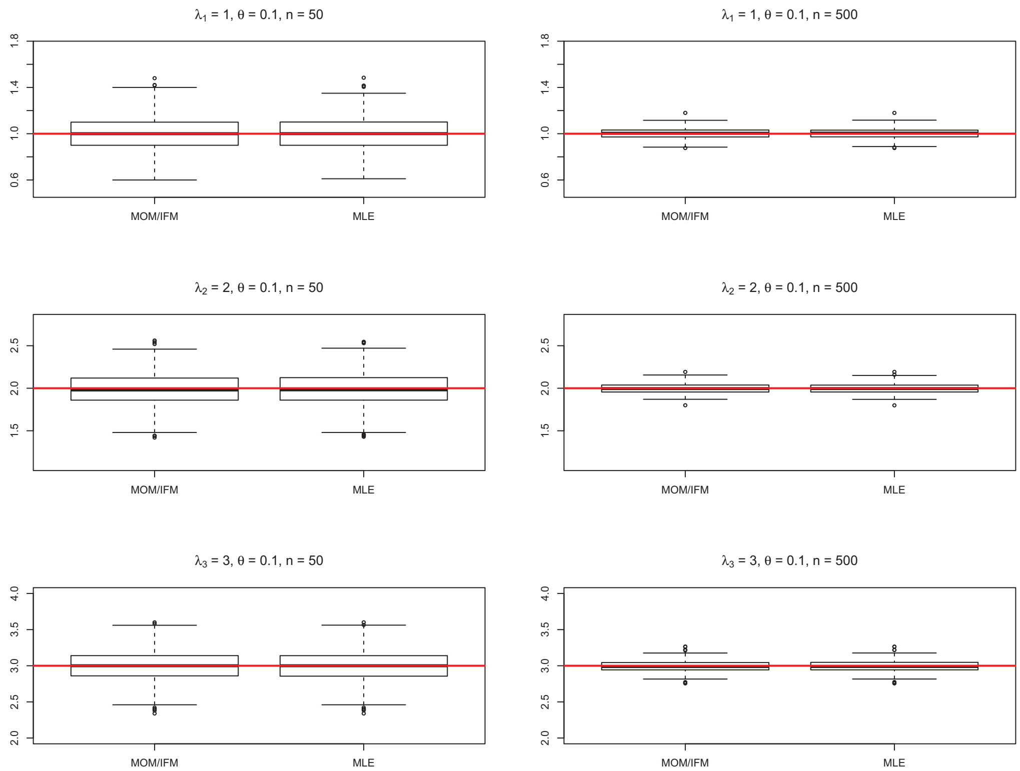

Figure 6.7: Estimation results for $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$ from the method of moments/inference function for margins (left) and maximum likelihood estimation (right) in indicated scenario.

in Chapter 7.

Table 6.1: Estimation results for the proposed trivariate Poisson model applied to the BIXI data. Estimates are given with 95% bootstrap confidence intervals in parentheses.

	λ_1	λ_2	λ_3	θ
MM	1.45 (1.26, 1.65)	1.61 (1.39, 1.86)	3.14 (2.89, 3.41)	0.31 (0.19, 0.43)
IFM	1.45 (1.26, 1.65)	1.61 (1.39, 1.86)	3.14 (2.89, 3.41)	0.24 (0.14, 0.35)
MLE	1.44 (1.26, 1.65)	1.63 (1.39, 1.87)	3.15 (2.89, 3.42)	0.24 (0.14, 0.35)

	ρ_{12}	ρ_{13}	ρ_{23}
MM	0.30 (0.18, 0.41)	0.27 (0.15, 0.39)	0.27 (0.15, 0.39)
IFM	0.23 (0.13, 0.33)	0.20 (0.11, 0.31)	0.20 (0.11, 0.31)
MLE	0.23 (0.13, 0.33)	0.20 (0.11, 0.31)	0.20 (0.11, 0.31)

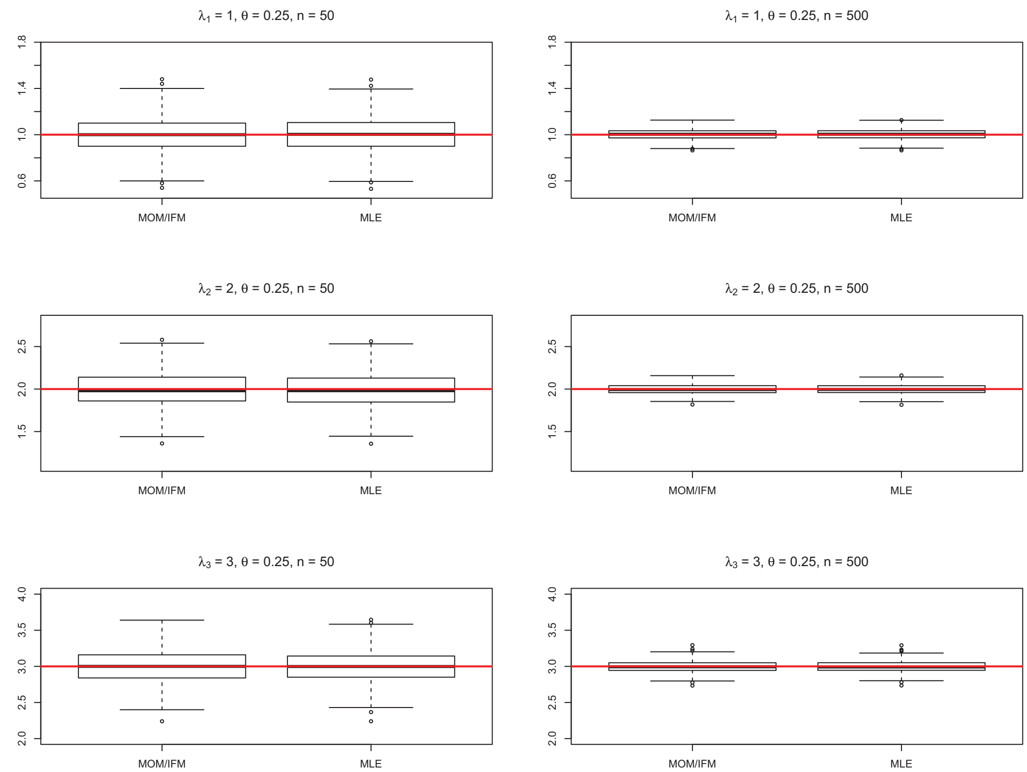

Figure 6.8: Estimation results for $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$ from the method of moments/inference function for margins (left) and maximum likelihood estimation (right) in indicated scenario.

