

Conditional least squares estimators for multitype Galton–Watson processes

FANNI NEDÉNYI*

Communicated by Gy. Pap

Abstract. The goal of the paper is to estimate the first four moments of the offspring and innovation distributions of subcritical, time-homogeneous multitype Galton–Watson processes. We apply the CLS (Conditional Least Squares) and the WCLS (Weighted Conditional Least Squares) methods for this purpose. It is also shown that under the proper moment conditions the estimators are strongly consistent and the ones of the first two moments are asymptotically normal.

1. Introduction

Galton–Watson models have been the subject of intense investigation since the 1950's. General properties of the processes — such as generating functions, asymptotical results, and ergodicity — are discussed in [1], [10], and [8] in detail. Simultaneously, the estimation of the expected values of the offspring and the innovation distributions has become an important statistical problem, and several results have been delivered separately for the three cases of single-type Galton–Watson processes. In the subcritical case the most known estimators are discussed by [5, 6]. Their formulae are derived from the moments of the invariant distribution, which method is analogous to the parameter estimation of AR(1) models based on the Yule–Walker

Received July 19, 2014, and in revised form February 5, 2015.

AMS Subject Classifications (2010): 60J80, 62F10, 62F12.

Key words and phrases: branching processes, Galton–Watson, moments, parameter estimation, conditional least squares.

*This research was realized in the frames of TÁMOP 4.2.4. A/2-11-1-2012-0001 “National Excellence Program – Elaborating and operating an inland student and researcher personal support system.” The project was subsidized by the European Union and co-financed by the European Social Fund.

equations. In the subcritical case, in the paper [17] estimations are given also for the variances of the offspring and the innovation distributions based on the already defined estimators of their first moments. This is a technique that is applied in this paper as well. Their estimators are consistent, asymptotically normal, and they obey the law of the iterated logarithm. In the book [4] an optimality is defined among parameter estimating functions and the estimators that come from equating them to zero. These optimal parameter estimators are called the quasi-likelihood estimators. They are discussed for single-type Galton–Watson processes, see the pages 15–16, 35–36, 69–70. As a generalization, the estimation of the mean matrix has been worked out also for multitype Galton–Watson processes in [13]. Similar formulae are derived by [2] for GINAR (Generalized INteger-valued AutoRegressive) models, that are special cases of multitype Galton–Watson processes.

In our paper we use another estimation method, the CLS (Conditional Least Squares) estimation, which was introduced by [9]. This method was worked out for an arbitrary discrete time stochastic process X_0, X_1, \dots defined along with the corresponding generated filtration $\mathcal{F}_0, \mathcal{F}_1, \dots$. Suppose that $E[X_n | \mathcal{F}_{n-1}] = f(\theta, \mathcal{F}_{n-1})$ holds for every n with a known function f and an unknown parameter θ . The CLS estimator of θ based on the sample X_0, \dots, X_m is defined as the minimum point of the sum of squares $\sum_{n=1}^m [X_n - f(\theta, \mathcal{F}_{n-1})]^2$. Besides other examples, Klimko and Nelson illustrated the method by estimating the means of the offspring and the innovation distributions of single-type Galton–Watson processes. As a modification of the CLS method, [12] suggested the so-called WCLS (Weighted Conditional Least Squares) estimation. In this case the terms of the sum of squares above are divided by a suitable weight function $w_n = w_n(\theta, X_n, \mathcal{F}_{n-1})$, and the estimation is performed by minimizing the weighted sum. Nelson highlighted that using a correctly chosen weight function the asymptotic properties of the estimators can be improved compared to the CLS ones.

Based on these ideas [15] and [16] applied the CLS and the WCLS methods to estimate both the means and the variances of the offspring and the innovation distributions of single-type Galton–Watson processes that are not even necessarily subcritical. In these papers the weight function $w_n = X_{n-1} + 1$ was introduced, with which the WCLS estimation can be considered also as the CLS estimation based on the weighted process $X_n / \sqrt{X_{n-1} + 1}$. Let us note that [15] provides a detailed overview on the most important earlier results about the estimations of the moments including the subcritical, the critical, and the supercritical cases as well. The CLS estimation was applied also for INAR (INteger-valued AutoRegressive) models by [3].

Our aim is to determine the CLS and the WCLS estimators of the first four

moments of the offspring and the innovation variables of multitype Galton–Watson processes, in the latter case by applying the multitype version of the weight function of [15]. We show that, in the subcritical case, under the proper moment conditions these estimators are strongly consistent and the ones of the expected values and variances are asymptotically normal as well. The results of the paper are stated in Section 2, and the proofs are presented in Section 3. We note that an application of these estimators is detailed in [11], where a change detection test is presented for multitype Galton–Watson models.

2. Main results

2.1. Notations and assumptions

Consider a process $\mathbb{X}_n = [X_{n,1}, \dots, X_{n,p}]^\top$, $n = 0, 1, \dots$, on the state space \mathbb{Z}_+^p with a fixed positive integer parameter p and a random or deterministic initial vector \mathbb{X}_0 . (In our paper the notation \mathbb{Z}_+ stands for the set of the nonnegative integers.) We say that the process \mathbb{X}_n , $n = 0, 1, \dots$, is a *time-homogeneous multitype Galton–Watson process* if it can be represented in the form

$$\mathbb{X}_n = \sum_{k=1}^{X_{n-1,1}} \boldsymbol{\xi}_1(n, k) + \dots + \sum_{k=1}^{X_{n-1,p}} \boldsymbol{\xi}_p(n, k) + \boldsymbol{\eta}(n), \quad n = 1, 2, \dots,$$

where

$$\boldsymbol{\xi}_i(n, k), \quad \boldsymbol{\eta}(n), \quad k, n = 1, 2, \dots, \quad i = 1, \dots, p, \quad (1)$$

are \mathbb{Z}_+^p -valued random vectors being independent of each other and of the initial variable \mathbb{X}_0 , the offspring variables $\boldsymbol{\xi}_i(n, k)$, $n, k = 1, 2, \dots$, are identically distributed for every i , and the innovation variables $\boldsymbol{\eta}(n)$, $n = 1, 2, \dots$, are identically distributed as well. For simplicity we refer to the distributions of these random vectors by $\boldsymbol{\xi}_i$ and $\boldsymbol{\eta}$ with components $\xi_{1,i}, \dots, \xi_{p,i}$ and η_1, \dots, η_p , respectively, as their distributions do not depend on the indices n and k . As a regularity condition we assume that the random vectors $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_p, \boldsymbol{\eta}$ have finite second moments. To shorten the notations we introduce the process and the matrix

$$\mathbb{Y}_n := \begin{bmatrix} \mathbb{X}_n \\ 1 \end{bmatrix} = [X_{n,1}, \dots, X_{n,p}, 1]^\top \quad \text{and} \quad \mathbb{I}_{m,r} = \sum_{n=1}^m \frac{\mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^r},$$

along with the generated filtration $\mathcal{F}_n = \sigma(\mathbb{X}_k : k \leq n) = \sigma(\mathbb{Y}_k : k \leq n)$, for $n, r \in \mathbb{Z}_+$, $m = 1, 2, \dots$, where $\mathbf{1} = [1, \dots, 1]^\top \in \mathbb{R}^{p+1}$. In Subsection 3.3 we show that under our assumptions the process \mathbb{Y}_n , $n \in \mathbb{Z}_+$, is ergodic with an invariant

distribution having finite second moment. The notation $\tilde{\mathbb{Y}}$ stands for an arbitrary variable with this unique invariant distribution.

The main goal of the paper is to estimate the moments of the variables $\xi_1, \dots, \xi_p, \eta$. The means of the components of the offspring and the innovation variables are denoted by $\mu_{i,j} := E(\xi_{i,j})$ and $\mu_{i,\eta} := E(\eta_i)$, $i, j = 1, \dots, p$. Also, we introduce the matrices

$$\mathbf{m} := \begin{bmatrix} \mu_{1,1} & \dots & \mu_{1,p} \\ \vdots & \ddots & \vdots \\ \mu_{p,1} & \dots & \mu_{p,p} \end{bmatrix},$$

$$\mathbf{M} := \begin{bmatrix} \boldsymbol{\mu}_1^\top \\ \vdots \\ \boldsymbol{\mu}_p^\top \end{bmatrix} := \begin{bmatrix} \mu_{1,1} & \dots & \mu_{1,p} & \mu_{1,\eta} \\ \vdots & \ddots & \vdots & \vdots \\ \mu_{p,1} & \dots & \mu_{p,p} & \mu_{p,\eta} \end{bmatrix} \in \mathbb{R}^{p \times (p+1)},$$

where $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_p \in \mathbb{R}^{p+1}$ are the transposes of the rows of \mathbf{M} . Similarly to the definition of \mathbf{M} we define the matrices $\mathbf{V}, \mathbf{A}, \mathbf{B} \in \mathbb{R}^{p \times (p+1)}$ of the second, third, and fourth central moments with rows $\mathbf{v}_i^\top, \boldsymbol{\alpha}_i^\top, \boldsymbol{\beta}_i^\top$, $i = 1, \dots, p$, respectively. We introduce the vectors of the second and the fourth order mixed central moments by the form

$$\mathbf{v}_{(i,i')} := [v_{(i,i'),1}, \dots, v_{(i,i'),p}, v_{(i,i'),\eta}]^\top, \quad \boldsymbol{\beta}_{(i,i')} := [\beta_{(i,i'),1}, \dots, \beta_{(i,i'),p}, \beta_{(i,i'),\eta}]^\top,$$

with the components $v_{(i,i'),j} := \text{Cov}(\xi_{i,j}, \xi_{i',j})$, $v_{(i,i'),\eta} := \text{Cov}(\eta_i, \eta_{i'})$,

$$\beta_{(i,i'),j} := E[(\xi_{i,j} - E\xi_{i,j})(\xi_{i',j} - E\xi_{i',j})^2], \quad \beta_{(i,i'),\eta} := E[(\eta_i - E\eta_i)(\eta_{i'} - E\eta_{i'})^2],$$

where $i, i', j = 1, \dots, p$. Let us note that $\mathbf{v}_{(i,i)} = \mathbf{v}_i$ and $\boldsymbol{\beta}_{(i,i)} = \boldsymbol{\beta}_i$.

In our paper the vector norm is the Euclidean norm, and the notation $\mathbf{0}$ stands for the null matrix of arbitrary dimensions. The Hadamard product of two arbitrary vectors $\mathbf{c} = [c_1, \dots, c_{p+1}]^\top$ and $\mathbf{d} = [d_1, \dots, d_{p+1}]^\top$ is defined as

$$\mathbf{c} \circ \mathbf{d} := [c_1 d_1, \dots, c_{p+1} d_{p+1}]^\top \in \mathbb{R}^{p+1}.$$

Similarly, for any $r \in \mathbb{Z}_+$ and random or deterministic vector or matrix \mathbf{N} , the notation $\mathbf{N}^{(r)}$ stands for the matrix of the r -th powers of the components.

