A multivariate Poisson model based on comonotonic shocks

Juliana Schulz\textsuperscript{1}, Christian Genest\textsuperscript{2} and Mhamed Mesfioui\textsuperscript{3}

\textsuperscript{1}Département de sciences de la décision, HEC Montréal, Montréal, Québec, Canada H3T 2A7
\textsuperscript{2}Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada H3A 0B9
\textsuperscript{3}Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, G9A 5H7, Québec, Canada
E-mail: juliana.schulz@hec.ca

Summary

Multivariate count data arise naturally in practice. In analysing such data, it is critical to define a model that can accurately capture the underlying dependence structure between the counts. To this end, this paper develops a multivariate model wherein correlated Poisson margins are generated by a comonotonic shock vector. The proposed model allows for greater flexibility in the dependence structure than that of the classical construction, which relies on the convolution of vectors of common Poisson shock variables. Several probabilistic properties of the multivariate comonotonic shock Poisson model are established, and various estimation strategies are discussed in detail. The model is further studied through simulations, and its utility is highlighted using a real data application.

Key words: comonotonic shock; correlation; dependence structure; multivariate count data; multivariate Poisson model; positive quadrant dependence.

1 Introduction

Multivariate count data are ubiquitous in statistical applications to medicine, risk management and many other fields. An accurate analysis of such data relies critically on the specification of a probabilistic representation of the dependence structure between the counts. A typical approach is to assume Poisson margins and use an appropriate modelling technique to define their joint behaviour. Numerous approaches can be taken to accomplish this, and as a result, there is no unique definition of a multivariate Poisson distribution. Some of the most common solutions are briefly reviewed.

1.1 Standard Approach

A well-known method for constructing correlated Poisson random variables is the multivariate reduction technique. In the most general setting, this approach is based on the convolution of independent Poisson random variables, which can be viewed as a set of multiple common
shocks to the margins. This framework was first described by Mahamunulu (1967) and revisited by several authors, including Kawamura (1976, 1979), Karlis & Meligkotsidou (2007) and Inouye et al. (2017).

In the trivariate case, for example, the multivariate reduction formulation amounts to expressing the components of a vector \((X_1, X_2, X_3)\) of Poisson random variables as sums of mutually independent Poisson random shocks \(Y_1, Y_2, Y_3, Y_{12}, Y_{13}, Y_{23}\) and \(Y_{123}\), namely

\[
X_1 = Y_1 + Y_{12} + Y_{13} + Y_{123}, \quad X_2 = Y_2 + Y_{12} + Y_{23} + Y_{123}, \quad X_3 = Y_3 + Y_{13} + Y_{23} + Y_{123}.
\]

In arbitrary dimension \(d \geq 2\), correlated Poisson margins can be generated by convoluting overlapping subsets of a \((2^d - 1)\)-variate shock vector of mutually independent Poisson random variables.

While the above-mentioned model construction is intuitive and provides some flexibility in the implied correlation structure, deriving the probability distribution is tedious, thereby leading to intractable estimation, as pointed out by Karlis (2003), Karlis & Meligkotsidou (2007) and Inouye et al. (2017). A simpler version of this approach is thus typically considered wherein a single common shock variable is used to induce dependence. In arbitrary dimension \(d\), the model can be written as

\[
X_1 = Y_1 + Z, \ldots, X_d = Y_d + Z
\]

in terms of mutually independent Poisson random variables \(Y_1, \ldots, Y_d, Z\). This formulation is considered by several authors, such as Loukas & Papageorgiou (1991), Loukas (1993), Tsionas (1999) and Karlis (2003). We will hereafter refer to this construction as the classical, or common shock, multivariate Poisson model. Note that when only two variables are involved, the model reduces to the classical bivariate Poisson model of Campbell (1934). This family of bivariate Poisson distributions has been extensively studied in the literature and applied in numerous contexts; for reviews, refer to, for example, Johnson et al. (1997), Inouye et al. (2017) and Schulz (2018).

### 1.2 Limitations

In the classical multivariate Poisson model, the common shock construction in Eqn (1) inherently imposes restrictions on the implied dependence structure: pairwise covariances must be positive and identically equal to the common shock rate \(E(Z)\), which itself cannot exceed the corresponding marginal Poisson rates. Specifically, the correlation of the pair \((X_j, X_k)\), with marginal rates \(\lambda_j\) and \(\lambda_k\), respectively, is restricted to fall within the interval \([0, \min(\lambda_j, \lambda_k)/\sqrt{\lambda_j\lambda_k}]\) for any distinct \(j, k \in \{1, \ldots, d\}\). Griffiths et al. (1979) show, however, that for an arbitrary pair of Poisson random variables with respective marginal means \(\lambda_j\) and \(\lambda_k\), the correlation can fall anywhere in the interval \([\rho_{\text{min}}(\lambda_j, \lambda_k), \rho_{\text{max}}(\lambda_j, \lambda_k)]\) with

\[
\rho_{\text{min}}(\lambda_j, \lambda_k) = \frac{1}{\sqrt{\lambda_j\lambda_k}} \left[ -\lambda_j\lambda_k - \sum_{m \in \mathbb{N}} \sum_{n \in \mathbb{N}} \min\{0, G_{\lambda_j}(m) + G_{\lambda_k}(n) - 1\} \right],
\]

\[
\rho_{\text{max}}(\lambda_j, \lambda_k) = \frac{1}{\sqrt{\lambda_j\lambda_k}} \left[ -\lambda_j\lambda_k + \sum_{m \in \mathbb{N}} \sum_{n \in \mathbb{N}} \min\{\tilde{G}_{\lambda_j}(m), \tilde{G}_{\lambda_k}(n)\} \right],
\]

where \(G_\lambda\) denotes the cumulative distribution function of a Poisson random variable with rate \(\lambda\), \(\tilde{G}_\lambda\) is the corresponding survival function and \(\mathbb{N} = \{0, 1, \ldots\}\).
It can be checked easily that the implied pairwise correlations in the classical multivariate Poisson model based on a common shock will fall short of \( \rho_{\text{max}} \), except in the case where the marginal rates are identical. Moreover, the classical model cannot account for negative dependence. Further note that while the use of multiple shocks in the more general model of Mahamunulu (1967) alleviates the restriction that all pairwise covariances coincide, the implied pairwise correlations are still restricted to the interval \( [0, \min(\lambda_j, \lambda_k) / \sqrt{\lambda_j \lambda_k}] \).

### 1.3 Alternative Solutions

There are many alternative methods that can be used to establish a multivariate Poisson model. For example, copulas provide a flexible framework for constructing multivariate distributions with arbitrary margins. As such, a multivariate Poisson model can be defined by combining Poisson margins with a particular copula family. This approach was considered by several authors, such as Van Ophem (1999), Pfeifer & Nešlehová (2004), Nikoloulopoulos & Karlis (2009), Smith & Khaled (2012), Panagiotelis et al. (2012) and Inouye et al. (2017).

As another option, a multivariate Poisson distribution can be derived in terms of mixture models. Examples are given by Karlis & Meligkotsidou (2007), Sarabia & Gómez-Déniz (2011), Gómez-Déniz et al. (2012) and Sellers et al. (2016), to name a few. For an extensive review of these alternative multivariate Poisson model formulations, refer to, for example, Inouye et al. (2017) and Schulz (2018). Despite the flexibility that ensues from such model constructions, the resulting distributions are often very complex and difficult to interpret.

### 1.4 Purpose of the Present Paper

The goal of this work is to introduce a multivariate Poisson model allowing for a flexible dependence structure, as is the case for copula and mixture models, while retaining the simplicity and interpretability that ensues from the multivariate reduction technique. To this end, we propose a construction analogous to the latter, but which relies on the notion of a comonotonic shock vector rather than independent common shocks.

By definition, a random vector \((Z_1, \ldots, Z_d)\) is comonotonic if and only if each component can be represented in terms of a common underlying uniform random variable on the unit interval. That is, for a uniform random variable \(U \sim \mathcal{U}(0,1)\), the comonotonic vector can be written as

\[
(Z_1, \ldots, Z_d) = (F_1^{-1}(U), \ldots, F_d^{-1}(U)),
\]

where for each \(j \in \{1, \ldots, d\}\), \(F_j\) denotes the distribution function of \(Z_j\). It can be shown easily that a comonotonic vector follows the upper Fréchet–Hoeffding boundary distribution, that is, for all \(z_1, \ldots, z_d \in \mathbb{R}\),

\[
\Pr(Z_1 \leq z_1, \ldots, Z_d \leq z_d) = \min \{F_1(z_1), \ldots, F_d(z_d)\}.
\]

Analogously to the method of Mahamunulu (1967), correlated Poisson random variables can be generated by convoluting a random vector \((Y_1, \ldots, Y_d)\) of mutually independent Poisson components with varying overlaps of a set of comonotonic shocks. In the trivariate case, for example, this construction can be expressed as

\[
X_1 = Y_1 + Z_{12} + Z_{13} + Z_1, \quad X_2 = Y_2 + Z_{21} + Z_{23} + Z_2, \quad X_3 = Y_3 + Z_{31} + Z_{32} + Z_3,
\]

where the vectors \((Z_{12}, Z_{21}), (Z_{13}, Z_{31}), (Z_{23}, Z_{32})\) and \((Z_1, Z_2, Z_3)\) are mutually independent and each of them is comonotonic.
More generally, for given $k \in \{2, \ldots, d\}$, there are $\binom{d}{k}$ possible $k$-variate comonotonic shocks. By letting $k$ vary from 2 to $d$, one then has a collection of $2^d - d - 1$ comonotonic shock vectors of varying lengths. The use of comonotonic rather than common shocks as building blocks in the model has the advantage that the pairwise correlations can attain the upper bound given by Griffiths et al. (1979). Indeed, it can be shown that a pair of comonotonic Poisson random variables will have correlation $\rho_{\text{max}}$ as given in (2).

Although the proposed multiple comonotonic shock construction allows for considerable flexibility in the inherent dependence structure, its complexity is rather daunting, as was the case in the common shock setting. In this paper, therefore, we limit ourselves to a thorough examination of the simpler model based on a single comonotonic shock vector. Specifically, let $(Y_1, \ldots, Y_d)$ be a vector of mutually independent Poisson random variables, which is independent of the vector $(Z_1, \ldots, Z_d)$ of comonotonic random variables with Poisson margins. A vector $(X_1, \ldots, X_d)$ of correlated Poisson random variables can then be generated by setting

$$X_1 = Y_1 + Z_1, \ldots, X_d = Y_d + Z_d.$$  

(3)

We will refer to this model as the comonotonic shock multivariate Poisson model.