Table 6.2: Estimation results for the classical trivariate Poisson model applied to the BIXI data. Estimates are given with 95% bootstrap confidence intervals in parentheses.

	λ_1	λ_2	λ_3	ξ
MM	$1.45\ (1.26, 1.65)$	$1.61\ (1.39, 1.86)$	$3.14\ (2.89, 3.41)$	$0.56\ (0.33, 0.78)$
MLE	$1.45\ (1.26, 1.65)$	$1.61\ (1.39, 1.86)$	$3.14\ (2.89, 3.41)$	$0.37\ (0.20, 0.55)$

	ρ_{12}	ρ_{13}	ρ_{23}
MM	$0.37\ (0.22, 0.52)$	$0.26\ (0.16, 0.37)$	$0.25\ (0.15, 0.35)$
MLE	$0.24\ (0.13, 0.36)$	$0.17\ (0.09, 0.26)$	$0.16\ (0.09, 0.25)$

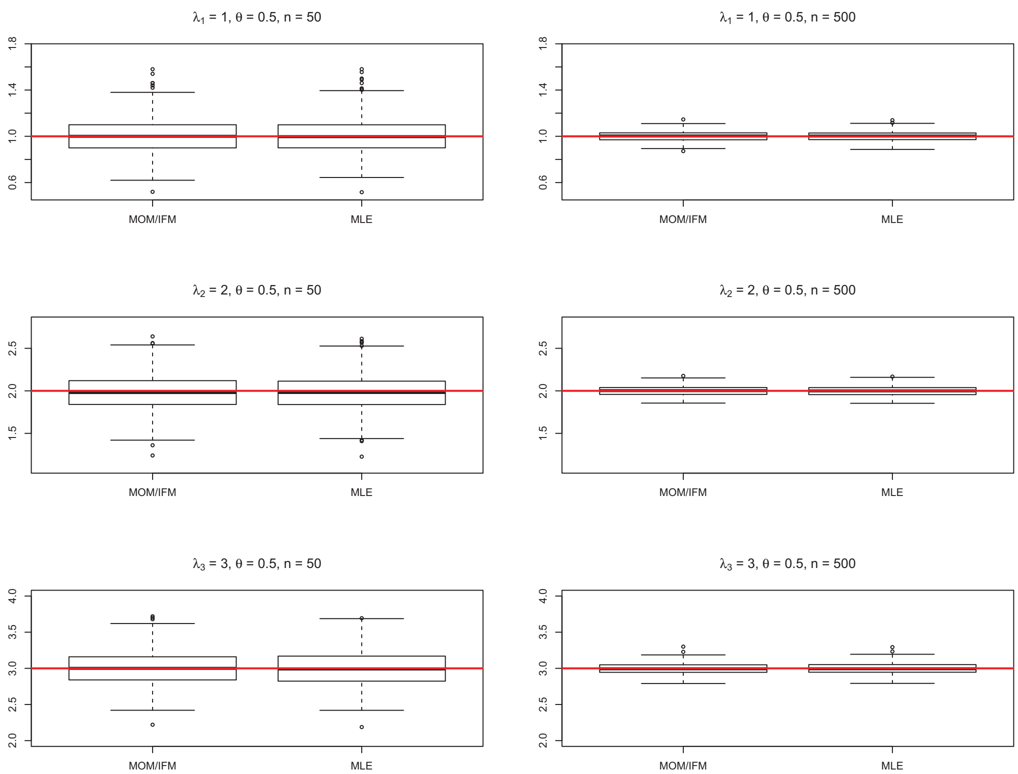

Figure 6.9: Estimation results for $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$ from the method of moments/inference function for margins (left) and maximum likelihood estimation (right) in indicated scenario.