Consider an arbitrary type $j = 1, \dots, p$. We say that type j dies out if $(\mathbf{m}^n)_{j,i} = 0$ for every $n \in \mathbb{Z}_+$ and for every type $i = 1, \dots, p$ satisfying $E(\eta_i) > 0$. This property means that there is no innovation in type j , furthermore, a type j individual can not be a descendant of an individual of any type that arrived to the population through immigration. Since in our paper the Galton–Watson process is subcritical

by assumption, types that die out disappear from the population in finitely many steps with probability one.

We summarize the previously mentioned conditions in the following assumption.

Assumption 1. Unless stated otherwise we assume that the multitype Galton–Watson process $\mathbb{X}_0, \mathbb{X}_1, \dots$ fulfills the following assumptions.

- (i) The process is subcritical meaning that the spectral radius of the mean matrix \mathbf{m} is strictly less than 1.
- (ii) The variables in (1) all have finite second moments.
- (iii) None of the types die out.
- (iv) There exists no vector $\mathbf{c} \in \mathbb{R}^p \setminus \{\mathbf{0}\}$ such that the variables $\mathbf{c}^\top \boldsymbol{\xi}_1, \dots, \mathbf{c}^\top \boldsymbol{\xi}_p, \mathbf{c}^\top \boldsymbol{\eta}$ are all degenerate and $\mathbf{c}^\top \boldsymbol{\xi}_1 = \dots = \mathbf{c}^\top \boldsymbol{\xi}_p = 0$ almost surely.

Assumptions (i) and (ii) imply that the process is ergodic and the invariant distribution has finite second moment. The remaining conditions are required to ensure that the estimators introduced in the next subsection are well defined if the sample size is large enough.

2.2. The estimators of the moments

In our paper we determine the Conditional Least Squares (CLS) and the Weighted Conditional Least Squares (WCLS) estimators of the matrices $\mathbf{M}, \mathbf{V}, \mathbf{A}$, and \mathbf{B} . The CLS estimation was introduced by [9] to estimate the parameter of parameterized discrete time stochastic processes. To perform the estimation we must consider two martingale difference sequences

$$\mathbb{U}_n := \mathbb{X}_n - E(\mathbb{X}_n | \mathcal{F}_{n-1}) \quad \text{and} \quad \mathbb{V}_n := \mathbb{U}_n^{(2)} - E(\mathbb{U}_n^{(2)} | \mathcal{F}_{n-1}), \quad n = 1, 2, \dots$$

As we suppose that the offspring and innovation variables have finite second moments, these martingale differences are well defined and in Subsection 3.1 we show that they can be written as

$$\mathbb{U}_n = \mathbb{X}_n - \mathbf{M}\mathbb{Y}_{n-1}, \quad \mathbb{V}_n = \mathbb{U}_n^{(2)} - \mathbf{V}\mathbb{Y}_{n-1}, \quad n = 1, 2, \dots \tag{2}$$

We also introduce the process $\mathbb{K}_n := \mathbb{U}_n^{(4)} - 3(\mathbf{V}\mathbb{Y}_{n-1})^{(2)} + 3\mathbf{V}^{(2)}\mathbb{Y}_{n-1}$, $n = 1, 2, \dots$

Based on the calculations presented in Subsection 3.2, the formulae for the CLS estimators based on the sample $\mathbb{X}_0, \dots, \mathbb{X}_m$ are

$$\widehat{\mathbf{M}}_m = \left[\sum_{n=1}^m \mathbb{X}_n \mathbb{Y}_{n-1}^\top \right] \mathbb{I}_{m,0}^{-1}, \quad \widehat{\mathbf{V}}_m = \left[\sum_{n=1}^m \widehat{\mathbb{U}}_{m,n}^{(2)} \mathbb{Y}_{n-1}^\top \right] \mathbb{I}_{m,0}^{-1}$$

$$\widehat{\mathbf{A}}_m = \left[\sum_{n=1}^m \widehat{\mathbf{U}}_{m,n}^{(3)} \mathbf{Y}_{n-1}^\top \right] \mathbb{I}_{m,0}^{-1}, \quad \widehat{\mathbf{B}}_m = \left[\sum_{n=1}^m \widehat{\mathbb{K}}_{m,n} \mathbf{Y}_{n-1}^\top \right] \mathbb{I}_{m,0}^{-1},$$

where

$$\widehat{\mathbf{U}}_{m,n} := \mathbb{X}_n - \widehat{\mathbf{M}}_m \mathbb{Y}_{n-1}, \quad \widehat{\mathbb{K}}_{m,n} := \widehat{\mathbf{U}}_{m,n}^{(4)} - 3(\widehat{\mathbf{V}}_m \mathbb{Y}_{n-1})^{(2)} + 3\widehat{\mathbf{V}}_m^{(2)} \mathbb{Y}_{n-1}$$

are the natural CLS estimators of \mathbb{U}_n and \mathbb{K}_n , respectively, $n = 1, 2, \dots$

We also define another type of parameter estimators called the Weighted Conditional Least Squares (WCLS) estimators. The weighted version of the CLS estimation was introduced by [12] with a general weight function to estimate the parameters in multivariate linear regression models. The WCLS estimation used in our paper is a special case of Nelson’s method and it is defined as the CLS estimation based on the weighted process $\mathbb{X}'_n := \mathbb{X}_n / \sqrt{\mathbf{1}^\top \mathbb{Y}_{n-1}}$, $n = 1, 2, \dots$. Our definition is originated from [15] and [16] who used the WCLS estimation to estimate the mean and the variance of the offspring and the innovation distributions in single-type Galton–Watson processes. In Subsection 3.2 we show that the WCLS estimators of the moments based on the sample $\mathbb{X}_0, \dots, \mathbb{X}_m$ are

$$\begin{aligned} \widehat{\mathbf{M}}'_m &= \left[\sum_{n=1}^m \frac{\mathbb{X}_n \mathbf{Y}_{n-1}^\top}{\mathbf{1}^\top \mathbb{Y}_{n-1}} \right] \mathbb{I}_{m,1}^{-1}, & \widehat{\mathbf{V}}'_m &= \left[\sum_{n=1}^m \frac{\widehat{\mathbf{U}}_{m,n}^{(2)} \mathbf{Y}_{n-1}^\top}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^2} \right] \mathbb{I}_{m,2}^{-1}, \\ \widehat{\mathbf{A}}'_m &= \left[\sum_{n=1}^m \frac{\widehat{\mathbf{U}}_{m,n}^{(3)} \mathbf{Y}_{n-1}^\top}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^3} \right] \mathbb{I}_{m,3}^{-1}, & \widehat{\mathbf{B}}'_m &= \left[\sum_{n=1}^m \frac{\widehat{\mathbb{K}}'_{m,n} \mathbf{Y}_{n-1}^\top}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^4} \right] \mathbb{I}_{m,4}^{-1}, \end{aligned}$$

with the WCLS estimators

$$\widehat{\mathbf{U}}'_{m,n} := \mathbb{X}_n - \widehat{\mathbf{M}}'_m \mathbb{Y}_{n-1}, \quad \widehat{\mathbb{K}}'_{m,n} := \widehat{\mathbf{U}}'_{m,n}^{(4)} - 3(\widehat{\mathbf{V}}'_m \mathbb{Y}_{n-1})^{(2)} + 3\widehat{\mathbf{V}}'^{(2)}_m \mathbb{Y}_{n-1}$$

of \mathbb{U}_n and \mathbb{K}_n , respectively, $n = 1, 2, \dots$

Let us note that for a given m the CLS and the WCLS estimators of the moments exist only with a probability lower than 1 because the matrix inverses in the formulae above are not well defined for every sample $\mathbb{X}_0, \dots, \mathbb{X}_m$. However, in the next theorem we show that the estimators do exist with asymptotic probability 1 as the sample size goes to infinity. Also, the estimators are strongly consistent under appropriate moment conditions.

Theorem 2.1. *Assume that Assumption 1 holds.*

- (i) *The CLS and the WCLS estimators exist with probability tending to 1 as $m \rightarrow \infty$.*
- (ii) *If the variables in (1) have finite second, third, fourth, and fifth moments, then the estimators $\widehat{\mathbf{M}}_m, \widehat{\mathbf{V}}_m, \widehat{\mathbf{A}}_m$, and $\widehat{\mathbf{B}}_m$ are strongly consistent, respectively.*

(iii) *The estimators $\widehat{\mathbf{M}}'_m$ and $\widehat{\mathbf{V}}'_m$ are strongly consistent. If additionally the variables in (1) have finite third and fourth moments, then the estimators $\widehat{\mathbf{A}}'_m$ and $\widehat{\mathbf{B}}'_m$ are also strongly consistent, respectively.*

Let $\widehat{\boldsymbol{\mu}}_{m,1}, \dots, \widehat{\boldsymbol{\mu}}_{m,p}$ stand for the transposes of the row vectors of the matrix $\widehat{\mathbf{M}}_m$, which can be considered as the CLS estimators of the vectors $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_p$ based on the sample $\mathbb{X}_0, \dots, \mathbb{X}_m$. In several multitype Galton–Watson models some of the rows of \mathbf{M} are known and the goal is to estimate the remaining rows under this *a priori* information. Such an example is the generalized integer-valued autoregressive (GINAR) process where only the first row of the matrix \mathbf{M} is unknown. (See [2] for the model.) In Remark 3.1 of Subsection 3.2 we show that the estimators $\widehat{\boldsymbol{\mu}}_{m,1}, \dots, \widehat{\boldsymbol{\mu}}_{m,p}$ can be calculated independently meaning that the knowledge of the exact values of some of the rows in \mathbf{M} does not change the estimators of the remaining ones. Furthermore, the same statement is true for the row vectors of all of the matrix estimators presented in this subsection.

Consider the random variables

$$Z_{i,j} := \boldsymbol{\beta}_{(i,j)}^\top \widetilde{\mathbf{Y}} - (\mathbf{v}_i \circ \mathbf{v}_j)^\top \widetilde{\mathbf{Y}} + 2(\mathbf{v}_{(i,j)}^\top \widetilde{\mathbf{Y}})^2 - 2(\mathbf{v}_{(i,j)}^{(2)})^\top \widetilde{\mathbf{Y}}, \quad i, j = 1, \dots, p,$$

and let $\widehat{\mathbf{v}}_{m,i}$, $\widehat{\boldsymbol{\mu}}'_{m,i}$, and $\widehat{\mathbf{v}}'_{m,i}$ denote the transpose of the i -th row of $\widehat{\mathbf{V}}_m$, $\widehat{\mathbf{M}}'_m$, and $\widehat{\mathbf{V}}'_m$, respectively. (For formulae see Remark 3.1.) In our next theorem we investigate the asymptotic distribution of these estimators.