### 1.5 Plan of the Paper

The purpose of this paper is to explore the properties of the comonotonic shock multivariate Poisson model (3), which generalises to arbitrary dimension the construction first considered for $d = 2$ by Genest et al. (2018). The basic properties of the model, which are derived in Section 2, are relatively straightforward extensions of the results therein. However, estimation in higher dimension is considerably more complex than in the bivariate case. This is the topic of Section 3, which considers various estimation strategies and compares their performance through several simulations reported in Section 4. The merits of the proposed model are highlighted through a data illustration presented in Section 5. Finally, concluding remarks and future work are outlined in Section 6.

### 2 The Proposed Model

The multivariate Poisson model described in the following definition is based on the notion of a comonotonic Poisson random vector. Recall that a vector $(Z_1, \ldots, Z_d)$ is said to be comonotonic if it can be written in the form $(Z_1, \ldots, Z_d) = (F_1^{-1}(U), \ldots, F_d^{-1}(U))$, where $U \sim U(0, 1)$ is a standard uniform random variable, and for all $j \in \{1, \ldots, d\}$ and $u \in (0, 1)$,

$$F_j^{-1}(u) = \inf\{t \in \mathbb{R} : F_j(t) \geq u\},$$

where $F_j$ is the distribution function of $Z_j$. In what follows, $\mathcal{P}(\lambda)$ denotes a univariate Poisson distribution with rate $\lambda$.

**Definition 1.** A random vector $(X_1, \ldots, X_d)$ is said to have a multivariate comonotonic shock Poisson distribution with parameters $\Lambda = (\lambda_1, \ldots, \lambda_d) \in (0, \infty)^d$ and $\theta \in [0, 1]$, denoted $(X_1, \ldots, X_d) \sim \mathcal{MP}_d(\Lambda, \theta)$, if and only if it can be expressed in the form (3), namely

$$(X_1, \ldots, X_d) = (Y_1, \ldots, Y_d) + (Z_1, \ldots, Z_d),$$

in terms of two independent random vectors $(Y_1, \ldots, Y_d)$ and $(Z_1, \ldots, Z_d)$ such that

(i) $Y_1, \ldots, Y_d$ are mutually independent and, for all $j \in \{1, \ldots, d\}$, $Y_j \sim \mathcal{P}(1 - \theta)\lambda_j$;

(ii) $Z_1, \ldots, Z_d$ are comonotonic and, for all $j \in \{1, \ldots, d\}$, $Z_j \sim \mathcal{P}(\theta \lambda_j)$.

© 2020 The Authors. International Statistical Review published by John Wiley & Sons Ltd on behalf of International Statistical Institute.
In this representation, the components of the parameter vector \( \Lambda = (\lambda_1, \ldots, \lambda_d) \in (0, \infty)^d \) are the marginal Poisson means, and the parameter \( \theta \in [0, 1] \) represents the portion of the overall mean which is due to the comonotonic shock, that is,

\[
\theta = \frac{E(Z_1 + \ldots + Z_d)}{E(X_1 + \ldots + X_d)}. \tag{4}
\]

It is also easily checked that \( \theta = E(Z_j)/E(X_j) \) for all \( j \in \{1, \ldots, d\} \).

2.1 Connections with Previous Proposals

When \( \theta = 0 \) in the \( MP_d(\Lambda, \theta) \) model, the margins are independent while \( \theta = 1 \) results in perfectly positively dependent components, that is, the variables \( X_1, \ldots, X_d \) are comonotonic. When the marginal Poisson rates are equal, that is, \( \lambda_1 = \ldots = \lambda_d = \lambda \), the comonotonic shock given in Definition 1 reduces to the common shock defined in (1), given that \( Z_1 = \ldots = Z_d = Z \) with \( Z \sim P(\theta \lambda) \). In dimension \( d = 2 \), the \( MP_d(\Lambda, \theta) \) model amounts to the bivariate comonotonic shock Poisson model of Genest et al. (2018), denoted by \( BP(\Lambda, \theta) \). Moreover, it is clear from the formulation given in (3) that any subset \( (X_j, X_k) \) of \( (X_1, \ldots, X_d) \sim MP_d(\Lambda, \theta) \) follows the \( BP(\lambda_j, \lambda_k, \theta) \) distribution.

2.2 Interpretation of \( \theta \) as a Dependence Parameter

The following lemma, which is proved in Appendix A, states that the \( MP_d(\Lambda, \theta) \) family of distributions is ordered by \( \theta \) in the positive quadrant dependence (PQD) ordering introduced by Joe (1990). In what follows, \( F_{\Lambda, \theta} \) denotes the cumulative distribution function of a random vector \( X = (X_1, \ldots, X_d) \sim MP_d(\Lambda, \theta) \), and \( \tilde{F}_{\Lambda, \theta} \) denotes the corresponding survival function.

**Proposition 1.** Suppose \( X = (X_1, \ldots, X_d) \sim MP_d(\Lambda, \theta) \) and \( X' = (X'_1, \ldots, X'_d) \sim MP_d(\Lambda, \theta') \). If \( \theta < \theta' \), then \( X \preceq_{PQD} X' \), i.e., one has, for all \( x_1, \ldots, x_d \in \mathbb{R} \),

\[
F_{\Lambda, \theta}(x_1, \ldots, x_d) \leq F_{\Lambda, \theta'}(x_1, \ldots, x_d) \quad \text{and} \quad \tilde{F}_{\Lambda, \theta}(x_1, \ldots, x_d) \leq \tilde{F}_{\Lambda, \theta'}(x_1, \ldots, x_d).
\]

In view of this result, the parameter \( \theta \) regulates the strength of the dependence in the model. Indeed, it was shown by Joe (1990) that multivariate versions of standard measures of dependence such as Kendall's tau or Spearman's rho are monotone with respect to the PQD order. Furthermore, Theorem 9.A.5 of Shaked & Shanthikumar (2007) states that the PQD order is closed under marginalisation. Therefore, if \( X \sim MP_d(\Lambda, \theta) \) and \( X' \sim MP_d(\Lambda, \theta') \), where \( \theta, \theta' \in (0, 1) \) with \( \theta < \theta' \), then for any distinct \( j, k \in \{1, \ldots, d\} \), one has \( (X_j, X_k) \preceq_{PQD} (X'_j, X'_k) \) and thus

\[
\text{cov}(X_j, X_k) \leq \text{cov}(X'_j, X'_k) \quad \text{and} \quad \text{corr}(X_j, X_k) \leq \text{corr}(X'_j, X'_k).
\]

Accordingly, for fixed marginal parameters \( \Lambda \in (0, \infty)^d \), the pairwise correlation between any pair \( (X_j, X_k) \) from \( X \sim MP_d(\Lambda, \theta) \), denoted as \( \rho_{X_j, X_k}(\theta) \), is an increasing function of \( \theta \). This result also follows immediately from the fact that \( (X_j, X_k) \sim BP(\Lambda, \theta) \) and \( (X'_j, X'_k) \sim BP(\Lambda, \theta') \), as the PQD ordering in the \( BP \) family was established by Genest et al. (2018).
2.3 Correlation

It was already shown by Genest et al. (2018) that if \((X_j, X_k) \sim \mathcal{BP}(\lambda_j, \lambda_k, \theta)\), then

\[
\text{cov}(X_j, X_k) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \min\{\tilde{G}_{\theta\lambda_j}(m), \tilde{G}_{\theta\lambda_k}(n)\} - \theta^2\lambda_j\lambda_k = m_{\lambda_j,\lambda_k}(\theta),
\]

and hence

\[
\text{corr}(X_j, X_k) = \frac{1}{\sqrt{\lambda_j\lambda_k}} \left[ \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \min\{\tilde{G}_{\theta\lambda_j}(m), \tilde{G}_{\theta\lambda_k}(n)\} - \theta^2\lambda_j\lambda_k \right] \equiv \rho_{\lambda_j,\lambda_k}(\theta).
\]

For \(X \sim \mathcal{MP}(\Lambda, \theta)\), it is clear that \(\theta = 0\) results in \(\rho_{\lambda_j,\lambda_k}(\theta) = 0\) for each pair \((X_j, X_k)\) from \(X\) with distinct \(j, k \in \{1, \ldots, d\}\). At the other end of the spectrum, it is easily verified that \(\theta = 1\) yields \(\rho_{\lambda_j,\lambda_k}(\theta) = \rho_{\text{max}}\) for all distinct \(j, k \in \{1, \ldots, d\}\), that is, all pairwise correlations reach the upper bound given in (2). In this case, \(X\) is a comonotonic random vector, and its distribution reaches the upper Fréchet–Hoeffding bound. Thus, the proposed multivariate comonotonic shock Poisson model yields pairwise correlations spanning the full range \([0, \rho_{\text{max}}]\), as regulated by the dependence parameter \(\theta \in [0, 1]\).

By construction, the proposed multivariate Poisson comonotonic shock model allows for greater flexibility in the covariance structure than its classical analogue. Indeed, the proposed model results in pairwise covariances given by \(m_{\lambda_j,\lambda_k}(\theta)\), for any distinct \(j, k \in \{1, \ldots, d\}\). Thus, although there is a single dependence parameter \(\theta\) that determines the strength of the dependence, the magnitude of its effect is regulated by the marginal means \(\lambda_j\) and \(\lambda_k\). In contrast, the classical model restricts all pairwise covariances to be identical.