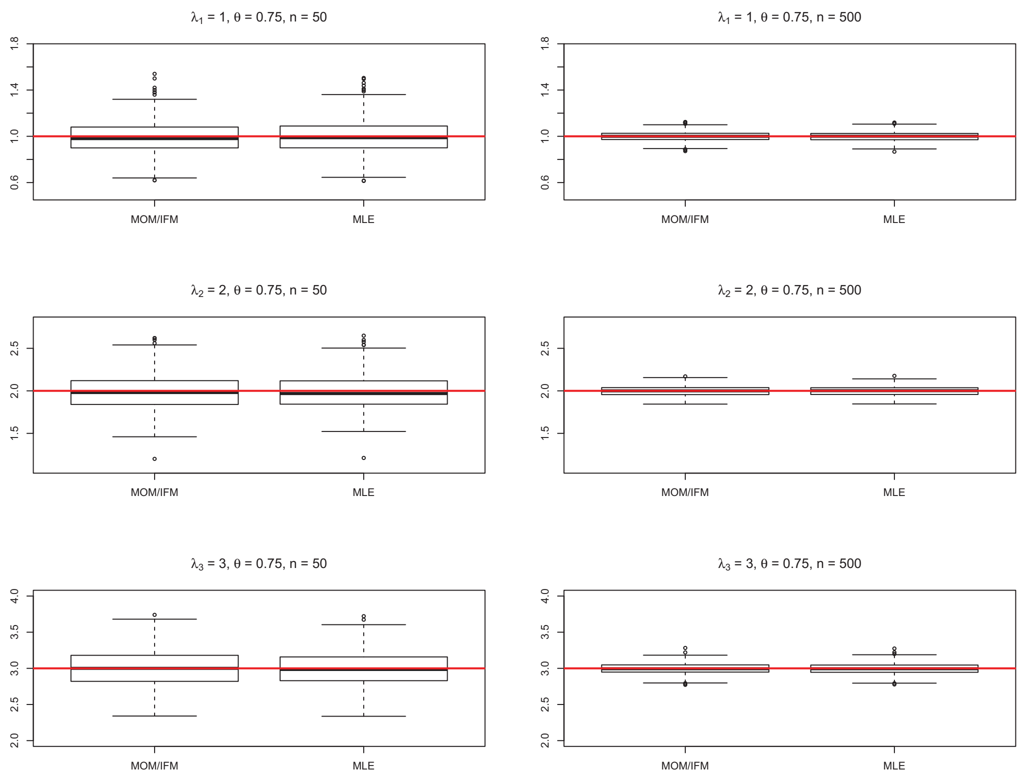

Figure 6.10: Estimation results for $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$ from the method of moments/inference function for margins (left) and maximum likelihood estimation (right) in indicated scenario.

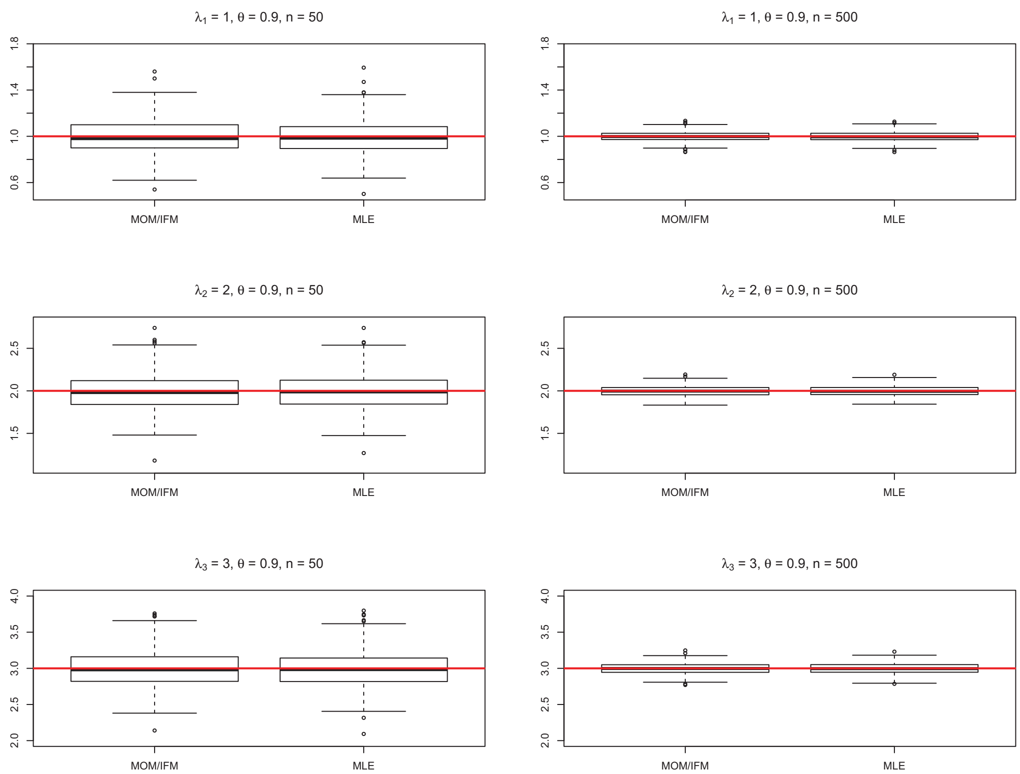

Figure 6.11: Estimation results for $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$ from the method of moments/inference function for margins (left) and maximum likelihood estimation (right) in indicated scenario.