Theorem 2.2. *Assume that Assumption 1 holds.*

(i) *If for some $\varepsilon > 0$ the $(4 + \varepsilon)$ -th moments of the variables in (1) exist, then*

$$\sqrt{m} \begin{bmatrix} \widehat{\boldsymbol{\mu}}_{m,1} - \boldsymbol{\mu}_1 \\ \vdots \\ \widehat{\boldsymbol{\mu}}_{m,p} - \boldsymbol{\mu}_p \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_{p(p+1)} \left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{1,1}^{\mathbf{M}} & \cdots & \boldsymbol{\Sigma}_{1,p}^{\mathbf{M}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{p,1}^{\mathbf{M}} & \cdots & \boldsymbol{\Sigma}_{p,p}^{\mathbf{M}} \end{bmatrix} \right), \quad m \rightarrow \infty,$$

where

$$\boldsymbol{\Sigma}_{i,j}^{\mathbf{M}} = \left(E[\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top] \right)^{-1} E \left[(\mathbf{v}_{(i,j)}^\top \widetilde{\mathbf{Y}}) \widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top \right] \left(E[\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top] \right)^{-1}, \quad i, j = 1, \dots, p.$$

If additionally the components of the variables in (1) are pairwise uncorrelated, then the estimators $\widehat{\boldsymbol{\mu}}_{m,1}, \dots, \widehat{\boldsymbol{\mu}}_{m,p}$ are asymptotically independent.

(ii) *If for some $\varepsilon > 0$ the $(2 + \varepsilon)$ -th moments of the variables in (1) exist, then*

$$\sqrt{m} \begin{bmatrix} \widehat{\boldsymbol{\mu}}'_{m,1} - \boldsymbol{\mu}_1 \\ \vdots \\ \widehat{\boldsymbol{\mu}}'_{m,p} - \boldsymbol{\mu}_p \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_{p(p+1)} \left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{1,1}^{\mathbf{M}'} & \cdots & \boldsymbol{\Sigma}_{1,p}^{\mathbf{M}'} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{p,1}^{\mathbf{M}'} & \cdots & \boldsymbol{\Sigma}_{p,p}^{\mathbf{M}'} \end{bmatrix} \right), \quad m \rightarrow \infty,$$

where

$$\Sigma_{i,j}^{\mathbf{M}'} = \left(E \left[\frac{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top}{\mathbf{1}^\top \widetilde{\mathbf{Y}}} \right] \right)^{-1} E \left[(\mathbf{v}_{(i,j)}^\top \widetilde{\mathbf{Y}}) \frac{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top}{(\mathbf{1}^\top \widetilde{\mathbf{Y}})^2} \right] \left(E \left[\frac{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top}{\mathbf{1}^\top \widetilde{\mathbf{Y}}} \right] \right)^{-1}, \quad i, j = 1, \dots, p.$$

If additionally the components of the variables in (1) are pairwise uncorrelated, then the estimators $\widehat{\boldsymbol{\mu}}'_{m,1}, \dots, \widehat{\boldsymbol{\mu}}'_{m,p}$ are asymptotically independent.

(iii) If for some $\varepsilon > 0$ the $(6 + \varepsilon)$ -th moments of the variables in (1) exist, then

$$\sqrt{m} \begin{bmatrix} \widehat{\mathbf{v}}_{m,1} - \mathbf{v}_1 \\ \vdots \\ \widehat{\mathbf{v}}_{m,p} - \mathbf{v}_p \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_{p(p+1)} \left(\mathbf{0}, \begin{bmatrix} \Sigma_{1,1}^{\mathbf{V}} & \cdots & \Sigma_{1,p}^{\mathbf{V}} \\ \vdots & \ddots & \vdots \\ \Sigma_{p,1}^{\mathbf{V}} & \cdots & \Sigma_{p,p}^{\mathbf{V}} \end{bmatrix} \right), \quad m \rightarrow \infty,$$

where

$$\Sigma_{i,j}^{\mathbf{V}} = \left(E[\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top] \right)^{-1} E \left[Z_{i,j} \widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top \right] \left(E[\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top] \right)^{-1}, \quad i, j = 1, \dots, p.$$

If additionally the components of the variables in (1) are pairwise independent, then the estimators $\widehat{\mathbf{v}}_{m,1}, \dots, \widehat{\mathbf{v}}_{m,p}$ are asymptotically independent.

(iv) If the fourth moments of the variables in (1) exist, then

$$\sqrt{m} \begin{bmatrix} \widehat{\mathbf{v}}'_{m,1} - \mathbf{v}_1 \\ \vdots \\ \widehat{\mathbf{v}}'_{m,p} - \mathbf{v}_p \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_{p(p+1)} \left(\mathbf{0}, \begin{bmatrix} \Sigma_{1,1}^{\mathbf{V}'} & \cdots & \Sigma_{1,p}^{\mathbf{V}'} \\ \vdots & \ddots & \vdots \\ \Sigma_{p,1}^{\mathbf{V}'} & \cdots & \Sigma_{p,p}^{\mathbf{V}'} \end{bmatrix} \right), \quad m \rightarrow \infty,$$

where

$$\Sigma_{i,j}^{\mathbf{V}'} = \left(E \left[\frac{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top}{(\mathbf{1}^\top \widetilde{\mathbf{Y}})^2} \right] \right)^{-1} E \left[Z_{i,j} \frac{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top}{(\mathbf{1}^\top \widetilde{\mathbf{Y}})^4} \right] \left(E \left[\frac{\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^\top}{(\mathbf{1}^\top \widetilde{\mathbf{Y}})^2} \right] \right)^{-1}, \quad i, j = 1, \dots, p.$$

If additionally the components of the variables in (1) are pairwise independent, then the estimators $\widehat{\mathbf{v}}'_{m,1}, \dots, \widehat{\mathbf{v}}'_{m,p}$ are asymptotically independent.

Let us note that for any type i we have $\mathbf{v}_{(i,i)} = \mathbf{v}_i$ and $Z_{i,i} = (\boldsymbol{\beta}_i - 3\mathbf{v}_i^{(2)})^\top \widetilde{\mathbf{Y}} + 2(\mathbf{v}_i^\top \widetilde{\mathbf{Y}})^2$. As a consequence, the asymptotic covariance matrices $\Sigma_{i,i}^{\mathbf{M}}$, $\Sigma_{i,i}^{\mathbf{M}'}$, $\Sigma_{i,i}^{\mathbf{V}}$, and $\Sigma_{i,i}^{\mathbf{V}'}$ of the estimators $\widehat{\boldsymbol{\mu}}_{m,i}$, $\widehat{\boldsymbol{\mu}}'_{m,i}$, $\widehat{\mathbf{v}}_{m,i}$, and $\widehat{\mathbf{v}}'_{m,i}$ depend only on the vectors \mathbf{v}_i and $\boldsymbol{\beta}_i$ but not on the mixed moments of the variables of any order.

3. Theoretical details and proofs

3.1. Some properties of the martingale differences

In this subsection we compute certain conditional expectations that are required to perform the CLS and the WCLS parameter estimations in the next subsection. All equations are understood for every $n = 1, 2, \dots$ and $i, i' = 1, \dots, p$, and in almost sure sense. Let $X_{n,i}, U_{n,i}$, and $V_{n,i}$ stand for the i -th component of the vector $\mathbb{X}_n, \mathbb{U}_n$, and \mathbb{V}_n , respectively. To shorten the notations we introduce the centered variables

$$\bar{\xi}_{i,j}(n, k) := \xi_{i,j}(n, k) - \mu_{i,j}, \quad \bar{\eta}_i(n) := \eta_i(n) - \mu_{i,\eta}, \quad i, j = 1, \dots, p, \quad n, k = 1, 2, \dots$$

The conditional expected value of $X_{n,i}$ with respect to \mathcal{F}_{n-1} is

$$E[X_{n,i} | \mathcal{F}_{n-1}] = \sum_{j=1}^p \sum_{k=1}^{X_{n-1,j}} E\xi_{i,j}(n, k) + E\eta_i(n) = \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1}, \tag{3}$$

meaning that $E[\mathbb{X}_n | \mathcal{F}_{n-1}] = \mathbf{M}\mathbb{Y}_{n-1}$, proving the first equation in (2). Similarly, applying the independence of the offspring and innovation variables, we get that

$$\begin{aligned} & E[U_{n,i}U_{n,i'} | \mathcal{F}_{n-1}] \\ &= E\left[\left(\sum_{j=1}^p \sum_{k=1}^{X_{n-1,j}} \bar{\xi}_{i,j}(n, k) + \bar{\eta}_i(n)\right)\left(\sum_{j=1}^p \sum_{k=1}^{X_{n-1,j}} \bar{\xi}_{i',j}(n, k) + \bar{\eta}_{i'}(n)\right) \middle| \mathbb{X}_{n-1}\right] \\ &= \sum_{j=1}^p X_{n-1,j} E(\bar{\xi}_{i,j}\bar{\xi}_{i',j}) + E(\bar{\eta}_i\bar{\eta}_{i'}) = \sum_{j=1}^p X_{n-1,j} v_{(i,i'),j} + v_{(i,i'),\eta} \\ &= \mathbf{v}_{(i,i')}^\top \mathbb{Y}_{n-1}. \end{aligned} \tag{4}$$

In the case $i = i'$ equation (4) can be written in the form $E[U_{n,i}^2 | \mathcal{F}_{n-1}] = \mathbf{v}_i^\top \mathbb{Y}_{n-1}$ which implies $E[\mathbb{U}_n^{(2)} | \mathcal{F}_{n-1}] = \mathbf{V}\mathbb{Y}_{n-1}$ and the second identity of (2). Also, if the variables in (1) have finite third moments, then by the independence of these variables

$$\begin{aligned} E[U_{n,i}^3 | \mathcal{F}_{n-1}] &= E\left[\left(\sum_{j=1}^p \sum_{k=1}^{X_{n-1,j}} \bar{\xi}_{i,j}(n, k) + \bar{\eta}_i(n)\right)^3 \middle| \mathbb{X}_{n-1}\right] \\ &= \sum_{j=1}^p X_{n-1,j} E\bar{\xi}_{i,j}^3 + E\bar{\eta}_i^3 = \boldsymbol{\alpha}_i^\top \mathbb{Y}_{n-1}, \end{aligned} \tag{5}$$

leading to the equations $E[U_n^{(3)} | \mathcal{F}_{n-1}] = \mathbf{A} \mathbb{Y}_{n-1}$ and

$$\begin{aligned} & E[U_{n,i} V_{n,i} | \mathcal{F}_{n-1}] \\ &= E[U_{n,i} (U_{n,i}^2 - E[U_{n,i}^2 | \mathcal{F}_{n-1}]) | \mathcal{F}_{n-1}] \\ &= E[U_{n,i}^3 | \mathcal{F}_{n-1}] - E[U_{n,i} | \mathcal{F}_{n-1}] E[U_{n,i}^2 | \mathcal{F}_{n-1}] = \boldsymbol{\alpha}_i^\top \mathbb{Y}_{n-1}. \end{aligned} \tag{6}$$