2.4 Multivariate Distribution and Probability Mass Function

As detailed in Schulz (2018), the comonotonic random vector \((Z_1, \ldots, Z_d)\) introduced in Definition 1 has probability mass function given, for all \(z_1, \ldots, z_d \in \mathbb{N}\), by

\[
c_{\Lambda,\theta}(z_1, \ldots, z_d) = \left[ \min_{j \in \{1, \ldots, d\}} G_{\theta\lambda_j}(z_j) - \max_{j \in \{1, \ldots, d\}} G_{\theta\lambda_j}(z_j - 1) \right]_+,
\]

where \([x]_+ = x 1(x > 0)\). Using this fact, one can show readily that the cumulative distribution function of the \(\mathcal{MP}(\Lambda, \theta)\) model is given, for all \(x_1, \ldots, x_d \in \mathbb{N}\), by

\[
F_{\Lambda,\theta}(x_1, \ldots, x_d) = \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) \min\{G_{\theta\lambda_1}(z_1), \ldots, G_{\theta\lambda_d}(z_d)\}
\]

\[
= \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \prod_{j=1}^{d} G_{(1-\theta)\lambda_j}(x_j - z_j) c_{\Lambda,\theta}(z_1, \ldots, z_d),
\]

where \(g_{\lambda}(x) = e^{-\lambda} \lambda^x / x!\). The corresponding probability mass function is given, for all \(x_1, \ldots, x_d \in \mathbb{N}\), by

\[
f_{\Lambda,\theta}(x_1, \ldots, x_d) = \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \prod_{j=1}^{d} g_{(1-\theta)\lambda_j}(x_j - z_j) c_{\Lambda,\theta}(z_1, \ldots, z_d).
\]
Figure 1. Probability mass function for the classical model (left), the proposed model (centre) and the difference between the two (right) for given values of $\xi$ and $\theta$, respectively, with marginal parameters fixed at $(\lambda_1, \lambda_2) = (1, 2)$. 
This result will be used in the estimation procedures detailed in Section 3.

In comparison, the classical multivariate Poisson model in (1) generates correlated margins by setting \((X_1, \ldots, X_d) = (Y_1 + Z, \ldots, Y_d + Z)\) for mutually independent Poisson random variables \(Y_1, \ldots, Y_d, Z\) with respective marginal rates \(\lambda_1 - \xi, \ldots, \lambda_d - \xi, \xi\). In this formulation, probabilities can be computed, for any \(x_1, \ldots, x_d \in \mathbb{N}\), according to the formula

\[
h_{\Lambda, \xi}(x_1, \ldots, x_d) = \sum_{z=0}^{\min(x_1, \ldots, x_d)} \prod_{j=1}^{d} g_{\lambda_j - \xi}(x_j - z)g_\xi(z). \tag{6}
\]

As previously noted, when the marginal rates coincide, the comonotonic multivariate Poisson model reduces to the classical common shock model and the probability mass functions (5) and (6) will coincide. When the degree of dependence is low (i.e., \(\theta\) and \(\xi\) are near zero), there is very little difference between the classical and comonotonic models given that in both constructions, the vector \((X_1, \ldots, X_d)\) is predominantly generated from the mutually independent components \(Y_1, \ldots, Y_d\). For strong degrees of dependence, however, the difference between the two formulations becomes more pronounced. When \(\theta\) is large, in particular such that \(m_{\lambda_j, \lambda_k}(\theta) > \min(\lambda_1, \ldots, \lambda_d)\) for any distinct \(j, k \in \{1, \ldots, d\}\), the classical multivariate Poisson model is in fact completely inappropriate as the covariance \(\xi\) must necessarily be smaller than the marginal parameters.

In order to readily contrast the classical common shock construction of (1) with the proposed comonotonic shock approach of (3), consider the bivariate setting. Comparisons can be made between the \(BP(\Lambda, \theta)\) model and an analogue form in the classical setup by choosing the same marginal rates \(\Lambda = (\lambda_1, \lambda_2)\) and covariance parameter \(\xi = m_{\lambda_1, \lambda_2}(\theta)\). Figure 1 illustrates the difference between the probability mass functions (5) and (6) with \((\lambda_1, \lambda_2)\) fixed at \((1, 2)\) and \(\theta \in \{0.10, 0.75\}\), or, correspondingly, \(\xi \in \{0.11, 0.97\}\). Note that for any \(\theta > 0.77\) in the \(BP\) model with \(\Lambda = (1, 2)\), there is no equivalent form in the classical model as this results in a covariance \(m_{1,2}(\theta) > 1\). As evidenced by the plots, the difference between the two models becomes more obvious as the degree of dependence increases.

3 Estimation

Let \(X_1, \ldots, X_n\) be a random sample from the proposed \(MP_d(\Lambda, \theta)\) model. There are several estimation strategies that will yield consistent estimators of the model parameters \(\lambda_1, \ldots, \lambda_d\) and \(\theta\), provided that they are all strictly positive and that \(\theta \in (0, 1)\). In particular, this section will detail three techniques: the method of moments (MM), maximum likelihood (ML) estimation (MLE), and the inference function for margins (IFM) approach.

3.1 Method of Moments

The MM approach establishes parameter estimates by matching theoretical moments to sample moments. In the \(MP_d\) model, the marginal parameters are simply the marginal means, and thus, the MM estimator \(\hat{\Lambda} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_d)\) is given by the vector \((\hat{X}_1, \ldots, \hat{X}_d)\) of sample means. The central limit theorem ensures that for each \(j \in \{1, \ldots, d\}\), the estimator \(\hat{\lambda}_j\) is consistent and asymptotically Gaussian, that is, as \(n \to \infty\),

\[
\sqrt{n}(\hat{\lambda}_j - \lambda_j) \sim \mathcal{N}(0, \lambda_j),
\]
where \( \rightsquigarrow \) denotes convergence in distribution. Note that the MM estimators are also the ML estimators in the independence model where \( \theta = 0 \).

Estimation of the dependence parameter \( \theta \) can be based on mixed moments. Suppose \( \mathbf{X} \sim MP_d(\Lambda, \theta) \), then for any distinct \( j, k \in \{1, \ldots, d\} \), the pair \( (X_j, X_k) \) has covariance \( m_{\lambda_j, \lambda_k}(\theta) \). The PQD ordering in the proposed \( MP_d \) family ensures that \( m_{\lambda_j, \lambda_k}(\theta) \) is a strictly increasing function of \( \theta \). For arbitrary \( j, k \in \{1, \ldots, d\} \), let \( S_{jk} \) denote the sample covariance of \( (X_j, X_k) \), namely

\[
S_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k).
\]

For any \( j, k \in \{1, \ldots, d\} \) with \( j \neq k \), one has \( S_{jk} \in (0, m_{\bar{X}_j, \bar{X}_k}(1)) \) almost surely as \( n \to \infty \). The MM estimate \( \tilde{\theta}_{jk} \) is then the unique root of the equation \( S_{jk} = m_{\bar{X}_j, \bar{X}_k}(\theta) \). In small samples, \( \tilde{\theta}_{jk} \) may occasionally fall beyond the permissible bounds; in such instances, one could set \( \tilde{\theta}_{jk} = 0 \) or 1 as the case may be. However, if \( S_{jk} \) is significantly smaller than 0, one should consider an alternative model specification to account for this negative dependence.

As stated in, for example, Theorem 8 of Ferguson (1996), the sample covariance is asymptotically Gaussian with asymptotic variance

\[
\sigma^2(\theta, \lambda_j, \lambda_k) = \text{var}\{ (X_j - \lambda_j)(X_k - \lambda_k) \}.
\]

Let \( \gamma_{jk} \) denote the mapping \( \gamma_{jk} : x \mapsto m_{\lambda_j, \lambda_k}^{-1}(x) \) such that \( \tilde{\theta}_{jk} = \gamma_{jk}(S_{jk}) \). Then, an application of the delta method leads to

\[
\sqrt{n}(\tilde{\theta}_{jk} - \theta) \rightsquigarrow \mathcal{N}\left[0, (\gamma'_{jk}(\sigma_{jk}))^2 \sigma^2(\theta, \lambda_j, \lambda_k)\right],
\]

where \( \sigma_{jk} = m_{\lambda_j, \lambda_k}(\theta) \) and \( \gamma'_{jk} \) denotes the first derivative of \( \gamma_{jk} \).

It is clear that in dimensions \( d > 2 \), there are in fact \( d(d-1)/2 \) distinct possible estimators of the form \( \tilde{\theta}_{jk} \) for all distinct \( j, k \in \{1, \ldots, d\} \). Each of these estimators is consistent and asymptotically Gaussian. A moment-based estimator \( \tilde{\theta} \) can then be defined as the average of these estimators, namely

\[
\tilde{\theta} = \frac{2}{d(d-1)} \sum_{j=1}^{d-1} \sum_{k=j+1}^{d} \tilde{\theta}_{jk}.
\]

As the dimension increases, the aforementioned estimation procedure may become rather long to carry out. As an alternative, one could consider averaging in (7) over a large subset of estimates \( \tilde{\theta}_{jk} \).

The following theorem establishes the asymptotic behaviour of the moment-based estimator \( \tilde{\theta} \). Its proof is given in Appendix B.

**Proposition 2.** Let \( X_1, \ldots, X_n \) be a random sample with distribution \( MP_d(\Lambda, \theta) \). For each choice of \( j, k \in \{1, \ldots, d\} \) with \( j < k \), let \( \tilde{\theta}_{jk} \) denote the unique solution to the equation \( m_{\bar{X}_j, \bar{X}_k}(\theta) = S_{jk} \). Let \( \tilde{\theta} \) be the average of the set of estimators \( \tilde{\theta}_{jk} \) with \( j, k \in \{1, \ldots, d\} \) and \( j < k \). Then, as \( n \to \infty \),

\[
\sqrt{n}(\tilde{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, \hat{V}),
\]

\( \hat{V} \) is the covariance matrix of \( \tilde{\theta} \).
where the asymptotic variance is given by

\[
\tilde{\nu} = \frac{2}{d(d-1)} \{ \Gamma'(\Sigma^*) \} V^* \{ \Gamma'(\Sigma^*) \}^T
\]

\[
= \frac{2}{d(d-1)} \sum_{j=1}^{d-1} \sum_{k=j+1}^{d} \gamma'_{jk}(\sigma_{jk}) \gamma'_{rs}(\sigma_{rs}) \times \text{cov} \{ (X_{ij} - \lambda_j)(X_{ik} - \lambda_k), (X_{ir} - \lambda_r)(X_{is} - \lambda_s) \}.
\]

In the aforementioned expression, \( \Gamma'(\Sigma^*) \) denotes the vectorisation of the inverse covariance function \( \gamma_{jk}(\sigma_{jk}) \) with \( d^2 = d(d-1)/2 \) elements and \( \Gamma' \) denotes the corresponding component-wise derivative, that is,

\[
\Gamma'(\Sigma^*) = (\gamma'_{12}(\sigma_{12}), \gamma'_{13}(\sigma_{13}), \ldots, \gamma'_{1d}(\sigma_{1d}), \gamma'_{23}(\sigma_{23}), \ldots, \gamma'_{d-1d}(\sigma_{d-1d})).
\]