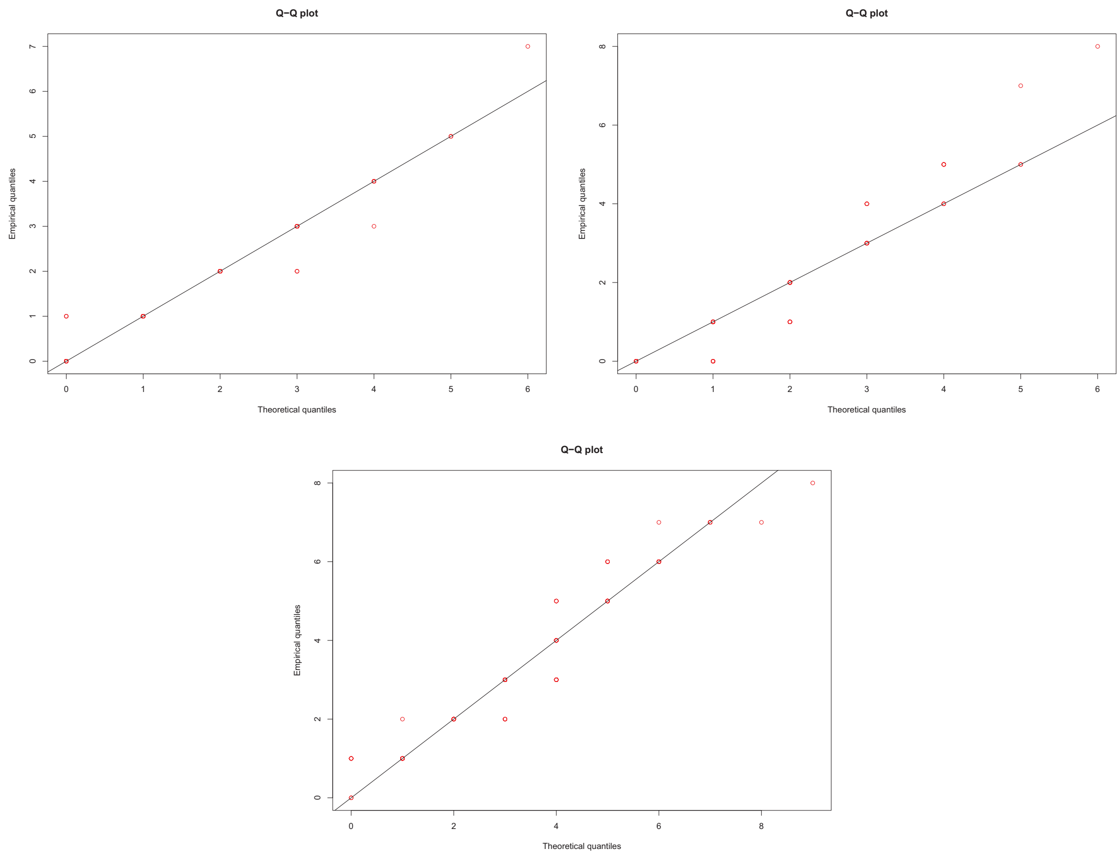

Figure 6.12: QQ plots assessing the marginal Poisson assumption for variables X_1 (top left), X_2 (top right) and X_3 (bottom). Theoretical and empirical quantiles are on the x- and y-axis, respectively.

7

Final Conclusion and Future Work

In this thesis, a novel construction for a fully flexible bivariate Poisson model with positive dependence was presented. The proposed $\mathcal{BP}$ family spans the full spectrum of possible positive dependence, with correlation ranging from 0 to $\rho_{\max}(\lambda_1, \lambda_2)$. It was shown that as the dependence parameter θ increases, the strength of the dependence increases accordingly in the sense of the PQD ordering. Several distributional properties were developed for the $\mathcal{BP}$ family and various estimation techniques were described in detail. Simulation studies as well as a data illustration validated the utility of the $\mathcal{BP}$ model. The proposed comonotonic shock model provides an appealing alternative to the classical bivariate Poisson model as its added flexibility accommodates strong degrees of (positive) dependence while retaining an intuitive and interpretable stochastic construction.

The proposed $\mathcal{BP}$ model stems from the notion of comonotonic shocks and as such extends naturally to higher dimensions. This allowed to define a multivariate Poisson model for positive dependence. It was shown that many of the distributional properties in the bivariate setting remain valid in arbitrary dimension $d \geqslant 2$. Estimation techniques were adapted for higher dimensions and tested through various simulations. To illustrate the usefulness of the proposed multivariate model, the $\mathcal{MP}_3$ model was applied to data provided by BIXI, Montréal's bike sharing network. As was the case in dimension 2, the proposed $\mathcal{MP}$ family provides a more flexible multivariate Poisson model than the classical version.

Analogously to the comonotonic shock model, the notion of counter-monotonicity allowed to define a bivariate Poisson model for negative dependence. Many of the properties in the proposed $\mathcal{BP}^-$ family parallel those in the comonotonic shock model. In particular, the dependence parameter regulates the strength of the dependence, with $\theta = 0$ corresponding to independence and $\theta = 1$ yielding perfect negative dependence. Estimation proved to be somewhat more challenging in the case of the $\mathcal{BP}^-$ model, as was highlighted through various simulation studies. Nonetheless, the proposed counter-monotonic shock model accommodates any degree of negative dependence while retaining an interpretable stochastic representation. This is a considerable advantage of the proposed model as an alternative to other constructions, such as copulas or mixture models.

7.1 Future work

Of course there is still room for improvements and further extensions in the proposed multivariate Poisson model framework. Certainly, there are numerous projects that stem naturally from the work done in this thesis. The following subsections outline some of these potential future research subjects.

7.1.1 A more flexible multivariate Poisson model

Although the proposed multivariate Poisson model discussed in Chapter 6 provides a more flexible definition than the classical common shock representation (see Chapter 2), the $\mathcal{MP}_d(\Lambda, \theta)$ family is somewhat restrictive in that there is a single dependence parameter and, for $d > 2$, is in fact not fully flexible. For example, according to the construction outlined in (6.1), if any pair (X_i, X_j), $i \neq j$, exhibits perfect positive dependence such that $\rho_\theta(X_i, X_j) = \rho_{\max}$, the dependence parameter θ must necessarily be equal to 1 so that $\mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$ represents a comonotonic random vector where all pairwise correlations attain the upper bound $\rho_{\max}$. Similarly, when $\theta = 0$, the random vector $\mathbf{X} \sim \mathcal{MP}_d(\Lambda, \theta)$ consists of independent components so that all pairwise correlations are 0. This follows from Lemma 6.1, which ensures that the pairwise correlations $\rho_\theta(X_i, X_j)$ are increasing functions θ. Clearly, a model based on a single dependence parameter imposes certain restrictions on the covariance structure.