Let us recall that $V_{n,i}$ is the i -th component of the random vector \mathbb{V}_n . Our last goal is to determine the conditional expected value of $U_n^{(4)}$ and $V_{n,i}^2$ under the assumption that the variables in (1) have finite fourth moments. Let $v_{i,j}$ and $v_{i,\eta}$ stand for the variances of $\xi_{i,j}$ and η_i , and similarly, let $\beta_{i,j}$ and $\beta_{i,\eta}$ denote the fourth central moments of these variables. With a calculation similar to the ones in (4) and (5) one can easily show that

$$\begin{aligned} & E[U_{n,i}^2 U_{n,i'}^2 | \mathcal{F}_{n-1}] \\ &= E\left[\left(\sum_{j=1}^p \sum_{k=1}^{X_{n-1,j}} \bar{\xi}_{i,j}(n, k) + \bar{\eta}_i(n)\right)^2 \left(\sum_{j=1}^p \sum_{k=1}^{X_{n-1,j}} \bar{\xi}_{i',j}(n, k) + \bar{\eta}_{i'}(n)\right)^2 \middle| \mathbb{X}_{n-1}\right] \\ &= \left[\sum_{j=1}^p X_{n-1,j} \beta_{(i,i'),j} + \beta_{(i,i'),\eta}\right] + \\ &\quad + \left[\sum_{j,j'=1}^p X_{n-1,j} X_{n-1,j'} v_{i,j} v_{i',j'} - \sum_{j=1}^p X_{n-1,j} v_{i,j} v_{i',j}\right] + \\ &\quad + 2 \left[\sum_{j,j'=1}^p X_{n-1,j} X_{n-1,j'} v_{(i,i'),j} v_{(i,i'),j'} - \sum_{j=1}^p X_{n-1,j} v_{(i,i'),j}^2\right] + \\ &\quad + \sum_{j=1}^p X_{n-1,j} [v_{i,j} v_{i',\eta} + v_{i,\eta} v_{i',j}] + 4 \sum_{j=1}^p X_{n-1,j} v_{(i,i'),j} v_{(i,i'),\eta} \\ &= \boldsymbol{\beta}_{(i,i')}^\top \mathbb{Y}_{n-1} + \mathbf{v}_i^\top \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \mathbf{v}_{i'} - (\mathbf{v}_i \circ \mathbf{v}_{i'})^\top \mathbb{Y}_{n-1} + 2(\mathbf{v}_{(i,i')}^\top \mathbb{Y}_{n-1})^2 - \\ &\quad - 2(\mathbf{v}_{(i,i')}^{(2)})^\top \mathbb{Y}_{n-1}. \end{aligned} \tag{7}$$

In the case $i = i'$ we have $\boldsymbol{\beta}_{(i,i')} = \boldsymbol{\beta}_i$ and $\mathbf{v}_{(i,i')} = \mathbf{v}_i$, and formula (7) implies the equation

$$E[U_{n,i}^4 | \mathcal{F}_{n-1}] = \boldsymbol{\beta}_i^\top \mathbb{Y}_{n-1} + 3(\mathbf{v}_i^\top \mathbb{Y}_{n-1})^2 - 3(\mathbf{v}_i^{(2)})^\top \mathbb{Y}_{n-1},$$

which implies the identities

$$\begin{aligned} E[U_n^{(4)} | \mathcal{F}_{n-1}] &= \mathbf{B} \mathbb{Y}_{n-1} + 3(\mathbf{V} \mathbb{Y}_{n-1})^{(2)} - 3\mathbf{V}^{(2)} \mathbb{Y}_{n-1}, \\ E[\mathbb{K}_n | \mathcal{F}_{n-1}] &= \mathbf{B} \mathbb{Y}_{n-1}. \end{aligned} \tag{8}$$

Also, from (7) it follows that

$$\begin{aligned}
 E[V_{n,i}V_{n,i'} | \mathcal{F}_{n-1}] &= E[(U_{n,i}^2 - E[U_{n,i}^2 | \mathcal{F}_{n-1}])(U_{n,i'}^2 - E[U_{n,i'}^2 | \mathcal{F}_{n-1}] | \mathcal{F}_{n-1}] \\
 &= E[U_{n,i}^2 U_{n,i'}^2 | \mathcal{F}_{n-1}] - E[U_{n,i}^2 | \mathcal{F}_{n-1}] E[U_{n,i'}^2 | \mathcal{F}_{n-1}] \\
 &= \boldsymbol{\beta}_{(i,i')}^\top \mathbb{Y}_{n-1} - (\mathbf{v}_i \circ \mathbf{v}_{i'})^\top \mathbb{Y}_{n-1} + 2(\mathbf{v}_{(i,i')}^\top \mathbb{Y}_{n-1})^2 - 2(\mathbf{v}_{(i,i')}^{(2)})^\top \mathbb{Y}_{n-1},
 \end{aligned} \tag{9}$$

leading to the equation

$$E[V_{n,i}^2 | \mathcal{F}_{n-1}] = \boldsymbol{\beta}_i^\top \mathbb{Y}_{n-1} - 3(\mathbf{v}_i^{(2)})^\top \mathbb{Y}_{n-1} + 2(\mathbf{v}_i^\top \mathbb{Y}_{n-1})^2. \tag{10}$$

3.2. Parameter estimations

In this subsection we calculate the Conditional Least Squares (CLS) and the Weighted Conditional Least Squares (WCLS) estimators of the first four moments of the offspring and the innovation distributions. Since in a multitype Galton–Watson model for a given $r \in \mathbb{Z}_+$ the conditional expectation $E[\mathbb{U}_n^{(r)} | \mathcal{F}_{n-1}]$ can be expressed as a function of \mathbb{Y}_{n-1} and the moments of the variables in (1) of order at most r , we can estimate the r -th moments based on the sample $\mathbb{X}_0, \dots, \mathbb{X}_m$ by minimizing the sum

$$Q_r^2 := \frac{1}{2} \sum_{n=1}^m \|\mathbb{U}_n^{(r)} - E[\mathbb{U}_n^{(r)} | \mathcal{F}_{n-1}]\|^2.$$

Unfortunately, in most cases the minimum point depends not only on the sample but also on certain lower moments of the offspring and the innovation variables. To solve this problem we estimate the moments recursively and we replace the lower moments in the formulae by their estimators which are already obtained.

To get the estimator of \mathbf{M} we minimize the sum

$$Q_1^2 = \frac{1}{2} \sum_{n=1}^m \|\mathbb{U}_n\|^2 = \frac{1}{2} \sum_{n=1}^m [\mathbb{X}_n - \mathbf{M}\mathbb{Y}_{n-1}]^\top [\mathbb{X}_n - \mathbf{M}\mathbb{Y}_{n-1}].$$

Let $\nabla_{\mathbf{M}} Q_1^2$ denote the matrix of the partial derivatives of Q_1^2 with respect to the components of \mathbf{M} . Then, the CLS estimator $\widehat{\mathbf{M}}_m$ can be obtained as the solution of the normal equation

$$\mathbf{0} = \nabla_{\mathbf{M}} Q_1^2 = \sum_{n=1}^m [\mathbb{X}_n - \mathbf{M}\mathbb{Y}_{n-1}] \mathbb{Y}_{n-1}^\top$$

leading to

$$\widehat{\mathbf{M}}_m = \left[\sum_{n=1}^m \mathbb{X}_n \mathbb{Y}_{n-1}^\top \right] \left[\sum_{n=1}^m \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \right]^{-1}.$$

We must note that for some realizations of the sample $\mathbb{X}_0, \dots, \mathbb{X}_m$ the estimator $\widehat{\mathbf{M}}_m$ may not exist because the matrix $\sum_{n=1}^m \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top$ may not be invertible. However, we show in the next subsection that $\widehat{\mathbf{M}}_m$ is well defined with asymptotic probability 1 as $m \rightarrow \infty$, and the same statement is true for every estimator introduced in this subsection.

Similarly, we define the estimator of \mathbf{V} as the matrix that minimizes the function

$$Q_2^2 = \frac{1}{2} \sum_{n=1}^m \|\mathbb{U}_n^{(2)} - E[\mathbb{U}_n^{(2)} | \mathcal{F}_{n-1}]\|^2 = \frac{1}{2} \sum_{n=1}^m [\mathbb{U}_n^{(2)} - \mathbf{V}\mathbb{Y}_{n-1}]^\top [\mathbb{U}_n^{(2)} - \mathbf{V}\mathbb{Y}_{n-1}],$$

and by differentiating Q_2^2 with respect to \mathbf{V} we get that

$$\mathbf{0} = \nabla_{\mathbf{V}} Q_2^2 = \sum_{n=1}^m [\mathbb{U}_n^{(2)} - \mathbf{V}\mathbb{Y}_{n-1}] \mathbb{Y}_{n-1}^\top.$$

After solving this normal equation with respect to \mathbf{V} and replacing the theoretical martingale difference $\mathbb{U}_n = \mathbb{X}_n - \mathbf{M}\mathbb{Y}_{n-1}$ with its natural CLS estimator $\widehat{\mathbb{U}}_{m,n} = \mathbb{X}_n - \widehat{\mathbf{M}}_m \mathbb{Y}_{n-1}$ we get that

$$\widehat{\mathbf{V}}_m = \left[\sum_{n=1}^m \widehat{\mathbb{U}}_{m,n}^{(2)} \mathbb{Y}_{n-1}^\top \right] \left[\sum_{n=1}^m \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \right]^{-1}. \tag{11}$$

The formula for $\widehat{\mathbf{A}}_m$, the CLS estimator of \mathbf{A} , follows similarly if we minimize

$$Q_3^2 = \frac{1}{2} \sum_{n=1}^m \|\mathbb{U}_n^{(3)} - E[\mathbb{U}_n^{(3)} | \mathcal{F}_{n-1}]\|^2 = \frac{1}{2} \sum_{n=1}^m [\mathbb{U}_n^{(3)} - \mathbf{A}\mathbb{Y}_{n-1}]^\top [\mathbb{U}_n^{(3)} - \mathbf{A}\mathbb{Y}_{n-1}].$$

By solving the normal equation $\nabla_{\mathbf{A}} Q_3^2 = \mathbf{0}$ and by replacing $\mathbb{U}_n^{(3)}$ with $\widehat{\mathbb{U}}_{m,n}^{(3)}$ we obtain

$$\widehat{\mathbf{A}}_m = \left[\sum_{n=1}^m \widehat{\mathbb{U}}_{m,n}^{(3)} \mathbb{Y}_{n-1}^\top \right] \left[\sum_{n=1}^m \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \right]^{-1}.$$

Finally, to determine the CLS estimator of \mathbf{B} , we minimize the sum

$$Q_4^2 = \frac{1}{2} \sum_{n=1}^m \|\mathbb{U}_n^{(4)} - E[\mathbb{U}_n^{(4)} | \mathcal{F}_{n-1}]\|^2 = \frac{1}{2} \sum_{n=1}^m [\mathbb{K}_n - \mathbf{B}\mathbb{Y}_{n-1}]^\top [\mathbb{K}_n - \mathbf{B}\mathbb{Y}_{n-1}].$$