In general, the asymptotic variance of the moment-based estimator \( \tilde{\theta} \) does not result in a compact expression. Nevertheless, simplifications occur in some special cases. For each \( \ell \in \{1, \ldots, d\} \), write \( X_\ell - \lambda_\ell = \hat{Y}_\ell + \hat{Z}_\ell \), where \( \hat{Y}_\ell \) and \( \hat{Z}_\ell \) represent centred variables, respectively, given by \( \hat{Y}_\ell = Y_\ell - (1 - \theta)\lambda_\ell \), and \( \hat{Z}_\ell = Z_\ell - \theta\lambda_\ell \). One then has

\[
\text{cov} \{ (X_j - \lambda_j)(X_k - \lambda_k), (X_r - \lambda_r)(X_s - \lambda_s) \} = \text{cov}(\hat{Y}_j\hat{Y}_k + \hat{Y}_j\hat{Z}_k + \hat{Z}_j\hat{Y}_k + \hat{Z}_j\hat{Z}_k, \hat{Y}_r\hat{Y}_s + \hat{Y}_r\hat{Z}_s + \hat{Z}_r\hat{Y}_s + \hat{Z}_r\hat{Z}_s).
\]

This expression can then be further decomposed by considering the overlaps between the indices \( j, k, r, s \). Additional simplifications ensue when \( \lambda_1 = \ldots = \lambda_d = \lambda \). For example, when \( j < k, r < s \) and the four indices are distinct, it can be shown that

\[
\text{cov} \{ (X_j - \lambda_j)(X_k - \lambda_k), (X_r - \lambda_r)(X_s - \lambda_s) \} = \text{cov}(\hat{Z}_j\hat{Z}_k, \hat{Z}_r\hat{Z}_s) = \theta\lambda(1 + 2\theta\lambda).
\]

### 3.2 Maximum Likelihood Estimation

Likelihood-based estimation is also possible in the proposed \( \mathcal{MP}_d \) family. Based on the formulation of the joint probability mass function given in (5), the log-likelihood can be expressed as

\[
\ell(\Lambda, \theta) = \sum_{i=1}^{n} \ln \left\{ \prod_{z_{1i}=0}^{x_{1i}} \ldots \prod_{z_{di}=0}^{x_{di}} \prod_{j=1}^{d} g(1-\theta)\lambda_j (x_{ij} - z_{ij}) c_{\Lambda,\theta}(\bar{z}_1, \ldots, \bar{z}_d) \right\}.
\]

Maximum likelihood estimates \( \hat{\Psi} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_d, \hat{\theta}) \) are then determined as \( \arg \max_{\lambda_1, \ldots, \lambda_d, \theta} \ell(\Lambda, \theta) \), or, equivalently, by solving the system

\[
\frac{\partial}{\partial \lambda_1} \ell(\Lambda, \theta) = \ldots = \frac{\partial}{\partial \lambda_d} \ell(\Lambda, \theta) = 0, \quad \frac{\partial}{\partial \theta} \ell(\Lambda, \theta) = 0,
\]

subject to the constraints \( \lambda_1, \ldots, \lambda_d \in (0, \infty) \) and \( \theta \in (0, 1) \). While no analytic solution exists to the score equations outlined in (8), this system can be solved numerically. Note that the optimisation procedure can be simplified by considering a reparametrisation wherein parameter constraints are removed. In the \( \mathcal{MP}_d(\Lambda, \theta) \) model, this is achieved by setting \( \xi_1 = \ln(\lambda_1), \ldots, \xi_d = \ln(\lambda_d), \eta = \ln\{\theta/(1-\theta)\} \) and optimising the reparametrised log-likelihood.
The ML estimators are then given by $(\hat{\lambda}_1, \ldots, \hat{\lambda}_d) = (e^{\hat{\lambda}_1}, \ldots, e^{\hat{\lambda}_d})$ and $\hat{\theta} = e^{\hat{\theta}}/(1 + e^{\hat{\theta}})$.

It can be seen that these estimators are asymptotically Gaussian, specifically, as $n \to \infty$, 
\[ \sqrt{n}(\hat{\Psi} - \Psi) \sim N(0, I^{-1}) , \]
where $I$ denotes the Fisher information matrix. In practice, the asymptotic variance can be estimated via bootstrap.

### 3.3 Inference Function for Margins

Although ML estimation is valid, its actual implementation may become numerically infeasible in high dimensions. The IFM approach is an appealing alternative likelihood-based estimation technique wherein marginal parameters are estimated by their respective marginal ML estimators, say $\hat{\lambda}_1, \ldots, \hat{\lambda}_d$, and the dependence parameter estimator is determined by $\hat{\theta} = \arg \max_\theta \ell(\hat{\lambda}_1, \ldots, \hat{\lambda}_d, \theta)$. This method is often used, e.g., in inference for copula models, where there is a natural separation between the marginal and dependence parameters; refer to Joe (2005, 2014).

In adapting the IFM approach to the $\mathcal{M}P_d(\Lambda, \theta)$ model, first consider rewriting the probability mass function given in (5) as
\[
f_{\Lambda, \theta}(x_1, \ldots, x_d) = \left\{ \prod_{j=1}^d g_{\lambda_j}(x_j) \right\} \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \omega_{\Lambda, \theta}(z; x) c_{\Lambda, \theta}(z_1, \ldots, z_d), \tag{9} \]
where $\omega_{\Lambda, \theta}(z; x)$ can be regarded as a weight function defined by
\[
\omega_{\Lambda, \theta}(z; x) = \prod_{j=1}^d b_{x_j, \theta}(z_j) / g_{\theta \lambda_j}(z_j) = \prod_{j=1}^d \left( x_j / z_j \right) \theta^{z_j} (1 - \theta)^{x_j - z_j} / g_{\theta \lambda_j}(z_j).
\]

The expression given in (9) leads to an alternative formulation of the log-likelihood, namely
\[ \ell(\Lambda, \theta) = \ell_1(\lambda_1) + \ldots + \ell_d(\lambda_d) + \ell_D(\Lambda, \theta) . \]
Here, $\ell_j(\lambda_j)$ is the marginal log-likelihood for the $j$-th component $X_j \sim \mathcal{P}(\lambda_j)$ and the term $\ell_D(\Lambda, \theta)$ encompasses the dependence between the margins and is given by
\[ \ell_D(\Lambda, \theta) = \ln \left\{ \sum_{z_1=0}^{x_1} \cdots \sum_{z_d=0}^{x_d} \omega_{\Lambda, \theta}(z; x) c_{\Lambda, \theta}(z_1, \ldots, z_d) \right\} . \]

The IFM estimators $\hat{\Psi} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_d, \hat{\theta})$ can be defined as the solution to the set of equations
\[ \mathbf{s}^T = \left( \frac{\partial}{\partial \lambda_1} \ell(\lambda_1), \ldots, \frac{\partial}{\partial \lambda_d} \ell(\lambda_d), \frac{\partial}{\partial \theta} \ell(\Lambda, \theta) \right) = (0, \ldots, 0). \tag{10} \]

The implementation of the IFM method results in a two-stage estimation procedure. In the first step, the marginal Poisson rates are simultaneously and independently estimated by their respective marginal ML estimates, that is, $\hat{\lambda}_j = \hat{X}_j$ for all $j \in \{1, \ldots, d\}$. The second step involves solving
\[ \frac{\partial}{\partial \theta} \ell(\hat{\lambda}, \theta) = \frac{\partial}{\partial \theta} \ell_D(\hat{\lambda}, \theta) = 0 . \]
As outlined in, for example, Joe (2005), the estimators that result from the system given in (10) are consistent and asymptotically Gaussian. In particular, as \( n \to \infty \), \( \sqrt{n}(\Psi - \Psi) \to N(0, \hat{V}) \), where the asymptotic variance is given by

\[
\hat{V} = (-D_\Psi ^{-1})M_\Psi (-D_\Psi ^{-1})^T,
\]

where \( M_\Psi = \text{cov}\{\hat{s}(\mathbf{X}; \Psi)\} \) and \( D_\Psi = \mathbb{E}\left\{\partial \hat{s}(\mathbf{X}; \Psi) / \partial \Psi^T\right\} \).

Because the elements of the estimating equation (10) are composed of marginal and joint log-likelihoods, the asymptotic variance can be further simplified. Set \( q = d + 1 \) and, as previously introduced, let \( \mathcal{I} \) denote the Fisher information matrix based on the full log-likelihood \( \ell(\Lambda, \theta) \). Using the convention that \( M_{jk} \) denotes the \((j, k)\)-th entry of a matrix \( M \), the following simplifications, as shown in Schulz (2018), follow for the asymptotic variance matrix:

\[
\hat{V}_{jj} = \lambda_j, \quad \hat{V}_{jk} = m_{\lambda_j, \lambda_k}(\theta), \quad \hat{V}_{jq} = -\mathcal{I}_{qq}^{-1} \sum_{k=1}^{d} m_{\lambda_j, \lambda_k}(\theta) \mathcal{I}_{kq},
\]

and

\[
\hat{V}_{qq} = \mathcal{I}_{qq}^{-1} + (\mathcal{I}_{qq}^{-1})^2 \sum_{j=1}^{d} \sum_{k=1}^{d} \mathcal{I}_{jq} \mathcal{I}_{kq} m_{\lambda_j, \lambda_k}(\theta).
\]

### 3.4 Comparison Between the Three Methods

Both ML estimation and the IFM method derive estimators according to a set of estimating equations, as given in (8) and (10), respectively. In comparing these two sets of estimating equations, it is clear that the difference between these two likelihood-based approaches is that the IFM method systematically ignores the information pertaining to the marginal parameters that is contained in the dependence portion of the log-likelihood \( \ell_D(\Lambda, \theta) \). As a result, there is a loss in efficiency in the IFM framework. However, when \( \theta = 0 \), independence ensues and thus \( \ell_D(\Lambda, 0) = 0 \) so that the IFM estimation procedure coincides with a full ML approach. In the case of independence, as previously noted, the MM will also yield identical estimators as the ML approach.