In order to address this shortcoming, the construction in (6.1) can be adapted to define a more complex model. In particular, the added flexibility should allow for each pairwise correlation $\text{corr}(X_i, X_j)$, $i \neq j$, to take on any value in $[0, \rho_{\max}]$ without implying restrictions on the remaining pairs (X_r, X_s), $r \neq s \neq i \neq j \in \{1, \ldots, d\}$. In dimension $d = 3$, a more flexible model could be achieved via the following construction:

$$\begin{aligned} X_1 &= Y_1 + Z_{12} + Z_{13} + Z_1, \\ X_2 &= Y_2 + Z_{21} + Z_{23} + Z_2, \\ X_3 &= Y_3 + Z_{31} + Z_{32} + Z_3. \end{aligned} \tag{7.1}$$

In this representation, Y_1, Y_2, Y_3 consist of independent Poisson random variables with respective rates $(1 - \alpha - \beta - \theta)\lambda_1$, $(1 - \alpha - \gamma - \theta)\lambda_2$ and $(1 - \beta - \gamma - \theta)\lambda_3$, which are also independent of each

$$\begin{aligned}
(Z_{12}, Z_{21}) &\sim \mathcal{M}\left\{\mathcal{P}(\alpha\lambda_1), \mathcal{P}(\alpha\lambda_2)\right\}, \\
(Z_{13}, Z_{31}) &\sim \mathcal{M}\left\{\mathcal{P}(\beta\lambda_1), \mathcal{P}(\beta\lambda_3)\right\}, \\
(Z_{23}, Z_{32}) &\sim \mathcal{M}\left\{\mathcal{P}(\gamma\lambda_2), \mathcal{P}(\gamma\lambda_3)\right\}, \\
(Z_1, Z_2, Z_3) &\sim \mathcal{M}\left\{\mathcal{P}(\theta\lambda_1), \mathcal{P}(\theta\lambda_2), \mathcal{P}(\theta\lambda_3)\right\}.
\end{aligned}$$

In terms of independent $\mathcal{U}(0,1)$ random variables V_1, V_2, V_3, U_{12}, U_{13}, U_{23}, U, the formulation in (7.1) can be written as

$$\begin{aligned}
X_1 &= G^{-1}_{(1-\alpha-\beta-\theta)\lambda_1}(V_1) + G^{-1}_{\alpha\lambda_1}(U_{12}) + G^{-1}_{\beta\lambda_1}(U_{13}) + G^{-1}_{\theta\lambda_1}(U), \\
X_2 &= G^{-1}_{(1-\alpha-\gamma-\theta)\lambda_2}(V_2) + G^{-1}_{\alpha\lambda_2}(U_{12}) + G^{-1}_{\gamma\lambda_2}(U_{23}) + G^{-1}_{\theta\lambda_2}(U), \\
X_3 &= G^{-1}_{(1-\beta-\gamma-\theta)\lambda_3}(V_3) + G^{-1}_{\beta\lambda_3}(U_{13}) + G^{-1}_{\gamma\lambda_3}(U_{23}) + G^{-1}_{\theta\lambda_3}(U),
\end{aligned}$$

subject to the constraints

$$0 \leqslant \alpha + \beta + \theta \leqslant 1, \quad 0 \leqslant \alpha + \gamma + \theta \leqslant 1, \quad 0 \leqslant \beta + \gamma + \theta \leqslant 1,$$

where $\alpha, \beta, \gamma, \theta \in [0,1]$ and $\lambda_s > 0$ for $s \in \{1,2,3\}$.

The above construction thus consists of convoluting a vector of independent Poisson random variables (Y_1, Y_2, Y_3) with all pairwise comonotonic Poisson shocks (Z_{12}, Z_{21}), (Z_{13}, Z_{31}), (Z_{23}, Z_{32}) as well as a global comonotonic Poisson shock (Z_1, Z_2, Z_3). In this trivariate model, there are four dependence parameters α, β, γ, and θ in addition to the three marginal parameters $(\lambda_1, \lambda_2, \lambda_3)$. When $\theta = 1$, it follows that $(X_1, X_2, X_3) \sim \mathcal{M}\left\{\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2), \mathcal{P}(\lambda_3)\right\}$, i.e., the upper Fréchet–Hoeffding bound is attained. In contrast if, say, $\alpha = 1$, then (X_1, X_2) are comonotonic and independent of the third component X_3. Conceivably, the benefit of the added flexibility in this more complex model could be outweighed by difficulties in estimation. A thorough examination of this more complex model is left for future research.

7.1.2 Regression models

The data illustrations considered in this thesis were rather simplistic applications of the proposed multivariate Poisson shock models. In more practical settings, data usually consists of not only the response variables of interest but also several covariates. Although not explicitly detailed in this thesis, it is possible to extend each of the proposed models to allow for covariate effects. As outlined in Section 10.2 of Joe (1997), under certain regularity conditions, the IFM estimation approach can accommodate the inclusion of covariates. Moreover, maximum likelihood estimation could also readily allow for covariate effects by rewriting the parameters (Λ, θ) in terms of regression parameters in the log-likelihood. Of course, in doing so the estimation procedure will become more complex and potentially numerically infeasible.

As pointed in Joe (1997), although it is typically straightforward to incorporate covariate effects into marginal parameters (e.g., via marginal generalized linear models), it is not always clear whether the dependence parameter should be a function of the covariates and moreover what form that function should assume. Regression models in the proposed classes of multivariate Poisson shock models is also deferred to future work.

7.1.3 A multivariate model for negative dependence

Extending the bivariate counter-monotonic Poisson shock model to higher dimensions is not immediately obvious. This follows from the difficulty in defining the notion of perfect negative dependence in dimensions $d > 2$. As was discussed in Chapter 3, the concept of counter-monotonicity does not extend to higher dimensions as the Fréchet-Hoeffding lower boundary distribution F_L is, in general, not a proper cumulative distribution function for $d > 2$.