Again, by solving $\nabla_{\mathbf{B}} Q_4^2 = \mathbf{0}$ and by replacing \mathbb{K}_n with its CLS estimator $\widehat{\mathbb{K}}_{m,n}$ defined in Subsection 2.2 we get that

$$\widehat{\mathbf{B}}_m = \left[\sum_{n=1}^m \widehat{\mathbb{K}}_{m,n} \mathbb{Y}_{n-1}^\top \right] \left[\sum_{n=1}^m \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \right]^{-1}.$$

The WCLS estimation is defined as the CLS estimation based on the weighted process $\mathbb{X}'_n = \mathbb{X}_n / \sqrt{\mathbf{1}^\top \mathbb{Y}_{n-1}}$, $n = 1, 2, \dots$. That is, by introducing the variables

$$\mathbb{U}'_n := \mathbb{X}'_n - E[\mathbb{X}'_n \mid \mathbb{X}_{n-1}] = \frac{\mathbb{U}_n}{\sqrt{\mathbf{1}^\top \mathbb{Y}_{n-1}}}, \quad \mathbb{K}'_n := \frac{\mathbb{K}_n}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^2}, \quad n = 1, 2, \dots, \tag{12}$$

the r -th moments of the offspring and the innovation distributions can be estimated by minimizing the expression

$$\frac{1}{2} \sum_{n=1}^m \|\mathbb{U}'^{(r)}_n - E[\mathbb{U}'^{(r)}_n \mid \mathcal{F}_{n-1}]\|^2.$$

Using (12) the conditional expectation of $\mathbb{U}'^{(r)}_n$ can be calculated based on the results of Subsection 3.1. Since the WCLS estimators can be obtained by similar calculations as the unweighted ones we omit the details. The results are presented in Subsection 2.2.

Remark 3.1. Let us note that the sum Q_1^2 can be expressed as

$$Q_1^2 = \frac{1}{2} \sum_{n=1}^m \sum_{i=1}^p U_{n,i}^2 = \frac{1}{2} \sum_{i=1}^p \sum_{n=1}^m (X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1})^2,$$

which means that the CLS estimators of the rows of the matrix \mathbf{M} can be computed independently by minimizing the sums

$$\frac{1}{2} \sum_{n=1}^m (X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1})^2, \quad i = 1, \dots, p.$$

This remark is true not only for the first moments but for the higher ones as well, in both the CLS and the WCLS case. Let $\widehat{U}_{m,n,i}$ and $\widehat{K}_{m,n,i}$ stand for the i -th component of the vector $\widehat{\mathbb{U}}_{m,n}$ and $\widehat{\mathbb{K}}_{m,n}$, respectively, and similarly, let $\widehat{U}'_{m,n,i}$ and $\widehat{K}'_{m,n,i}$ denote the components of the weighted processes. Then the transposes of the i -th rows of the moment matrices can be estimated by

$$\begin{aligned} \widehat{\boldsymbol{\mu}}_{m,i} &= \mathbb{I}_{m,0}^{-1} \left[\sum_{n=1}^m X_{n,i} \mathbb{Y}_{n-1} \right], & \widehat{\boldsymbol{\mu}}'_{m,i} &= \mathbb{I}_{m,1}^{-1} \left[\sum_{n=1}^m \frac{X_{n,i} \mathbb{Y}_{n-1}}{\mathbf{1}^\top \mathbb{Y}_{n-1}} \right], \\ \widehat{\mathbf{v}}_{m,i} &= \mathbb{I}_{m,0}^{-1} \left[\sum_{n=1}^m \widehat{U}_{m,n,i}^2 \mathbb{Y}_{n-1} \right], & \widehat{\mathbf{v}}'_{m,i} &= \mathbb{I}_{m,2}^{-1} \left[\sum_{n=1}^m \frac{\widehat{U}_{m,n,i}^2 \mathbb{Y}_{n-1}}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^2} \right], \end{aligned}$$

$$\begin{aligned} \widehat{\alpha}_{m,i} &= \mathbb{I}_{m,0}^{-1} \left[\sum_{n=1}^m \widehat{U}_{m,n,i}^3 \mathbb{Y}_{n-1} \right], & \widehat{\alpha}'_{m,i} &= \mathbb{I}_{m,3}^{-1} \left[\sum_{n=1}^m \frac{\widehat{U}_{m,n,i}^3 \mathbb{Y}_{n-1}}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^3} \right], \\ \widehat{\beta}_{m,i} &= \mathbb{I}_{m,0}^{-1} \left[\sum_{n=1}^m \widehat{K}_{m,n,i} \mathbb{Y}_{n-1} \right], & \widehat{\beta}'_{m,i} &= \mathbb{I}_{m,4}^{-1} \left[\sum_{n=1}^m \frac{\widehat{K}'_{m,n,i} \mathbb{Y}_{n-1}}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^4} \right]. \end{aligned}$$

3.3. Proof of the existence and the strong consistency

To prove Theorem 2.1 we need some ergodic properties of the process $\mathbb{X}_n, n \in \mathbb{Z}_+$. By Theorem 1 of [14], if a multitype Galton–Watson process is subcritical, then it is ergodic in the sense that it has a unique invariant distribution concentrated on a positive recurrent class that the process reaches within finitely many steps with probability 1 in case of any initial distribution. Also, Theorem 3 of the same paper states that if all the random variables in (1) have finite r -th moments for some positive real number r , then so does the invariant distribution. It can be shown by standard methods that the Markov chain $(\mathbb{Y}_{n-1}, \mathbb{X}_n), n = 1, 2, \dots$, inherits these ergodic properties from the Galton–Watson process. Let us consider an arbitrary vector variable $(\widetilde{\mathbb{Y}}, \widetilde{\mathbb{X}})$ having the same distribution as $(\mathbb{Y}_0, \mathbb{X}_1)$ has under the invariant distribution. Let \widetilde{X}_i denote the i -th component of $\widetilde{\mathbb{X}}, i = 1, \dots, p$. Then by applying (3) we obtain the equation

$$E[\widetilde{\mathbb{X}} \mid \widetilde{\mathbb{Y}}] = \mathbf{M}\widetilde{\mathbb{Y}}. \tag{13}$$

Also, by ergodicity for any function $h : \mathbb{Z}_+^{p+1} \times \mathbb{Z}_+^p \rightarrow \mathbb{R}$, we have

$$\frac{1}{m} \sum_{n=1}^m h(\mathbb{Y}_{n-1}, \mathbb{X}_n) \rightarrow E(h(\widetilde{\mathbb{Y}}, \widetilde{\mathbb{X}})), \quad m \rightarrow \infty, \tag{14}$$

almost surely if the expectation is finite. By the ergodic property

$$\frac{\mathbb{I}_{m,r}}{m} = \frac{1}{m} \sum_{n=1}^m \frac{\mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^r} \rightarrow \widetilde{\mathbb{I}}_r := E \left[\frac{\widetilde{\mathbb{Y}} \widetilde{\mathbb{Y}}^\top}{(\mathbf{1}^\top \widetilde{\mathbb{Y}})^r} \right], \quad m \rightarrow \infty, \tag{15}$$

almost surely for any $r \in \mathbb{Z}_+$. The limit is finite as by the definition of $\widetilde{\mathbb{Y}}$ it holds that $1 \leq \mathbf{1}^\top \widetilde{\mathbb{Y}}$, and as the existence of the second moments of the variables in (1) is assumed, the variable $\widetilde{\mathbb{Y}}$ has finite second moment as well.

Proposition 3.2. *Assume that the process $\mathbb{X}_n, n = 0, 1, \dots$, is subcritical and the variables in (1) have finite second moments. Then the matrix $E(\widetilde{\mathbb{Y}} \widetilde{\mathbb{Y}}^\top)$ is invertible if and only if (iii) and (iv) of Assumption 1 hold.*

Proof. Theorem 2 of [14] states that the components of $\tilde{\mathbf{X}}$ are linearly independent if and only if (iii) and (iv) of Assumption 1 hold. Therefore we only have to show that the matrix $E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top)$ is singular exactly if the components of $\tilde{\mathbf{X}}$ are linearly dependent, meaning that there exists a vector $\mathbf{c} \in \mathbb{R}^p$, $\mathbf{c} \neq \mathbf{0}$, satisfying $\mathbf{c}^\top(\tilde{\mathbf{X}} - E(\tilde{\mathbf{X}})) = 0$ with probability 1.

Since the matrix $E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top)$ is positive semidefinite, it is singular exactly if there exists a vector

$$\mathbf{d} = \begin{bmatrix} \mathbf{c} \\ c' \end{bmatrix} \in \mathbb{R}^p \times \mathbb{R} = \mathbb{R}^{p+1}, \quad \mathbf{d} \neq \mathbf{0}, \tag{16}$$

such that $0 = \mathbf{d}^\top E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top)\mathbf{d} = E(\mathbf{d}^\top \tilde{\mathbf{Y}})^2$, which holds if and only if $\mathbf{d}^\top \tilde{\mathbf{Y}} = 0$ almost surely. That is, if $E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top)$ is singular then $\mathbf{c}^\top E(\tilde{\mathbf{X}}) + c' = 0$, therefore $\mathbf{c}^\top(\tilde{\mathbf{X}} - E(\tilde{\mathbf{X}})) = 0$. Let us note that currently $\mathbf{c} \neq \mathbf{0}$, since $\mathbf{c} = \mathbf{0}$ implies that $c' = 0$ and $\mathbf{d} = \mathbf{0}$. This means that the components of $\tilde{\mathbf{X}}$ are linearly dependent. For the contrary direction, assume that the components of $\tilde{\mathbf{X}}$ are linearly dependent with some vector $\mathbf{c} \neq \mathbf{0}$. Then with $c' = -\mathbf{c}^\top E(\tilde{\mathbf{X}})$ and with the vector $\mathbf{d} \neq \mathbf{0}$ defined in (16) it holds that

$$\mathbf{d}^\top \tilde{\mathbf{Y}} = \mathbf{c}^\top \tilde{\mathbf{X}} + c' = \mathbf{c}^\top(\tilde{\mathbf{X}} - E(\tilde{\mathbf{X}})) = 0$$

with probability 1. This means that $E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top)$ is not positive definite implying that the matrix is singular. ■

Proof of Theorem 2.1. (i) The CLS and the WCLS estimators of the moments based on the sample $\mathbb{X}_0, \dots, \mathbb{X}_m$ are well defined if and only if the matrices $\mathbb{I}_{m,r}$, $r = 0, \dots, 4$, introduced in Subsection 2.1 are invertible. That is, to prove the statement, it is enough to show that these matrices are nonsingular with asymptotic probability 1 as $m \rightarrow \infty$. By Proposition 3.2 under Assumption 1 the matrix $\tilde{\mathbb{I}}_0 = E(\tilde{\mathbf{Y}}\tilde{\mathbf{Y}}^\top)$ is nonsingular and by (15) we have that $\mathbb{I}_{m,0}/m \rightarrow \tilde{\mathbb{I}}_0$ almost surely as $m \rightarrow \infty$, implying that $\mathbb{I}_{m,0}$ is nonsingular with asymptotic probability 1.