### 3.5 Including Covariate Effects

In applications, it is often of interest to fit models that take into account covariate effects. In the univariate setting, this could be accomplished via generalised linear models (GLMs) wherein a transformation of the Poisson mean \( \lambda \) is expressed in terms of a linear predictor. This is typically performed by considering the log link function and setting \( \ln(\lambda) = \mathbf{T}\beta \), where \( \mathbf{T} \) is a \( 1 \times p \) vector of covariates and \( \beta \) is a \( p \times 1 \) vector of regression parameters.

Model (3) can be adapted to allow for the inclusion of explanatory variables. To this end, the parameters can be re-expressed in terms of covariates \( \mathbf{T} \), namely

\[
\lambda_1 = g_1^{-1}(\mathbf{T}\beta_1), \ldots, \lambda_d = g_d^{-1}(\mathbf{T}\beta_d), \quad \theta = g_{d+1}^{-1}(\mathbf{T}\beta_{d+1}),
\]

where for \( k \in \{1, \ldots, d + 1\} \), \( g_k^{-1} \) denotes the inverse link function for the \( k \)-th component. For univariate Poisson margins, an appropriate choice of link function is the natural logarithm. Hence, the exponential function is a suitable choice for \( g_1^{-1}, \ldots, g_d^{-1} \).

While it is natural to express the marginal means in terms of linear predictors, it is less clear whether the dependence parameter \( \hat{\theta} \) should depend on covariates and, moreover, what...
functional form the relation should assume, that is, the form of the function $g_{d+1}$. As previously stated in (4), the dependence parameter can be viewed as the portion of the overall mean that is due to the comonotonic shock. Accordingly, the logit link function could be a suitable choice, wherein $\ln \frac{\text{Pr}(Y=1)}{1-\text{Pr}(Y=1)} = T\alpha$. By design, the constraints $\lambda_1, \ldots, \lambda_d \in (0, \infty)$ and $\theta \in (0, 1)$ will be satisfied. In a full ML approach, estimators are then given by

$$\hat{\beta}_j = \arg\max_{\beta_j} \ell_j(\lambda_j(\beta_j))$$

for all $j \in \{1, \ldots, d\}$ and

$$\hat{\alpha} = \arg\max_{\alpha} \ell_1(\lambda_1(\hat{\beta}_1), \ldots, \lambda_d(\hat{\beta}_d), \theta(\alpha)).$$

4 Simulations

Simulations were conducted in order to assess the estimation methods outlined in Section 3. Several scenarios were considered wherein the values of the parameters, dimension, and sample size were varied. Specifically, 30 distinct scenarios were examined, each of which resulted from the combination of dimension $d \in \{2, 3, 4\}$, dependence parameter $\theta \in \{0.10, 0.25, 0.50, 0.75, 0.90\}$ and sample size $n \in \{50, 500\}$. Throughout, the marginal parameters were held fixed at $\lambda = (1, \ldots, d)$ as the primary interest was to assess the impact of dependence in the model. For each scenario, a random sample of size $n$ was generated from the specified $\mathcal{MP}_d(\Lambda, \theta)$ model according to Algorithm 1.

Algorithm 1 Simulation algorithm.

1. For each $i \in \{1, \ldots, n\}$, generate $(V_{i1}, \ldots, V_{id}, U_i)^{iid} \sim \mathcal{U}(0, 1)$.
2. For each $j \in \{1, \ldots, d\}$, set $X_{ij} = G^{-1}(1-\theta)\lambda_j(V_{ij}) + G^{-1}(\theta)\lambda_j(U_i)$.

In each of the 30 cases, the three estimation methods discussed in Section 3 were implemented and replicated 500 times. Sample code for the data generation process and the implementation of the three estimation procedures is provided in the supporting information.

Recall that under the MM, estimation of the dependence parameter involves solving for the unique root of $S_{jk} = m_{\tilde{x}_j, \tilde{x}_k}(\theta)$ for each $j, k \in \{1, \ldots, d\}$ with $j < k$. This can readily be implemented in R using the `uniroot` function. This, however, is only feasible for values of $S_{jk}$ falling within the permissible range $[0, m_{\tilde{x}_j, \tilde{x}_k}(1)]$ implied by the $\mathcal{MP}_d(\tilde{\Lambda}, \tilde{\theta})$ family. Whenever $S_{jk}$ fell above the threshold $m_{\tilde{x}_j, \tilde{x}_k}(1)$, the convention was to set $\tilde{\theta}_{jk} = 1$. Whenever $S_{jk} < 0$, however, the estimate $\tilde{\theta}_{jk}$ was assigned a value of $NA$ and was excluded from the average (7) yielding the MM estimate $\tilde{\theta}$. In over 90% of iterations, there were no instances of
negative pairwise correlations. Moreover, in only 1.5% of cases were all pairwise correlations
negative, rendering estimation of the dependence parameter impossible. Such issues occurred
mostly when \((d, \theta, n) = (2.50, 0.10)\).

Both likelihood-based methods (i.e., ML and IFM) rely on maximising some form of a
score function and were implemented using the \texttt{optim} function in R, where the MM esti-
mates were used as starting values for the procedure. Note that whenever the MM estimate \(\hat{\theta}\)
was found to be \(\mathcal{N}A\), the starting value was set to \(\theta^{(0)} = 0.01\), while when \(\hat{\theta} = 1\) a starting
value of \(\theta^{(0)} = 0.99\) was used instead. This was done to avoid issues at the boundary of the
parameter space of \(\theta\). In order to simplify the optimisation procedures, the reparametrisation
discussed in Section 3.2 was used. Recall that the latter involves the transformed parameters
\(\xi_1 = \ln(\lambda_1), \ldots, \xi_d = \ln(\lambda_d), \eta = \ln\{\theta/(1 - \theta)\}\). In what follows, the reported results are
transformed back to the original \((\Lambda, \theta)\) scale for convenience.

Figures 2 and 3 display the dependence parameter estimates obtained from the various esti-
mation methods for the specific set-ups \((d, \theta, n)\), where \(\theta = 0.25\) and 0.90, respectively. The
IFM and ML methods have similar performance across all scenarios. The results suggest that
the MM performs better for weaker levels of dependence, for all dimensions and sample sizes.
In particular, the variability in the resulting MM estimates is similar to that under the IFM and
ML methods for lower values of \(\theta\), while for stronger dependence levels, the variability in the
MM estimates is noticeably greater than that for the two likelihood approaches.

The results also show a decrease in variability for the two likelihood-based approaches as
the strength of the dependence increases. This phenomenon is in accordance with intuition:
for larger values of \(\theta\), the random vector \((X_1, \ldots, X_d) \sim \mathcal{MP}_d(\Lambda, \theta)\) will be predominately
generated from the underlying comonotonic shock \((Z_1, \ldots, Z_d)\), as in (3). This induces less
uncertainty as the components of the comonotonic shock vector are generated from a common

![Figure 2](image-url)

**Figure 2.** Estimation results for \(\theta\) from the method of moments (MM) (left), inference function for margins (IFM) method
(middle) and maximum likelihood estimation (MLE) (right) in indicated scenario

© 2020 The Authors. International Statistical Review published by John Wiley & Sons Ltd on behalf of International Statistical Institute.
uniform random variable. Moreover, the estimation results improve as \( n \) increases, as expected from standard theory.

Note that there were some instances where seemingly outlier estimates were obtained under the likelihood-based estimation methods. For example, this can be seen in the results for the scenarios with \( \theta = 0.90 \). These outlier occurrences tended to happen when large starting values were used in the optimisation procedure. As initialising the algorithm at a value \( \theta^{(0)} \) close to the boundary of the parameter space can sometimes lead to poor estimates, caution is advised.

Figure 4 displays the results for the marginal parameter estimates in dimension \( d \) for \( \theta = 0.50 \). Recall that the MM and IFM methods both result in estimators given by the marginal sample mean, that is, \( \hat{\lambda}_j = \bar{X}_j \) for all \( j \in \{1, \ldots, d\} \). Across all scenarios, the MM/IFM and ML results are very similar, suggesting that there is little information about the marginal parameters \( \Lambda \) in the dependence portion of the log-likelihood \( \ell_{DP}(\Lambda, \theta) \). This further reinforces the utility of the IFM approach as a likelihood-based estimation method.

Figure 5 displays the run times in R for the three estimation procedures considered, respectively shown for each \( d \in \{2, 3, 4\} \). The MM is the fastest across all dimensions, followed by the IFM approach and lastly ML estimation. The execution time for the full ML estimation became increasingly burdensome as the dimension grew. The IFM approach, however, remained relatively quick to implement in higher dimensions. Notably, in dimension 4, the average run time of the ML method is more than six times longer than the average IFM execution time and almost 17 times than that of the MM. Conceivably, a full ML approach could eventually become infeasible as the dimension increases. This highlights the appeal of the IFM method as an alternative likelihood-based estimation method in high-dimensional settings.

An additional set of simulations was carried out to investigate the performance of the estimation methods in higher dimensions, specifically the case where \( d = 10 \). As the previous
simulations suggested that both likelihood-based approaches yielded similar results, we focused here on the IFM approach as well as the MM, with similar implementations are previously discussed. Throughout the simulations, the sample size was held fixed at $n = 250$, the marginal parameters were set to $\Lambda = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$ and $\theta$ was varied in $\{0.1, 0.5, 0.9\}$. Note that in order to speed up the simulations, the convergence tolerance levels were lowered in this second set of simulations to $10^{-4}$. Each scenario was replicated 500 times, resulting in a total of 1500 iterations.

The results, which are summarised in Figure 6, are consistent with what was previously observed. In particular, for weaker levels of dependence there is seemingly little difference between the MM and IFM methods, while for larger values of $\theta$ the likelihood-based IFM approach shows considerably less variability. Note that estimation results for the marginal parameters $\Lambda$ are not shown as both the MM and IFM approaches yield identical estimates equal to the marginal sample means.

The implementation times (which are not shown) were substantially longer for IFM estimation in comparison with the MM in dimension 10. Indeed, the IFM estimation procedure took, on average, between 1.4 and 68.2 times longer than that of the MM. Not surprisingly, executing the estimation procedures for large $d$ becomes increasingly burdensome. In particular, the MM involves computing the average of the solutions to $d(d - 1)/2$ equations, while likelihood-based estimation involves evaluating a summation across $d$ variables. The latter stems from the form of the probability mass function, as given in (5) or equivalently in (9).