Nonetheless, it is possible to define a multivariate Poisson model exhibiting negative dependence in dimensions $d > 2$. In particular, consider the following trivariate Poisson model:

$$\begin{aligned}
X_1 &= Y_1 + Z_{12} + Z_{13} \\
&= G^{-1}_{(1-\alpha_{12}-\alpha_{13})\lambda_1}(V_1) + G^{-1}_{\alpha_{12}\lambda_1}(U_{12}) + G^{-1}_{\alpha_{13}\lambda_1}(U_{13}), \\
X_2 &= Y_2 + Z_{21} + Z_{23} \\
&= G^{-1}_{(1-\alpha_{12}-\alpha_{23})\lambda_2}(V_2) + G^{-1}_{\alpha_{12}\lambda_2}(1-U_{12}) + G^{-1}_{\alpha_{23}\lambda_2}(U_{23}), \\
X_3 &= Y_3 + Z_{31} + Z_{32} \\
&= G^{-1}_{(1-\alpha_{13}-\alpha_{23})\lambda_3}(V_3) + G^{-1}_{\alpha_{13}\lambda_3}(1-U_{13}) + G^{-1}_{\alpha_{23}\lambda_3}(1-U_{23}),
\end{aligned} \tag{7.2}$$

subject to the constraints

$$\alpha_{12}, \alpha_{13}, \alpha_{23} \in [0,1], \quad 0 \leqslant \alpha_{12} + \alpha_{13} \leqslant 1, \quad 0 \leqslant \alpha_{12} + \alpha_{23} \leqslant 1, \quad 0 \leqslant \alpha_{13} + \alpha_{23} \leqslant 1.$$

The above construction consists of convoluting a trivariate vector of independent Poisson random variables (Y_1, Y_2, Y_3) with all possible pairs of counter-monotonic shocks, (Z_{12}, Z_{21}), (Z_{13}, Z_{31}), (Z_{23}, Z_{32}). In this framework, each pair exhibits negative correlation. Indeed, it is straightfoward to see that each pair $(X_i, X_j) \sim \mathcal{BP}^-$ with marginal parameters (λ_i, λ_j) and dependence parameter α_{ij} for $i < j \in \{1, 2, 3\}$.

Note that in this representation, only one pair can exhibit perfect negative dependence since only one of the dependence parameters α_{12}, α_{13}, α_{23} can be equal to 1. If say, $\alpha_{12} = 1$, then necessarily $\alpha_{13} = \alpha_{23} = 0$ and thus $(X_1, X_2) \sim \mathcal{W}\{\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2)\}$ and $(X_1, X_2) \perp X_3$. Of course, this construction can be further extended to higher dimensions.

As a side note, an alternative trivariate Poisson model formulation consists of the following:

$$\begin{aligned}
X_1 &= Y_1 + Z_{12} + Z_{13} + Z_1 \\
&= G^{-1}_{(1-\alpha_{12}-\alpha_{13}-\theta)\lambda_1}(V_1) + G^{-1}_{\alpha_{12}\lambda_1}(U_{12}) + G^{-1}_{\alpha_{13}\lambda_1}(U_{13}) + G^{-1}_{\theta\lambda_1}(U), \\
X_2 &= Y_2 + Z_{21} + Z_{23} + Z_2 \\
&= G^{-1}_{(1-\alpha_{12}-\alpha_{23}-\theta)\lambda_2}(V_2) + G^{-1}_{\alpha_{12}\lambda_2}(1-U_{12}) + G^{-1}_{\alpha_{23}\lambda_2}(U_{23}) + G^{-1}_{\theta\lambda_2}(U), \\
X_3 &= Y_3 + Z_{31} + Z_{32} + Z_3 \\
&= G^{-1}_{(1-\alpha_{13}-\alpha_{23}-\theta)\lambda_3}(V_3) + G^{-1}_{\alpha_{13}\lambda_3}(1-U_{13}) + G^{-1}_{\alpha_{23}\lambda_3}(1-U_{23}) + G^{-1}_{\theta\lambda_3}(U),
\end{aligned}$$

subject to the constraints

$$\alpha_{12}, \alpha_{13}, \alpha_{23}, \theta \in [0,1], \quad 0 \leqslant \alpha_{12}+\alpha_{13}+\theta \leqslant 1, \quad 0 \leqslant \alpha_{12}+\alpha_{23}+\theta \leqslant 1, \quad 0 \leqslant \alpha_{13}+\alpha_{23}+\theta \leqslant 1.$$

In this setting, when $\theta = 0$ the above construction coincides with that given in (7.2). In contrast, when $\theta = 1$ the model reduces to the proposed $\mathcal{MP}_3(\Lambda, \theta)$ model discussed in Chapter 6. Clearly, there are many ways to construct flexible multivariate Poisson models based on the notion of comonotonic and counter-monotonic shocks. However, the greater the flexibility in the induced dependence structure, the more complex the model becomes.

List of Publications

Published:

C. Genest, M. Mesfioui and J. Schulz. “A new bivariate Poisson common shock model covering all possible degrees of dependence." Statistics & Probability Letters 140: 202–209, 2018.

Submitted:

J. Schulz, C. Genest and M. Mesfioui. “A multivariate Poisson model based on comonotonic shocks.” *In progress. Expected submission: December 2018*.

Bibliography

K. Adamidis and S. Loukas. ML estimation in the bivariate Poisson distribution in the presence of missing values via the EM algorithm. *Journal of Statistical Computation and Simulation*, 50: 163–172, 1994.

B. P. Akpoue and J.-F. Angers. Some contributions on the multivariate Poisson–Skellam probability distribution. *Communications in Statistics - Theory and Methods*, 46:49–68, 2017.

F. E. AlMuhayfith, A. A. Alzaid, and M. A. Omair. On bivariate Poisson regression models. *Journal of King Saud University - Science*, 28:178–189, 2016.

P. Berkhout and E. Plug. A bivariate Poisson count data model using conditional probabilities. *Statistica Neerlandica*, 58:349–364, 2004.

L. Bermúdez and D. Karlis. A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Computational Statistics & Data Analysis*, 56:3988–3999, 2012.

P. Billingsley. *Probability and Measure*. Wiley, New York, third edition, 1995.

T. Bouezmarni, M. Mesfioui, and A. Tajar. On concordance measures for discrete data and dependence properties of Poisson model. *Journal of Probability and Statistics*, pages 1–15, 2009.

J. T. Campbell. The Poisson correlation function. *Proceedings of the Edinburgh Mathematical Society*, 4:18–26, 1934.

G. Casella and R.L. Berger. *Statistical Inference – 2nd Edition*. Duxbury, Pacific Grove, CA, 2002.

S. Chib and R. Winkelmann. Markov Chain Monte Carlo analysis of correlated count data. *Journal of Business & Economic Statistics*, 19:428–435, 2001.