Since for any r and m the matrix $\mathbb{I}_{m,r}$ is positive semidefinite, it is invertible if and only if it is a positive definite matrix. Also note that $\mathbb{I}_{m,r}$ is defined as a finite sum of positive semidefinite matrices. Then, $\mathbb{I}_{m,r}$ is positive definite if and only if any of its terms is positive definite. We already know that $\mathbb{I}_{m,0}$ is positive definite with asymptotic probability 1. This implies that with probability 1 at least one of the terms $\mathbb{Y}_{n-1}\mathbb{Y}_{n-1}^\top$, $n \in \mathbb{Z}_+$, is positive definite, meaning that for any $r \in \mathbb{Z}_+$ at least one of the terms $\mathbb{Y}_{n-1}\mathbb{Y}_{n-1}^\top/(\mathbf{1}^\top \mathbb{Y}_{n-1})^r$, $n \in \mathbb{Z}_+$, is positive definite. From this we immediately get that for any $r \in \mathbb{Z}_+$ the matrix $\mathbb{I}_{m,r}$ is positive definite with asymptotic probability 1 as $m \rightarrow \infty$. This completes the proof of the existence of the estimators.

(ii)–(iii) Since all of the estimators are represented as products of two matrices, to prove their convergence it is enough to show that the components are convergent in almost sure sense. As $\tilde{\mathbb{I}}_r$ is nonsingular by part (i) of the current proof, it follows from (15) that $m\mathbb{I}_{m,r}^{-1} \rightarrow \tilde{\mathbb{I}}_r^{-1}$ as $m \rightarrow \infty$ for any $r \in \mathbb{Z}_+$ without any additional moment condition on the offspring and the innovation variables. Then by applying (14) and (13) we get that

$$\begin{aligned} \widehat{\mathbf{M}}_m &= \left[\frac{1}{m} \sum_{n=1}^m \mathbb{X}_n \mathbb{Y}_{n-1}^\top \right] (m\mathbb{I}_{m,0}^{-1}) \rightarrow E(\tilde{\mathbb{X}}\tilde{\mathbb{Y}}^\top) \tilde{\mathbb{I}}_0^{-1} = E\left(E[\tilde{\mathbb{X}} | \tilde{\mathbb{Y}}] \tilde{\mathbb{Y}}^\top\right) \tilde{\mathbb{I}}_0^{-1} \\ &= E(\mathbf{M}\tilde{\mathbb{Y}}\tilde{\mathbb{Y}}^\top) [E(\mathbb{Y}\tilde{\mathbb{Y}}^\top)]^{-1} = \mathbf{M}, \quad m \rightarrow \infty, \end{aligned}$$

almost surely, proving the strong consistency of the CLS estimator of \mathbf{M} . Similarly,

$$\widehat{\mathbf{V}}_m = \left[\frac{1}{m} \sum_{n=1}^m \widehat{\mathbb{U}}_{m,n}^{(2)} \mathbb{Y}_{n-1}^\top \right] (m\mathbb{I}_{m,0}^{-1}) \rightarrow E(\mathbf{V}\tilde{\mathbb{Y}}\tilde{\mathbb{Y}}^\top) [E(\mathbb{Y}\tilde{\mathbb{Y}}^\top)]^{-1} = \mathbf{V}, \quad m \rightarrow \infty,$$

since we have

$$\begin{aligned} &\frac{1}{m} \sum_{n=1}^m \widehat{U}_{m,n,i}^2 \mathbb{Y}_{n-1}^\top \\ &= \frac{1}{m} \sum_{n=1}^m \left[(X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1}) + (\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1} \right]^2 \mathbb{Y}_{n-1}^\top \\ &= \frac{1}{m} \sum_{n=1}^m (X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1})^2 \mathbb{Y}_{n-1}^\top + \\ &\quad + \frac{2}{m} \sum_{n=1}^m (X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1})(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top + \\ &\quad + \frac{1}{m} \sum_{n=1}^m ((\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1})^2 \mathbb{Y}_{n-1}^\top \rightarrow E\left[(\tilde{X}_i - \boldsymbol{\mu}_i^\top \tilde{\mathbb{Y}})^2 \tilde{\mathbb{Y}}^\top\right] + 0 + 0 \\ &= E\left(E\left[(\tilde{X}_i - \boldsymbol{\mu}_i^\top \tilde{\mathbb{Y}})^2 | \tilde{\mathbb{Y}}\right] \tilde{\mathbb{Y}}^\top\right) = E(\mathbf{v}_i^\top \tilde{\mathbb{Y}}\tilde{\mathbb{Y}}^\top), \quad i = 1, \dots, p, \end{aligned}$$

almost surely, where in the last steps we used equation (4), the definition of $(\tilde{\mathbb{Y}}, \tilde{\mathbb{X}})$, and the following two arguments. First, with the Frobenius matrix norm $\|\mathbf{C}\|_F := (\sum_{i,j=1}^{p+1} c_{i,j}^2)^{1/2}$, $\mathbf{C} = (c_{i,j})_{i,j=1,\dots,p+1} \in \mathbb{R}^{(p+1) \times (p+1)}$, we get that

$$\begin{aligned} &\left\| \frac{2}{m} \sum_{n=1}^m (X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1})(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \right\| \\ &\leq \|\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i}\| \left\| \frac{2}{m} \sum_{n=1}^m (X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1}) \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \right\|_F \rightarrow 0, \quad m \rightarrow \infty, \end{aligned}$$

almost surely, as the first term tends to zero almost surely by the strong consistency of $\hat{\boldsymbol{\mu}}_{m,i}$, while the second term converges to the finite limit $\|2E(\tilde{X}_i - \boldsymbol{\mu}_i^\top \tilde{Y})\tilde{Y}\tilde{Y}^\top\|_F$ almost surely by ergodicity. Similarly,

$$\begin{aligned} & \left\| \frac{1}{m} \sum_{n=1}^m ((\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1})^2 \mathbb{Y}_{n-1}^\top \right\| \\ & \leq \frac{1}{m} \sum_{n=1}^m ((\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1})^2 \|\mathbb{Y}_{n-1}\| \leq \|\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_{m,i}\|^2 \frac{1}{m} \sum_{n=1}^m \|\mathbb{Y}_{n-1}\|^3, \end{aligned}$$

which tends to zero almost surely since the first part converges to zero again, and the second tends to the finite limit $E(\|\tilde{Y}\|^3)$ by the ergodic theorem.

We do not detail the proofs of the consistency of the remaining estimators as they only require the standard steps presented above. However, it is important to note that the convergence of the WCLS estimators requires lower moment conditions than the CLS ones, because the weights reduce the degrees of the variables in the limit expectations. ■

3.4. Proof of the asymptotic normality

Our first proposition is a technical statement.

Proposition 3.3. *For any $\gamma \geq 1$ and $n = 1, 2, \dots$ the following inequalities hold:*

- (i)
$$E[\|\mathbb{X}_n\|^\gamma \mid \mathbb{X}_{n-1}] \leq (p + 1)^\gamma \|\mathbb{Y}_{n-1}\|^\gamma M_\gamma,$$
- (ii)
$$E[\|\mathbb{U}_n\|^\gamma \mid \mathbb{X}_{n-1}] \leq (p + 1)^\gamma \|\mathbb{Y}_{n-1}\|^\gamma C_\gamma,$$
- (iii)
$$E[\|\mathbb{V}_n\|^\gamma \mid \mathbb{X}_{n-1}] \leq 2^{\gamma+1} p^\gamma (p + 1)^{2\gamma} \|\mathbb{Y}_{n-1}\|^{2\gamma} C_{2\gamma},$$

where $M_\gamma = \max\{E\|\boldsymbol{\xi}_1\|^\gamma, \dots, E\|\boldsymbol{\xi}_p\|^\gamma, E\|\boldsymbol{\eta}\|^\gamma\}$ and

$$C_\gamma = \max\left\{E\|\boldsymbol{\xi}_1 - E\boldsymbol{\xi}_1\|^\gamma, \dots, E\|\boldsymbol{\xi}_p - E\boldsymbol{\xi}_p\|^\gamma, E\|\boldsymbol{\eta} - E\boldsymbol{\eta}\|^\gamma\right\}.$$

Proof. For the shorter notation let $\|\cdot\|_\gamma$ stand for the L^γ norm of the random vectors. Consider any $n \in \mathbb{Z}_+$ and $\mathbf{x} = [x_1, \dots, x_p]^\top \in \mathbb{Z}_+^p$, and introduce $\mathbf{y} = [x_1, \dots, x_p, 1]^\top \in \mathbb{Z}_+^{p+1}$. Applying the Minkowski inequality we get that

$$\begin{aligned} (E[\|\mathbb{X}_n\|^\gamma \mid \mathbb{X}_{n-1} = \mathbf{x}])^{1/\gamma} &= \left\| \sum_{i=1}^p \sum_{k=1}^{x_i} \boldsymbol{\xi}_i(n, k) + \boldsymbol{\eta}(n) \right\|_\gamma \\ &\leq \sum_{i=1}^p \sum_{k=1}^{x_i} \|\boldsymbol{\xi}_i(n, k)\|_\gamma + \|\boldsymbol{\eta}(n)\|_\gamma \leq (x_1 + \dots + x_p + 1) M_\gamma^{1/\gamma} \leq (p + 1) \|\mathbf{y}\| M_\gamma^{1/\gamma}, \end{aligned}$$

which immediately implies the first statement. The proof of the second inequality is analogous since

$$(E[\|\mathbb{U}_n\|^\gamma \mid \mathbb{X}_{n-1} = \mathbf{x}])^{1/\gamma} = \left\| \sum_{i=1}^p \sum_{k=1}^{x_i} [\boldsymbol{\xi}_i(n, k) - E\boldsymbol{\xi}_i] + [\boldsymbol{\eta}(n) - E\boldsymbol{\eta}] \right\|_\gamma.$$

For the last statement let us note that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ we have $\|\mathbf{x}^{(2)}\| \leq \|\mathbf{x}\|^\gamma$ and

$$\|\mathbf{x} + \mathbf{y}\|^\gamma \leq \left[\sum_{i=1}^p (|x_i| + |y_i|) \right]^\gamma \leq (2p \max_{1 \leq i \leq p} \{|x_i|, |y_i|\})^\gamma \leq 2^\gamma p^\gamma (\|\mathbf{x}\|^\gamma + \|\mathbf{y}\|^\gamma).$$

Then it follows that

$$\begin{aligned} E[\|\mathbb{V}_n\|^\gamma \mid \mathbb{X}_{n-1}] &= E[\|\mathbb{U}_n^{(2)} - E[\mathbb{U}_n^{(2)} \mid \mathbb{X}_{n-1}]\|^\gamma \mid \mathbb{X}_{n-1}] \\ &\leq 2^\gamma p^\gamma (E[\|\mathbb{U}_n^{(2)}\|^\gamma \mid \mathbb{X}_{n-1}] + \|E[\mathbb{U}_n^{(2)} \mid \mathbb{X}_{n-1}]\|^\gamma) \leq 2^{\gamma+1} p^\gamma E[\|\mathbb{U}_n\|^{2\gamma} \mid \mathbb{X}_{n-1}], \end{aligned}$$

and inequality (ii) of the current proposition completes the proof. ■

Proof of Theorem 2.2. At several points of our proof we will apply the multi-dimensional martingale central limit theorem (MCLT). For reference see, e.g., [7, Chapter VIII, Theorem 3.33].