Based on the simulations considered here, the implementation of the MM seemingly remains feasible even as the dimension becomes quite large. Nonetheless, a viable alternative for large $d$ would be to consider taking the average of only a subset of the $d(d - 1)/2$ estimators $\hat{\theta}_{jk}$ with $j, k \in \{1, \ldots, d\}$ and $j < k$. As each $\hat{\theta}_{jk}$ is consistent and asymptotically Gaussian, an
average of any number of these estimators will also be consistent and asymptotically Gaussian. Note that a practical advantage of the MM estimation approach in higher dimensions is that in averaging over \(d(d-1)/2\) estimators, there are less problematic occurrences wherein each \(S_{jk} < 0\), rendering estimation of \(\theta\) impossible. Indeed, in this second set of simulations, there were no instances wherein all pairwise covariances were negative leading to \(\hat{\theta} = NA\), as had been previously observed in the first set of simulations.

The implementation of the IFM method is more challenging for larger \(d\). Note that choosing lower values for the marginal rates, as was performed here with \(\Lambda = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)\), rendered the computation of the IFM estimates...
more feasible. Indeed, larger values of $\Lambda$ quickly increase the number of summands involved in the dependence portion of the log-likelihood $\ell_D(\Lambda, \theta)$ thereby increasing the computational difficulty. Despite the choice for $\Lambda$ used in this second set of simulations, there were some instances (specifically, 18 out of the 1500 iterations) in which the IFM estimation method failed because of size issues. Recall that the implementation of the IFM approach consists of a two-step procedure. As a first step, the marginal parameter estimates are obtained as the marginal ML estimates $\hat{\lambda}_1 = \bar{X}_1, \ldots, \hat{\lambda}_d = \bar{X}_d$. At the second step, the dependence parameter is estimated according to the score equation $\partial \ell_D(\Lambda, \theta)/\partial \theta = 0$. Notice that the optimisation problem for IFM estimation consists only of univariate score equations. Nonetheless, the implementation of the IFM approach proved to be difficult in higher dimensions. Further explorations of more efficient implementation strategies are necessary for large $d$.

5 Data Illustration

To illustrate the usefulness of the proposed model, an application of the $\mathcal{MP}_d$ class was carried out using open data provided by BIXI, Montréal’s bicycle sharing system (https://montreal.bixi.com/en/open-hyphen-data). The raw data detail each bike rental occurrence and provide information on the exact date and time of departure and arrival, along with the corresponding departure and arrival stations, the total bike usage time as well as a binary variable indicating whether the user was a BIXI member.

We considered data from the 2017 season, extending from 15 April to 15 November, inclusively, and focused on bike usage information collected at three nearby stations located on Montréal’s south shore town of Longueuil. In particular, we applied the trivariate model to the number of daily bike rentals taken from three specific departure stations, located at the intersection of St Charles and Charlotte streets (station 1), the intersection of St Charles and St Sylvestre streets (station 2) and at Collège Édouard-Montpetit (Station 3), denoted by the random vector $(X_1, X_2, X_3)$. We used a subset of the data consisting only of BIXI members, resulting in a total of 215 observations.

The stations exhibit moderate dependence, with sample correlations given by $R_{12} = 0.33$, $R_{13} = 0.29$ and $R_{23} = 0.37$, where $R_{jk} = S_{jk}/(\bar{X}_j \bar{X}_k)^{1/2}$ for distinct $j, k \in \{1, 2, 3\}$. The data showed evidence of Poisson margins, as supported by a standard $\chi^2$ goodness-of-fit test (using the `gofstat` function in R), which yielded $p$-values of 0.92, 0.14 and 0.12 for Stations 1, 2 and 3, respectively.

5.1 A First Model

The three estimation methods discussed in Section 3 were applied to the BIXI data, namely the MM, the IFM approach and ML estimation. The results are summarised in Table 1.

For comparisons, estimation results under the classical model (1) are also given; refer to Table 2. Recall that the latter model defines correlated Poisson margins in terms of a common shock. By letting $Z$ denote the common shock variable, the classical model sets $X_1 = Y_1 + Z$, $X_2 = Y_2 + Z$ and $X_3 = Y_3 + Z$, for mutually independent Poisson random variables $Y_1, Y_2, Y_3$ and $Z$ with respective rates $\lambda_1 - \xi$, $\lambda_2 - \xi$, $\lambda_3 - \xi$ and $\xi$. In implementing the likelihood-based estimation techniques, in both the common and the comonotonic shock models, the MM estimates were used as starting values for the optimisation procedures.

Tables 1 and 2 display the model parameter estimates and the implied pairwise correlations along with 99% bootstrap confidence intervals. The latter intervals are based on 999 bootstrap replications, as there was one iteration out of 1000 in which the IFM optimisation failed to converge within the maximum number of iterations (set to 100 by default).
Table 1. Estimation results for the proposed trivariate Poisson model applied to the BIXI data.

|    | \( \lambda_1 \)   | \( \lambda_2 \)   | \( \lambda_3 \)   | \( \theta \)       |
|----|-------------------|-------------------|-------------------|-------------------|
| MM | 1.39 (1.17, 1.63)  | 1.53 (1.28, 1.79) | 2.74 (2.45, 3.05) | 0.37 (0.22, 0.52) |
| IFM| 1.39 (1.17, 1.63) | 1.53 (1.28, 1.79) | 2.74 (2.45, 3.05) | 0.23 (0.00, 0.35) |
| ML | 1.38 (1.15, 1.63) | 1.55 (1.29, 1.79) | 2.74 (2.45, 3.06) | 0.23 (0.14, 0.35) |

Estimates are given with 99% bootstrap confidence intervals in parentheses. IFM, inference functions for margins; ML, maximum likelihood; MM, method of moments.

Table 2. Estimation results for the classical trivariate Poisson model applied to the BIXI data.

|    | \( \lambda_1 \)   | \( \lambda_2 \)   | \( \lambda_3 \)   | \( \xi \)      |
|----|-------------------|-------------------|-------------------|---------------|
| MM | 1.39 (1.17, 1.63)  | 1.53 (1.28, 1.79) | 2.74 (2.45, 3.05) | 0.60 (0.34, 0.89) |
| ML | 1.39 (1.17, 1.63) | 1.53 (1.28, 1.79) | 2.74 (2.45, 3.05) | 0.34 (0.18, 0.52) |

Estimates are given with 99% bootstrap confidence intervals in parentheses. ML, maximum likelihood; MM, method of moments.

In model (3), the MM and IFM methods lead to marginal parameter estimates that are simply the marginal sample means. This is also the case for both estimation techniques (MM and ML) in the classical common shock model. Only ML estimation in the proposed \( \mathcal{M}_{P_d} \) model leads to different estimators \( \hat{\lambda} \). However, as shown in Table 1, this difference is small. This is consistent with what was observed in the simulations, further demonstrating that there is very little information concerning the marginal parameters in the dependence portion of the log-likelihood.

In both the classical and proposed models, the MM seemingly implies a stronger dependence level than the likelihood-based methods. Note that for both models, the bootstrap confidence intervals for the dependence parameter, respectively \( \hat{\theta} \) and \( \hat{\xi} \), are wider for the MM than for ML estimation. In the \( \mathcal{M}_{P_3} \) model, the IFM approach leads to the widest bootstrap confidence intervals for \( \hat{\theta} \). This is the result of outlier values for \( \hat{\theta} \) that are sometimes obtained in the optimisation procedure. This was also observed in the simulations.

Recall that the sample correlations were found to be \( R_{12} = 0.33 \), \( R_{13} = 0.29 \) and \( R_{23} = 0.37 \). The results for the proposed \( \mathcal{M}_{P_3} \) model (Table 1) show that each of these values falls within their respective interval obtained under the MM. In comparison, the bootstrap intervals obtained from the likelihood-based methods contain both \( R_{12} \) and \( R_{13} \) but fail to encompass \( R_{23} \). The classical model (Table 2) performs slightly worse: while the MM intervals contain each of the sample correlations, the intervals produced by the ML approach only contain \( R_{12} \). It is not surprising that the intervals based on the MM include all the pairwise sample correlations as this estimation technique is based on the sample covariances. The seemingly poor performance of the likelihood-based methods suggests that the underlying probabilistic model is not entirely satisfactory. It may thus be that a single dependence parameter is insufficient to regulate the strength of the dependence between all components of \((X_1, X_2, X_3)\).
Interestingly, there is more evidence in favour of the proposed model if the sample correlation between variables $X_j$ and $X_k$ is computed using variance, rather than mean, estimates for the marginal Poisson rates. Specifically, suppose that for distinct $j, k \in \{1, 2, 3\}$, the correlation between variables $X_j$ and $X_k$ is given by $R_{jk}^* = S_{jk}/(S_{jj}S_{kk})^{1/2}$, where $S_{jj}$ denotes the sample variance. One then finds $R_{12}^* = 0.26$, $R_{13}^* = 0.25$ and $R_{23}^* = 0.29$. In the proposed $\mathcal{MP}_3$ model, each of these sample correlations is contained within their respective bootstrap confidence intervals, whatever the estimation technique. This is not the case, however, in the classical model. While the MM intervals still contain the three pairwise sample correlations, the bootstrap intervals obtained from ML estimation fail to include $R_{23}^*$.

5.2 Discussion

It is not surprising that the $\mathcal{MP}_3$ model as well as the classical model (1) both present difficulties in quantifying the dependence structure, particularly the association between $X_2$ and $X_3$. There are notable similarities in the implied pairwise correlations between stations 1 and 3 and between stations 2 and 3 in the proposed comonotonic shock model; refer to Table 1. Indeed, the estimates for $\lambda_1$ and $\lambda_2$ are comparable across all methods, thus leading to similar implied correlations $\rho_{13}$ and $\rho_{23}$. Specifically, under the MM and IFM methods, the estimated marginal rates at stations 1 and 2 are respectively given by $\tilde{X}_1 = 1.39$ and $\tilde{X}_2 = 1.53$, which, by construction of the $\mathcal{MP}_3$ family, leads to $m_{\tilde{X}_1, \tilde{X}_3}(\theta) \approx m_{\tilde{X}_2, \tilde{X}_3}(\theta)$ for any $\theta \in (0, 1)$.