J. Cuenin, B. Jørgensen, and C. Kokonendji. Simulations of full multivariate Tweedie with flexible dependence structure. *Computational Statistics*, 31:1477–1492, Dec 2016.

J. Dhaene, S. Vanduffel, M.J. Goovaerts, R. Kaas, Q. Tang, and D. Vyncke. Risk measures and comonotonicity: A review. *Stochastic Models*, 22:573–606, 2006.

P. Embrechts and M. Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77:423–432, 2013.

F. Famoye. A new bivariate generalized Poisson distribution. *Statistica Neerlandica*, 64:112–124, 2010a.

F. Famoye. On the bivariate negative binomial regression model. *Journal of Applied Statistics*, 37: 969–981, 2010b.

F. Famoye. Comparisons of some bivariate regression models. *Journal of Statistical Computation and Simulation*, 82:937–949, 2012.

F. Famoye. A multivariate generalized Poisson regression model. *Communications in Statistics - Theory and Methods*, 44:497–511, 2015.

F. Famoye and P. C. Consul. Bivariate generalized Poisson distribution with some applications. *Metrika*, 42:127–138, Dec 1995.

P. Faroughi and N. Ismail. Bivariate zero-inflated generalized Poisson regression model with flexible covariance. *Communications in Statistics - Theory and Methods*, 46:7769–7785, 2017.

T.S. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall, London, 1996.

C. Genest and J. Nešlehová. A primer on copulas for count data. *ASTIN Bulletin*, 37:475–515, 2007.

E. Gómez-Déniz, J. M. Sarabia, and N. Balakrishnan. A multivariate discrete Poisson–Lindley distribution: Extensions and actuarial applications. *ASTIN Bulletin*, 42:655–678, 2012.

R.C. Griffiths, R.K. Milne, and R. Wood. Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics*, 21:238–255, 1979.

S. Gurmu and J. Elder. Generalized bivariate count data regression models. *Economics Letters*, 68:31–36, 2000.

S. Gurmu and J. Elder. A bivariate zero-inflated count data regression model with unrestricted correlation. *Economics Letters*, 100:245–248, 2008.

P. Holgate. Estimation for the bivariate Poisson distribution. *Biometrika*, 51:241–287, 1964.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.

H. Joe. Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94:401–419, 2005.

H. Joe. *Dependence Modeling with Copulas*. CRC Press, London, 2015.

K. Kano and K. Kawamura. On recurrence relations for the probability function of multivariate generalized Poisson distribution. *Communications in Statistics - Theory and Methods*, 20:165–178, 1991.

D. Karlis. An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics*, 30:63–77, 2003.

D. Karlis and L. Meligkotsidou. Multivariate Poisson regression with covariance structure. *Statistics and Computing*, 15:255–265, 2005.

D. Karlis and L. Meligkotsidou. Finite mixtures of multivariate Poisson distributions with application. *Journal of Statistical Planning and Inference*, 137:1942–1960, 2007.

D. Karlis and E. Xekalaki. Mixed Poisson distributions. *International Statistical Review*, 73: 35–58, 2005.

K. Kawamura. The structure of bivariate Poisson distribution. *Kodai Mathematical Seminar Reports*, 25:246–256, 1973.

K. Kawamura. The structure of trivariate Poisson distribution. *Kodai Mathematical Seminar Reports*, 28:1–8, 1976.

K. Kawamura. The structure of multivariate Poisson distribution. *Kodai Mathematical Seminar Reports*, 2:337–345, 1979.

K. Kawamura. A note on the recurrent relations for the bivariate Poisson distribution. *Kodai Mathematical Journal*, 8:70–78, 1985.

S. Kocherlakota and K. Kocherlakota. Regression in the bivariate poisson distribution. *Communications in Statistics - Theory and Methods*, 30:815–825, 2001.

A.S. Krishnamoorthy. Multivariate binomial and Poisson distributions. *Sankhyā: The Indian Journal of Statistics (1933–1960)*, 11:117–124, 1951.

C. D. Lai. Construction of bivariate distributions by a generalised trivariate reduction technique. *Statistics & Probability Letters*, 25:265–270, 1995.

J. Lakshminarayana, S.N.N. Pandit, and K.S. Rao. On a bivariate Poisson distribution. *Communications in Statistics - Theory and Methods*, 28:267–276, 1999.

E.L. Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, 37:1137–1153, 1966.

C.-S. Li, J. C. Lu, J. Park, K. Kim, P. A. Brinkley, and J. P. Peterson. Multivariate zero-inflated Poisson models and their applications. *Technometrics*, 41:29–38, 1999.

F. Lindskog and A. J. McNeil. Common Poisson shock models: Applications to insurance and credit risk modelling. *ASTIN Bulletin*, 33:209–238, 2003.

S. Loukas. Some methods of estimation for a trivariate Poisson distribution. *Applicationes Mathematicae*, 21:503–510, 01 1993.

S. Loukas and H. Papageorgiou. On a trivariate Poisson distribution. *Applications of Mathematics*, 36:432–439, 1991.

S. Loukas, C. D. Kemp, and H. Papageorgiou. Even point estimation for the bivariate Poisson distribution. *Biometrika*, 73:222–223, 1986.

D. M. Mahamunulu. A note on regression in the multivariate Poisson distribution. *Journal of the American Statistical Association*, 62:251–258, 1967.

A. W. Marshall and I. Olkin. A family of bivariate distributions generated by the bivariate Bernoulli distribution. *Journal of the American Statistical Association*, 80:332–338, 1985.

G.J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley, Hoboken, NJ, second edition, 2008.

R.J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.

R.B. Nelsen. Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics - Simulation and Computation*, 16:199–208, 1987.