(i) Introduce the processes

$$\mathbb{S}_{m,i} := \frac{1}{\sqrt{m}} \sum_{n=1}^m U_{n,i} Y_{n-1}, \quad \mathbb{T}_n := \begin{bmatrix} U_{n,1} Y_{n-1} \\ \vdots \\ U_{n,p} Y_{n-1} \end{bmatrix}, \quad \mathbb{S}_m := \frac{1}{\sqrt{m}} \sum_{n=1}^m \mathbb{T}_n = \begin{bmatrix} \mathbb{S}_{m,1} \\ \vdots \\ \mathbb{S}_{m,p} \end{bmatrix},$$

where $n, m = 1, 2, \dots$ and $i = 1, \dots, p$. Note that we have $\|\mathbb{T}_n\| = \|\mathbb{U}_n\| \|Y_{n-1}\|$ because

$$\|\mathbb{T}_n\|^2 = \sum_{i=1}^p U_{n,i}^2 \|Y_{n-1}\|^2 = \|\mathbb{U}_n\|^2 \|Y_{n-1}\|^2, \quad n \in \mathbb{Z}_+. \tag{17}$$

From the identity $X_{n,i} - Y_{n-1}^\top \boldsymbol{\mu}_i = U_{n,i}$ it follows that

$$\sqrt{m} (\widehat{\boldsymbol{\mu}}_{m,i} - \boldsymbol{\mu}_i) = \sqrt{m} \mathbb{I}_{m,0}^{-1} \sum_{n=1}^m Y_{n-1} (X_{n,i} - Y_{n-1}^\top \boldsymbol{\mu}_i) = m \mathbb{I}_{m,0}^{-1} \mathbb{S}_{m,i}, \quad i = 1, \dots, p,$$

where $\mathbb{I}_{m,0}/m \rightarrow E[\widetilde{Y}\widetilde{Y}^\top]$ almost surely as $m \rightarrow \infty$ by formula (15). As a result, to prove the joint asymptotic normality of the estimators $\widehat{\boldsymbol{\mu}}_{m,1}, \dots, \widehat{\boldsymbol{\mu}}_{m,p}$, it is

enough to show the weak convergence of the process $\mathbb{S}_1, \mathbb{S}_2, \dots$ to a proper normal distribution by using the MCLT.

Let us recall that under the assumptions the $(4 + \varepsilon)$ -th moment of \tilde{Y} is finite by Theorem 3 of [14]. It is obvious that the random vectors $\mathbb{T}_1, \mathbb{T}_2, \dots$ form a martingale difference sequence with respect to the filtration $\mathcal{F}_0, \mathcal{F}_1, \dots$. First, we show that the Lindeberg condition is satisfied. By using ergodicity and statement (ii) of Proposition 3.3 we obtain the almost sure convergence

$$\begin{aligned} & \sum_{n=1}^m E \left[\left\| \frac{1}{\sqrt{m}} \mathbb{T}_n \right\|^2 \mathbb{1}_{\{\|\mathbb{T}_n/\sqrt{m}\| > \delta\}} \mid \mathcal{F}_{n-1} \right] \\ & \leq \frac{1}{\delta^{\varepsilon/2}} \sum_{n=1}^m E \left[\left\| \frac{1}{\sqrt{m}} \mathbb{T}_n \right\|^{2+\varepsilon/2} \mid \mathcal{F}_{n-1} \right] \\ & = \frac{1}{\delta^{\varepsilon/2} m^{1+\varepsilon/4}} \sum_{n=1}^m E \left[\|\mathbb{U}_n\|^{2+\varepsilon/2} \mid \mathbb{X}_{n-1} \right] \|\mathbb{Y}_{n-1}\|^{2+\varepsilon/2} \\ & \leq \frac{(p+1)^{2+\varepsilon/2}}{\delta^{\varepsilon/2} m^{1+\varepsilon/4}} \sum_{n=1}^m \|\mathbb{Y}_{n-1}\|^{4+\varepsilon} C_{2+\varepsilon/2} \rightarrow 0, \quad m \rightarrow \infty. \end{aligned} \tag{18}$$

By equation (4) and ergodicity we also have

$$\begin{aligned} & \frac{1}{m} \sum_{n=1}^m E \left[(U_{n,i} \mathbb{Y}_{n-1})(U_{n,j} \mathbb{Y}_{n-1})^\top \mid \mathcal{F}_{n-1} \right] \\ & = \frac{1}{m} \sum_{n=1}^m E [U_{n,i} U_{n,j} \mid \mathcal{F}_{n-1}] \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \\ & = \frac{1}{m} \sum_{n=1}^m (\mathbf{v}_{(i,j)}^\top \mathbb{Y}_{n-1}) \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \rightarrow \boldsymbol{\Sigma}_{i,j} := E \left[(\mathbf{v}_{(i,j)}^\top \tilde{Y}) \tilde{Y} \tilde{Y}^\top \right], \quad m \rightarrow \infty, \end{aligned} \tag{19}$$

with probability 1 for $i, j = 1, \dots, p$. Then, the MCLT implies the convergence

$$\mathbb{S}_m \xrightarrow{\mathcal{D}} \mathcal{N}_{p(p+1)} \left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \cdots & \boldsymbol{\Sigma}_{1,p} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{p,1} & \cdots & \boldsymbol{\Sigma}_{p,p} \end{bmatrix} \right), \quad m \rightarrow \infty,$$

proving the first statement in (i) of Theorem 2.2.

If the components of the offspring and innovation variables in (1) are pairwise uncorrelated, then $\mathbf{v}_{(i,j)} = \mathbf{0}$ and $\boldsymbol{\Sigma}_{i,j}^{\mathbf{M}} = \mathbf{0}$ by definition for any types $i \neq j$. From this the asymptotic independence of the estimators follows.

(ii) To show the asymptotic normality of the weighted estimators of the mean vectors let us introduce the weighted versions of the variables \mathbb{Y}_{n-1} , $\mathbb{S}_{m,i}$, \mathbb{T}_n , and \mathbb{S}_m in the form

$$\mathbb{Y}'_{n-1} := \frac{\mathbb{Y}_{n-1}}{\sqrt{\mathbf{1}^\top \mathbb{Y}_{n-1}}}, \quad \mathbb{S}'_{m,i} := \frac{1}{\sqrt{m}} \sum_{n=1}^m U'_{n,i} \mathbb{Y}'_{n-1}, \quad i = 1, \dots, p,$$

$$\mathbb{T}'_n := \frac{\mathbb{T}_n}{\mathbf{1}^\top \mathbb{Y}_{n-1}} = \begin{bmatrix} U'_{n,1} \mathbb{Y}'_{n-1} \\ \vdots \\ U'_{n,p} \mathbb{Y}'_{n-1} \end{bmatrix}, \quad \mathbb{S}'_m := \frac{1}{\sqrt{m}} \sum_{n=1}^m \mathbb{T}'_n = \begin{bmatrix} \mathbb{S}'_{m,1} \\ \vdots \\ \mathbb{S}'_{m,p} \end{bmatrix}, \quad n, m = 1, 2, \dots,$$

where $U'_{n,i}$ stands for the i th component of \mathbb{U}'_n , $i = 1, \dots, p$. By applying the identity $X_{n,i} - \mathbb{Y}_{n-1}^\top \boldsymbol{\mu}_i = U_{n,i}$ and the formula for the estimator $\hat{\boldsymbol{\mu}}'_{m,i}$ we get that

$$\sqrt{m}(\hat{\boldsymbol{\mu}}'_{m,i} - \boldsymbol{\mu}_i) = \sqrt{m} \mathbb{I}_{m,1}^{-1} \sum_{n=1}^m \frac{\mathbb{Y}_{n-1}(X_{n,i} - \mathbb{Y}_{n-1}^\top \boldsymbol{\mu}_i)}{\mathbf{1}^\top \mathbb{Y}_{n-1}} = m \mathbb{I}_{m,1}^{-1} \mathbb{S}'_{m,i}, \quad i = 1, \dots, p.$$

Since (15) implies the convergence $\mathbb{I}_{m,1}/m \rightarrow E[\widetilde{\mathbb{Y}}\widetilde{\mathbb{Y}}^\top / (\mathbf{1}^\top \widetilde{\mathbb{Y}})]$, $m \rightarrow \infty$, statement (ii) can be proven by showing the weak convergence of the process $\mathbb{S}'_1, \mathbb{S}'_2, \dots$