A similar phenomenon is seen in the classical common shock model; refer to Table 2. This feature is a drawback of both the classical and proposed models. Indeed, whether a common or comonotonic shock is used in the construction, the multivariate reduction technique implies that if $\lambda_j = \lambda_k$, then $\rho_{j\ell} = \rho_{k\ell}$ for any $\ell \in \{1, \ldots, d\} \setminus \{j, k\}$. This follows from the fact that in both model formulations, there is a single dependence parameter simultaneously regulating the correlation structure between all components. However, the data suggest that there is a stronger association between $X_2$ and $X_3$ in comparison with that between $X_1$ and $X_3$. The proposed comonotonic model, as well as the classical common shock model, is seemingly unable to capture this particular dependence structure. Nonetheless, the MM is able to better measure this dependence structure as this approach is based on the pairwise sample covariances.

5.3 Regression Model

In order to improve model fit, covariate effects were introduced into the proposed $\mathcal{MP}_3$ model. As mentioned in Section 3.5, there are numerous ways in which this can be performed. Here, we considered the simple case wherein the marginal Poisson rates are expressed in terms of explanatory variables through GLMs and the dependence parameter is held fixed. In the context of the BIXI data application, two categorical variables were included in the marginal GLMs: one for the day of the week (taking on seven levels) and the other for the month (taking on eight levels).

For computational ease, estimation was carried out using the IFM approach. The implementation of the latter consisted of first fitting marginal Poisson GLMs (with log link functions) to the data at each individual station to yield daily level estimated means $\hat{\lambda}_{ij} = \exp(t_i \hat{\beta}_j)$, where $t_i$ is used to denote the explanatory variables corresponding to observation $x_i$, for each $i \in \{1, \ldots, 215\}$ and $j \in \{1, 2, 3\}$. The next step in the estimation procedure then consisted of finding $\hat{\theta}$, which maximised $\ell(\hat{\Lambda}, \hat{\theta})$. This amounts to solving $\partial \ell_D(\hat{\Lambda}, \hat{\theta})/\partial \theta = 0$.

The fitted marginal GLMs showed similar patterns across stations. In particular, the estimated regression coefficients suggested that the months of July through September tended to have higher daily bike rental rates and that Thursdays were the busiest day of the week. An
analysis of deviance was used to assess the marginal model fits. For each of the three respective stations, comparisons with the null model confirmed that the inclusion of the covariate effects via GLMs provided significantly better fits. Furthermore, comparisons with the full (saturated) model indicated that, with the exception of station 2, the fitted models were inadequate simplifications of the full model. Indeed, the GLMs were rather simplistic, and it is likely that important covariates are missing from the marginal models.

The regression-based IFM approach resulted in \( \hat{\theta} = 0.18 \), with 99% bootstrap confidence interval given by (0.00, 0.35), based on 1000 bootstrap replications. Recall that in the covariate-free model considered in Section 5.1, the dependence parameter estimate was given by \( \hat{\theta} = 0.23 \) with a similar 99% bootstrap confidence interval of (0.00, 0.35). Adjusting for covariate effects in the model allows for varying daily level marginal Poisson rates for each site, which correspondingly allows for varying daily pairwise correlations \( \rho_{i,j,k} = m_{\lambda_{i,j},\lambda_{i,k}}(\theta)/\sqrt{\lambda_{i,j}\lambda_{i,k}} \), where \( \lambda_{i,j} \) is the \( i \)-th daily rate at Station \( j \). For the BIXI data application, the implied \( n = 215 \) daily pairwise correlations are summarised in Figure 7. The regression-based approach seems to imply slightly weaker levels of dependence across all stations. This suggests that perhaps the association between \( (X_1, X_2 \text{ and } X_3) \) can be partly explained by marginal covariate effects, thereby reducing the residual dependence.

The Akaike Information Criterion (AIC) can be used to compare the regression MP3 model with the null MP3 model considered in Section 5.1. In the regression-based MP3 model discussed here, there are 14 coefficients for the marginal GLMs at each of the three stations, in addition to the dependence parameter, leading to a total of 43 model parameters. In comparison, the null model only involves a total of four parameters. Despite the added complexity induced by the covariate effects, the AIC for the regression model was found to be 2083.208 in comparison to 2180.517 for the null model, suggesting that the inclusion of covariate effects does indeed lead to a better model. In contrast, when the Bayesian Information Criterion (BIC) is used, the regression model yields a value of 2228.145, while the null model results in a value of 2194.000, indicating that the simpler model is more appropriate. The differing conclusions based on the AIC and BIC measures are not surprising, as BIC tends to penalise more severely for model complexity.

Note that while these measures are convenient for contrasting the regression based model with the covariate-free model, it is important to bear in mind that the IFM estimation method

![Figure 7. Implied pairwise correlations for the regression based model using the inference function for margins (IFM) estimation approach](image-url)
Table 3. Model comparisons for models with and without covariates (labelled as regression and null model, respectively) applied to the BIXI data, according to different estimation methods.

|                      | Null Model | Regression Model |
|----------------------|------------|------------------|
| Log-likelihood:      | MM         | −1 092.270       |
|                      | IFM        | −1 086.259       |
|                      | MLE        | −1 086.233       |
| AIC:                 | MM         | 2 192.539        |
|                      | IFM        | 2 180.517        |
|                      | MLE        | 2 180.466        |
| BIC:                 | MM         | 2 206.022        |
|                      | IFM        | 42 194.000       |
|                      | MLE        | 2 193.948        |

AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; IFM, inference function for margins; MLE, maximum likelihood estimation; MM, method of moments.

does not optimise the full log-likelihood, but rather the marginal log-likelihoods and the modified dependence portion $\ell_D(\lambda, \theta)$, as detailed in Section 3.3. However, as demonstrated through the simulations, there is seemingly very little difference between the IFM and MLE approaches. For further comparisons, Table 3 provides the log-likelihood, AIC and BIC values for both the regression and null models, respectively, according to the various estimation methods considered.

6 Conclusion

This paper introduces a multivariate Poisson model based on a comonotonic shock. In considering a multivariate reduction technique with a comonotonic rather than common shock vector, the proposed model is able to capture strong degrees of (positive) dependence that the classical model fails to attain. Indeed, the proposed $\mathcal{MP}_d$ family allows for greater flexibility in the implied correlation structure while retaining an intuitive and interpretable stochastic representation akin to the classical setting. In particular, the proposed multivariate Poisson model was shown to be PQD ordered in terms of the dependence parameter $\theta$ and, moreover, can attain the upper Fréchet–Hoeffding bound. Several estimation techniques were discussed for the multivariate comonotonic shock model, each of which allows for consistent estimation of the model parameters. Through simulations as well as a real data application, each of the estimation methods was shown to perform well.

While the proposed model provides an improvement to the classical common shock construction, the $\mathcal{MP}_d$ family does not allow for a fully flexible representation, and indeed, the correlation structure suffers from certain inherent restrictions, as highlighted in the data application. First, the proposed model only allows for positive dependence, and there is no immediately obvious adaptation to the case of negatively correlated margins, except in dimension $d = 2$.

In the bivariate setting, the $\mathcal{BP}$ model can be adapted to allow for negative dependence by considering a counter-monotonic rather than comonotonic shock vector. In this setting, the shock vector has the form

$$ (Z_1, Z_2) = (G_{\theta \lambda_1}^{-1}(U), G_{\theta \lambda_2}^{-1}(1 - U)), $$

where $U \sim U(0, 1)$, thus yielding negatively correlated Poisson pairs. This model, denoted by $\mathcal{BP}^-$, was explored in Genest et al. (2018). It was shown there that the $\mathcal{BP}^-$ family is also

© 2020 The Authors. International Statistical Review published by John Wiley & Sons Ltd on behalf of International Statistical Institute.
A multivariate Poisson model based on comonotonic

PQD ordered in terms of $\theta$, with pairwise correlations spanning from $\rho_{\min}$ in the case of perfect negative dependence to 0 in the case of independence. The difficulty in extending the $BP^-$ model to higher dimensions stems from the fact that there is no convenient notion of counter-monotonicity in dimension $d > 2$. Although this problem may be circumvented by considering multiple comonotonic and counter-monotonic shocks, similar to what was outlined in Section 1, the benefit of the added flexibility will likely be outweighed by the resulting model complexities and difficulties in estimation.

Another shortcoming of the proposed comonotonic shock model is that the pairwise correlations are regulated by a single dependence parameter, thus imposing certain restrictions on the covariance structure. For example, suppose that the margins $X_j$ and $X_k$ have a common marginal parameter $\lambda_j = \lambda_k = \lambda$, for some distinct $j,k \in \{1, \ldots, d\}$. Then for any $\ell \in \{1, \ldots, d\}$, it follows that $\text{cov}(X_j, X_\ell) = \text{cov}(X_k, X_\ell)$ as both covariances are equal to $m_{\lambda, \lambda, \ell}\theta).$ This limitation of the model was highlighted in the data illustration. In the extremal cases, if there is a single pair $(X_r, X_t)$ within the vector $X \sim \mathcal{MP}_d(\lambda, \theta)$ exhibiting independence, then $\theta$ must necessarily be equal to 0, and the vector $X$ must be composed of mutually independent components. At the other end of the spectrum, if the pair $(X_r, X_t)$ exhibits perfect positive dependence, then $\theta = 1$ and the random vector $X$ must be comonotonic, that is, its distribution attains the upper Fréchet–Hoeffding bound. Clearly, there is a need to further develop the comonotonic shock construction to allow for more flexibility in the implied correlation structure. However, when $d = 2$ the resulting $BP$ model allows for a fully flexible (positive) dependence structure as it is based on a single pair to begin with.

Despite some shortcomings, the $\mathcal{MP}_d$ model is a significant improvement to the classical common shock representation and provides a novel modelling framework for multivariate count data. The intuitive form of the comonotonic shock construction allows for a straightforward understanding of the implied correlation structure with an interpretable parametrisation. The $\mathcal{MP}_d$ model will certainly prove to be a helpful tool for characterising dependence patterns exhibited by multivariate count data.

Acknowledgements

This paper is largely based on the doctoral dissertation of the first author, under the supervision of her coauthors. Funding in partial support of this work was provided by the Canada Research Chairs Programme, the Canadian Statistical Sciences Institute, HEC Montréal, l’Institut de valorisation des données (IVADO), as well as fellowships and grants from the Fonds de recherche du Québec—Nature et technologies, and the Natural Sciences and Engineering Research Council of Canada.