R.B. Nelsen. *An Introduction to Copulas, 2nd edition*. Springer, New York, 2006.

A. K. Nikoloulopoulos and D. Karlis. Modeling multivariate count data using copulas. *Communications in Statistics - Simulation and Computation*, 39:172–187, 2009.

A. K. Nikoloulopoulos and D. Karlis. Regression in a copula model for bivariate count data. *Journal of Applied Statistics*, 37:1555–1568, 2010.

A. Panagiotelis, C. Czado, and H. Joe. Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, 107:1063–1072, 2012.

H. Papageorgiou. On characterizing some bivariate discrete distributions. *Australian Journal of Statistics*, 25:136–144, 1983.

H. Papageorgiou and C. D. Kemp. A method of estimation for some bivariate discrete distributions. *Biometrical Journal*, 30:993–1001, 1988.

D. Pfeifer and J. Nešlehová. Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bulletin*, 34:333–360, 2004.

S.I. Resnick. *Extremes Values, Regular Variation, and Point Processes*. Springer, New York, 2008.

J. M. Sarabia and E. Gómez-Déniz. Multivariate Poisson-Beta distributions with applications. *Communications in Statistics - Theory and Methods*, 40:1093–1108, 2011.

M. Scarsini. On measures of concordance. *Stochastica*, 8:201–218, 1984.

K. F. Sellers, D. S. Morris, and N. Balakrishnan. Bivariate Conway–Maxwell–Poisson distribution: Formulation, properties, and inference. *Journal of Multivariate Analysis*, 150:152–168, 2016.

R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, New York, 1980.

M. Shaked and J.G. Shanthikumar. *Stochastic Orders*. Springer, New York, 2007.

M. S. Smith and M. A. Khaled. Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association*, 107:290–303, 2012.

E. G. Tsionas. Bayesian analysis of the multivariate Poisson distribution. *Communications in Statistics - Theory and Methods*, 28:431–451, 1999.

E. G. Tsionas. Bayesian multivariate Poisson regression. *Communications in Statistics - Theory and Methods*, 30:243–255, 2001.

H. Van Ophem. A general method to estimate correlated discrete random variables. *Econometric Theory*, 15:228–237, 1999.

J. Wakefield. *Bayesian and Frequentist Regression Methods*. Springer, New York, 2013.

J.-F. Walhin. Bivariate ZIP models. *Biometrical Journal*, 43:147–160, 2001.

H. Zamani, P. Faroughi, and N. Ismail. Bivariate generalized Poisson regression model: Applications on health care data. *Empirical Economics*, 51:1607–1621, 2016.

C. Zhang, G. Tian, and X. Huang. Two new bivariate zero-inflated generalized Poisson distributions with a flexible correlation structure. *Statistics Optimization & Information Computing*, 3: 105–137, 2015.

Acronyms and Notation

MM	Method of Moments
IFM	Inference Functions for the Margins
MLE	Maximum Likelihood Estimation
PMF / CDF	Probability mass function / Cumulative distribution function
$\mathcal{P}(\lambda)$	Univariate Poisson distribution with mean λ
$\mathcal{U}(0,1)$	Univariate (standard) Uniform distribution on the interval $(0,1)$
$\mathcal{N}(\mu,\sigma^2)$	Univariate Normal distribution with mean μ and variance σ^2
$\mathcal{M}(F_1,\ldots,F_d)$	Upper Fréchet-Hoeffding boundary distribution with marginal distributions $F_1,\ldots,F_d$
$\mathcal{W}(F_1,\ldots,F_d)$	Lower Fréchet-Hoeffding boundary distribution with marginal distributions $F_1,\ldots,F_d$
$\mathcal{BP}(\Lambda,\theta)$	Proposed bivariate Poisson distribution for positive dependence (Λ: marginal parameters, θ: dependence parameter)
$\mathcal{BP}^-(\Lambda,\theta)$	Proposed bivariate Poisson distribution for negative dependence (Λ: marginal parameters, θ: dependence parameter)
$\mathcal{MP}_d(\Lambda,\theta)$	Proposed multivariate Poisson distribution for positive dependence (d: dimension, Λ: marginal parameters, θ: dependence parameter)
$g_\lambda(x)$	Univariate Poisson PMF, i.e. $g_\lambda(x) = e^{-\lambda}\lambda^x/x!$
$G_\lambda(x)$	Univariate Poisson CDF, i.e. $G_\lambda(x) = \sum_{m=0}^{x} e^{-\lambda}\lambda^m/m!$
$\bar{G}_\lambda(x)$	Univariate Poisson survival function, i.e. $\bar{G}_\lambda(x) = 1 - G_\lambda(x)$
$b_{n,p}(x)$	Binomial PMF, i.e. $b_{n,p}(x) = \binom{n}{x}p^x(1-p)^{n-x}$
$f_{\Lambda,\theta}$, $F_{\Lambda,\theta}$	PMF and CDF in $\mathcal{MP}_d(\Lambda,\theta)$ family $(d = 2,3,\ldots)$
$f^-_{\Lambda,\theta}$, $F^-_{\Lambda,\theta}$	PMF and CDF in $\mathcal{BP}^-(\Lambda,\theta)$ family
$c_{\Lambda,\theta}$	PMF associated with $\mathcal{M}\{\mathcal{P}(\theta\lambda_1),\ldots,\mathcal{P}(\theta\lambda_d)\}$
$c^-_{\Lambda,\theta}$	PMF associated with $\mathcal{W}\{\mathcal{P}(\theta\lambda_1),\ldots,\mathcal{P}(\theta\lambda_d)\}$
$\mathbf{1}(\cdot)$	Indicator function
$[x]_+$	Function $[x]_+ = x\mathbf{1}(x>0)$
$=_d$	Equal in distribution
$\rightsquigarrow$	Convergence in distribution
$\mathbb{N}$	Positive integers $0,1,2,\ldots$