Appendix A: Proof of Proposition 1

Fix $\Lambda \in (0, \infty)^d$ and $\theta, \theta' \in [0, 1]$ with $\theta < \theta'$. By (3), the random vector $X$ can be expressed as $X = Y + Z$, where the summation is performed component-wise. Recall that in this formulation, the components of $Y = (Y_1, \ldots, Y_d)$ are mutually independent Poisson random variables with respective marginal means $(1-\theta)\lambda_1, \ldots, (1-\theta)\lambda_d$, while the components of $Z$ are comonotonic with respective marginal means $\lambda_1, \ldots, \lambda_d$. In an analogous way, write $X' = Y' + Z'$, where the components of $Y'$ are mutually independent Poisson random variables with respective marginal means $(1-\theta')\lambda_1, \ldots, (1-\theta')\lambda_d$, while the components of $Z'$ are comonotonic with respective marginal means $\theta'\lambda_1, \ldots, \theta'\lambda_d$.

Exploiting the fact that the univariate Poisson distribution is infinitely divisible, one can write $Y = T + S$, where $T$ is a vector of mutually independent Poisson random variables

International Statistical Review (2021), 89, 2, 323–348
© 2020 The Authors. International Statistical Review published by John Wiley & Sons Ltd on behalf of International Statistical Institute.
with respective marginal means \((1 - \theta')\lambda_1, \ldots, (1 - \theta')\lambda_d\), while \(S\) consists of mutually independent Poisson random variables with respective marginal means \((\theta' - \theta)\lambda_1, \ldots, (\theta' - \theta)\lambda_d\). Then the random vectors \(T\) and \(Y'\) are identically distributed, and thus trivially, \(T \sim_{PQD} Y'\).

By construction, both the random vectors \(S + Z\) and \(Z'\) have Poisson distributed margins with respective rates \(\theta'\lambda_1, \ldots, \theta'\lambda_d\). Given that \(Z'\) is comonotonic, it follows immediately from the Fréchet–Hoeffding bound inequalities that \(S + Z \sim_{PQD} Z'\). Theorem 9.A.4 of Shaked & Shanthikumar (2007) ensures that PQD ordering is closed under convolutions and thus \(T + S + Z \sim_{PQD} Y' + Z'\). This concludes the proof, because \(X = T + S + Z\) and \(X' = Y' + Z'\).

**Appendix B: Proof of Proposition 2**

Let \(\bar{X} = (X_1 + \ldots + X_n)/n\) be the sample mean vector and define the sample covariance matrix as

\[
S = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T.
\]

As stated in Corollary 1.2.18 of Muirhead (1982), it can be shown that \(S\) is asymptotically Gaussian. Following the notation of Muirhead (1982), define the vectorisation of the sample covariance matrix as \(\text{vec}(S) = (S_{11}, S_{12}, \ldots, S_{1d}, S_{22}, \ldots, S_{2d}, \ldots, S_{1d}, S_{2d}, \ldots, S_{dd})^T\).

For each \(i \in \{1, \ldots, n\}\), let \(\Sigma\) denote the theoretical covariance \(\text{cov}(X_i)\) and denote by \(\mu\) the mean vector \(E(X_i)\). Then, as \(n \to \infty\), \(\sqrt{n} \{\text{vec}(S - \Sigma)\} \sim N_d(0, V)\), where \(V = \text{cov}[^{\text{vec}}(X_i - \mu)(X_i - \mu)^\top]\). Now let \(\text{vec}(S^*)\) denote the vectorised sample covariance wherein repeat arguments as well as the sample variance terms are removed, that is,

\[
\text{vec}(S^*) = (S_{12}, S_{13}, \ldots, S_{1d}, S_{23}, \ldots, S_{2d}, \ldots, S_{1d}, S_{2d}, \ldots, S_{dd})^T,
\]

so that \(\text{vec}(S^*)\) is of dimension \(d^* = d(d - 1)/2\) with elements \(S_{jk}\) for all \(j, k \in \{1, \ldots, d\}\) with \(j < k\). Further, let \(\text{vec}(\Sigma^*)\) be the corresponding vectorised true covariance matrix, namely \(\text{vec}(\Sigma^*) = (\sigma_{12}, \sigma_{13}, \ldots, \sigma_{1d}, \sigma_{23}, \ldots, \sigma_{d-1d})^T\), where \(\sigma_{jk} = m_{\lambda_j, \lambda_k}(\theta)\).

By the Cramér–Wold theorem, \(\text{vec}(S^*)\) has an asymptotic normal distribution centred at \(\text{vec}(\Sigma^*)\) with asymptotic variance \(V^* = \text{cov}[^{\text{vec}}((X_i^* - \mu^*)(X_i^* - \mu^*)^\top)]\), where \(\text{vec}((X_i^* - \mu^*)(X_i^* - \mu^*)^\top)\) is the column vector given by

\[
\begin{bmatrix}
(X_{i1} - \lambda_1)(X_{i2} - \lambda_2) \\
(X_{i1} - \lambda_1)(X_{i1d} - \lambda_d) \\
(X_{i2} - \lambda_2)(X_{i13} - \lambda_3) \\
\vdots \\
(X_{i2} - \lambda_2)(X_{i1d} - \lambda_d) \\
(X_{id-1} - \lambda_{d-1})(X_{id} - \lambda_d)
\end{bmatrix}.
\]

Define the function \(\Gamma : \mathbb{R}^{d^*} \mapsto [0, 1]^{d^*}\) to be the vectorised inverse of the covariance function with components \(\gamma_{jk}(x) = m_{\lambda_j, \lambda_k}^{-1}(x)\) such that

\[
\Gamma(S^*) = (\gamma_{12}(S_{12}), \gamma_{13}(S_{13}), \ldots, \gamma_{1d}(S_{1d}), \gamma_{23}(S_{23}), \ldots, \gamma_{d-1d}(S_{d-1d}))
= (\tilde{\theta}_{12}, \tilde{\theta}_{13}, \ldots, \tilde{\theta}_{1d}, \tilde{\theta}_{23}, \ldots, \tilde{\theta}_{d-1d}).
\]
The multivariate delta method can then be applied to conclude that, as $n \to \infty$,

$$\sqrt{n} \left[ \text{vec} \{ \Gamma(S^*) - \Gamma(\Sigma^*) \} \right] \rightsquigarrow N_{d^*} \left[ 0, \{ \Gamma'(\Sigma^*) \}V^*\{ \Gamma'(\Sigma^*) \}^T \right],$$

where $\Gamma'$ denotes the component-wise derivative of the function $\Gamma$. Notice that each component of the vectorisation of $\Gamma(\Sigma^*)$ is identically equal to $\theta$. It then follows from the Cramér–Wold theorem that the estimator $\tilde{\theta}$ defined as the linear combination

$$\tilde{\theta} = \frac{1}{d^*} \sum_{j=1}^{d-1} \sum_{k=j+1}^{d} \tilde{\theta}_{jk}$$

is also asymptotically Gaussian. In particular, properties of the normal distribution ensure that, as $n \to \infty$, $\sqrt{n}(\tilde{\theta} - \theta) \rightsquigarrow N(0, \tilde{V})$, where

$$\tilde{V} = \frac{1}{(d^*)^4} \sum_{j=1}^{d-1} \sum_{k=j+1}^{d-1} \sum_{r=1}^{d-1} \sum_{s=r+1}^{d} \{ \gamma'_{jk}(\sigma_{jk}) \} \{ \gamma'_{rs}(\sigma_{rs}) \} \times \text{cov}\{(X_{ij} - \lambda_j)(X_{ik} - \lambda_k), (X_{ir} - \lambda_r)(X_{is} - \lambda_s)\}.$$ 

This concludes the argument.

References

Campbell, J.T. (1934). The Poisson correlation function. *P. Edinburgh Math. Soc.*, 4, 18–26.

Ferguson, T.S. (1996). *A Course in Large Sample Theory*. Chapman & Hall: London.

Genest, C., Mesfioui, M. & Schulz, J. (2018). A new bivariate Poisson common shock model covering all possible degrees of dependence. *Stat. Probab. Lett.*, 140, 202–209.

Gómez-Déniz, E., Sarabia, J.M. & Balakrishnan, N. (2012). A multivariate Poisson–Lindley distribution: extensions and actuarial applications. *ASTIN Bull.*, 42, 655–678.

Griffiths, R.C., Milne, R.K. & Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Aust. J. Stat.*, 21, 238–255.

Inouye, D.I., Yang, E., Allen, G.I. & Ravikumar, P. (2017). A review of multivariate distributions for count data derived from the Poisson distribution. *Wiley Interdiscip. Rev. Comput. Stat.*, 9(3), e1398. https://doi.org/10.1002/wics.1398

Joe, H. (1990). Multivariate concordance. *J. Multivariate Anal.*, 35, 12–30.

Joe, H. (2004). Modeling and generating dependent risk processes for IRM and DFA. *ASTIN Bull.*, 34, 333–360.

Johnson, N.L., Kotz, S. & Balakrishnan, N. (1997). *Continuous Univariate Distributions*. Vol. 2. Wiley: New York.
Sarabia, J.M. & Gómez-Déniz, E. (2011). Multivariate Poisson-Beta distributions with applications. *Comm. Statist. Theory Methods*, **40**, 1093–1108.

Schulz, J. (2018). *Multivariate Poisson Models Based on Comonotonic and Counter-Monotonic Shocks*. Doctoral Dissertation, McGill University: Montréal, QC, Canada.

Sellers, K.F., Morris, D.S. & Balakrishnan, N. (2016). Bivariate Conway–Maxwell–Poisson distribution: formulation, properties, and inference. *J. Multivariate Anal.*, **150**, 152–168.

Shaked, M. & Shanthikumar, J.G. (2007). *Stochastic Orders*. Springer: New York.

Smith, M.S. & Khaled, M.A. (2012). Estimation of copula models with discrete margins via Bayesian data augmentation. *J. Amer. Statist. Assoc.*, **107**, 290–303.

Tsionas, E.G. (1999). Bayesian analysis of the multivariate Poisson distribution. *Comm. Statist. Theory Methods*, **28**, 431–451.

Van Ophem, H. (1999). A general method to estimate correlated discrete random variables. *Economet. Theory*, **15**, 228–237.

[Received August 2019, Revised July 2020, Accepted August 2020]

**Supporting Information**

Supporting information may be found in the online version of this article.