The assumptions of the current statement imply that the random vector $\widetilde{\mathbb{Y}}$ has finite $(2 + \varepsilon)$ -th moment with some $\varepsilon > 0$. Using this fact and the identity $\|\mathbb{T}_n\| = \|\mathbb{U}_n\| \|\mathbb{Y}_{n-1}\|$ obtained from equation (17) we get the convergence

$$\begin{aligned} & \sum_{n=1}^m E \left[\left\| \frac{1}{\sqrt{m}} \mathbb{T}'_n \right\|^2 \mathbb{1}_{\{\|\mathbb{T}'_n/\sqrt{m}\| > \delta\}} \mid \mathcal{F}_{n-1} \right] \\ & \leq \frac{1}{\delta^\varepsilon} \sum_{n=1}^m E \left[\left\| \frac{1}{\sqrt{m}} \mathbb{T}'_n \right\|^{2+\varepsilon} \mid \mathcal{F}_{n-1} \right] \\ & = \frac{1}{\delta^\varepsilon m^{1+\varepsilon/2}} \sum_{n=1}^m E \left[\|\mathbb{T}_n\|^{2+\varepsilon} \mid \mathcal{F}_{n-1} \right] \frac{1}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^{2+\varepsilon}} \\ & = \frac{1}{\delta^\varepsilon m^{1+\varepsilon/2}} \sum_{n=1}^m E \left[\|\mathbb{U}_n\|^{2+\varepsilon} \mid \mathbb{X}_{n-1} \right] \frac{\|\mathbb{Y}_{n-1}\|^{2+\varepsilon}}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^{2+\varepsilon}} \\ & \leq \frac{(p+1)^{2+\varepsilon}}{\delta^\varepsilon m^{1+\varepsilon/2}} \sum_{n=1}^m \frac{\|\mathbb{Y}_{n-1}\|^{4+2\varepsilon}}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^{2+\varepsilon}} C_{2+\varepsilon} \leq \frac{(p+1)^{2+\varepsilon}}{\delta^\varepsilon m^{1+\varepsilon/2}} \sum_{n=1}^m \|\mathbb{Y}_{n-1}\|^{2+\varepsilon} C_{2+\varepsilon} \rightarrow 0, \end{aligned}$$

as $m \rightarrow \infty$. This means that the martingale difference sequence $\mathbb{T}'_1, \mathbb{T}'_2, \dots$ satisfies the Lindeberg condition. Also, by the ergodicity of the Galton–Watson process, for

any types $i, j = 1, \dots, p$, it follows that

$$\begin{aligned} & \frac{1}{m} \sum_{n=1}^m E \left[(U'_{n,i} \mathbb{Y}'_{n-1}) (U'_{n,j} \mathbb{Y}'_{n-1})^\top \mid \mathcal{F}_{n-1} \right] \\ &= \frac{1}{m} \sum_{n=1}^m E \left[U_{n,i} U_{n,j} \mid \mathcal{F}_{n-1} \right] \frac{\mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^2} \\ &= \frac{1}{m} \sum_{n=1}^m \frac{(\mathbf{v}_{(i,j)}^\top \mathbb{Y}_{n-1}) \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^2} \rightarrow \boldsymbol{\Sigma}'_{i,j} := E \left[\frac{(\mathbf{v}_{(i,j)}^\top \tilde{\mathbb{Y}}) \tilde{\mathbb{Y}} \tilde{\mathbb{Y}}^\top}{(\mathbf{1}^\top \tilde{\mathbb{Y}})^2} \right], \quad m \rightarrow \infty. \end{aligned}$$

The finiteness of $\boldsymbol{\Sigma}'_{i,j}$ follows from the fact that the sum of the absolute values of the entries of the matrix has finite expectation as

$$\begin{aligned} E \left[\left| \mathbf{v}_{(i,j)}^\top \tilde{\mathbb{Y}} \mathbf{1}^\top \frac{\tilde{\mathbb{Y}} \tilde{\mathbb{Y}}^\top}{(\mathbf{1}^\top \tilde{\mathbb{Y}})^2} \mathbf{1} \right| \right] &\leq E \left[\left| \mathbf{v}_{(i,j)}^\top \tilde{\mathbb{Y}} \right| \right] \\ &\leq \|\mathbf{v}_{(i,j)}\| E(\|\tilde{\mathbb{Y}}\|) \leq \|\mathbf{v}_{(i,j)}\| (E(\|\tilde{\mathbb{Y}}\|^2))^{1/2} < \infty, \end{aligned}$$

where we applied the Cauchy–Schwarz and Lyapunov inequalities.

As a result, from the MCLT we obtain that

$$\boldsymbol{\Sigma}'_m \xrightarrow{D} \mathcal{N}_{p(p+1)} \left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}'_{1,1} & \cdots & \boldsymbol{\Sigma}'_{1,p} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}'_{p,1} & \cdots & \boldsymbol{\Sigma}'_{p,p} \end{bmatrix} \right), \quad m \rightarrow \infty,$$

which implies the joint asymptotic normality of $\hat{\boldsymbol{\mu}}'_{m,1}, \dots, \hat{\boldsymbol{\mu}}'_{m,p}$ with the covariance matrix presented in the statement. Similarly to the unweighted case, if the variables in (1) have uncorrelated components, then $\mathbf{v}_{(i,j)} = \mathbf{0}$ and $\boldsymbol{\Sigma}^{\mathbf{M}'}_{i,j} = \mathbf{0}$ for any $i \neq j$, proving the asymptotic independence of the estimators.

(iii) First, note that

$$\sqrt{m}(\hat{\mathbf{v}}_{m,i} - \mathbf{v}_i) = \sqrt{m} \mathbb{I}_{m,0}^{-1} \sum_{n=1}^m \mathbb{Y}_{n-1} (\hat{U}_{m,n,i}^2 - \mathbb{Y}_{n-1}^\top \mathbf{v}_i), \quad i = 1, \dots, p,$$

where

$$\begin{aligned} \hat{U}_{m,n,i}^2 - \mathbb{Y}_{n-1}^\top \mathbf{v}_i &= (X_{n,i} - \hat{\boldsymbol{\mu}}_{m,i}^\top \mathbb{Y}_{n-1})^2 - \mathbf{v}_i^\top \mathbb{Y}_{n-1} \\ &= (U_{n,i} + (\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1})^2 - \mathbf{v}_i^\top \mathbb{Y}_{n-1} \\ &= V_{n,i} + 2U_{n,i} (\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1} + ((\boldsymbol{\mu}_i - \hat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1})^2. \end{aligned}$$

Then, using ergodicity, the asymptotic normality of $\widehat{\boldsymbol{\mu}}_{m,i}$, and standard Slutsky’s argument

$$\begin{aligned} & \frac{1}{\sqrt{m}} \sum_{n=1}^m \mathbb{Y}_{n-1} U_{n,i} (\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1} \\ &= \left[\frac{1}{m} \sum_{n=1}^m (X_{n,i} - \boldsymbol{\mu}_i^\top \mathbb{Y}_{n-1}) \mathbb{Y}_{n-1} \mathbb{Y}_{n-1}^\top \right] \sqrt{m} (\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i}) = o_P(1), \quad m \rightarrow \infty, \end{aligned}$$

since the first term tends to 0 almost surely by the ergodic theorem because

$$E[(\widetilde{X}_i - \boldsymbol{\mu}_i^\top \widetilde{\mathbb{Y}}) \widetilde{\mathbb{Y}} \widetilde{\mathbb{Y}}^\top] = E[(E[\widetilde{X}_i | \widetilde{\mathbb{Y}}] - \boldsymbol{\mu}_i^\top \widetilde{\mathbb{Y}}) \widetilde{\mathbb{Y}} \widetilde{\mathbb{Y}}^\top] = 0.$$

Similarly, by the ergodic theorem

$$\begin{aligned} & \left\| \frac{1}{\sqrt{m}} \sum_{n=1}^m \mathbb{Y}_{n-1} ((\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1})^2 \right\| \\ & \leq \frac{1}{\sqrt{m}} \sum_{n=1}^m \|\mathbb{Y}_{n-1}\| ((\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})^\top \mathbb{Y}_{n-1})^2 \\ & \leq \|\sqrt{m}(\boldsymbol{\mu}_i - \widehat{\boldsymbol{\mu}}_{m,i})\|^2 \frac{1}{m^{3/2}} \sum_{n=1}^m \|\mathbb{Y}_{n-1}\|^3 = o_P(1), \end{aligned}$$

as $m \rightarrow \infty$. By summing up for $i = 1, \dots, p$, it follows that

$$\sqrt{m}(\widehat{\mathbf{v}}_{m,i} - \mathbf{v}_i) = (m\mathbb{I}_{m,0}^{-1}) \left[\frac{1}{\sqrt{m}} \sum_{n=1}^m V_{n,i} \mathbb{Y}_{n-1} + o_P(1) \right], \quad m \rightarrow \infty. \tag{20}$$

Since the first component of the product converges to $(E[\widetilde{\mathbb{Y}} \widetilde{\mathbb{Y}}^\top])^{-1}$ with probability 1, to finish the proof it is enough to show the weak convergence of the second component. This can be done by using the MCLT analogously as in the proof of statement (i) of the current theorem. Using (iii) of Proposition 3.3 and the finiteness of the $(6 + \varepsilon)$ -th moment of the invariant distribution, one can show that the Lindeberg condition is satisfied. Also, (9) implies the almost sure convergence

$$\frac{1}{m} \sum_{n=1}^m E[(V_{n,i} \mathbb{Y}_{n-1})(V_{n,j} \mathbb{Y}_{n-1})^\top | \mathcal{F}_{n-1}] \rightarrow E[Z_{i,j} \widetilde{\mathbb{Y}} \widetilde{\mathbb{Y}}^\top], \quad m \rightarrow \infty.$$

Then the joint asymptotic normality of the estimators $\widehat{\mathbf{v}}_1, \dots, \widehat{\mathbf{v}}_p$ follows from the MCLT.

If additionally the components of the offspring and innovation variables are pairwise independent, then $\boldsymbol{\beta}_{(i,j)} = \mathbf{v}_i \circ \mathbf{v}_j$ and $\mathbf{v}_{(i,j)} = \mathbf{0}$ for any $i \neq j$. This means

that $Z_{i,j} = 0$ and $\Sigma_{i,j}^{\mathbf{V}} = \mathbf{0}$, proving the asymptotic independence of $\widehat{\mathbf{v}}_{m,i}$ and $\widehat{\mathbf{v}}_{m,j}$ in case of arbitrary types $i \neq j$. (We note that the reason of the equation $Z_{i,j} = 0$ is the conditional independence of the zero-mean variables $V_{n,i}$ and $V_{n,j}$ in formula (9).)

(iv) Since the last statement can be proved analogously as the earlier ones, we do not detail the steps. Using ergodicity and the asymptotic normality of the estimator $\widehat{\boldsymbol{\mu}}'_i$, one can show that

$$\sqrt{m}(\widehat{\mathbf{v}}'_{m,i} - \mathbf{v}_i) = (m\mathbb{I}_{m,2}^{-1}) \left[\frac{1}{\sqrt{m}} \sum_{n=1}^m \frac{V_{n,i} \mathbb{Y}_{n-1}}{(\mathbf{1}^\top \mathbb{Y}_{n-1})^2} + o_P(1) \right], \quad m \rightarrow \infty, \quad i = 1, \dots, p.$$

Then, the statement is implied by MCLT similarly as in the proof of (ii) and (iii). ■

Acknowledgement. The author is thankful to her supervisors, Gyula Pap and Gábor Szűcs for introducing her to the problem investigated in this paper. The careful reading of the manuscript and helpful comments of the referee are greatly appreciated.

References

- [1] K. B. ATHREYA and P. E. NEY, *Branching Processes*, Springer-Verlag, New York - Heidelberg, 1972.
- [2] J.-P. DION, G. GAUTHIER and A. LATOUR, Branching processes with immigration and integer-valued time series, *Serdica Math. J.*, **21** (1995), 123–136.
- [3] J. G. DU and Y. LI, The integer-valued autoregressive (INAR(p)) model, *J. Time Ser. Anal.*, **12** (1991), 129–142.
- [4] C. C. HEYDE, *Quasi-likelihood and its application: a general approach to optimal parameter estimation*, Springer-Verlag, New York, 1997.
- [5] C. C. HEYDE and E. SENETA, Estimation theory for growth and immigration rates in a multiplicative process, *J. Appl. Probability*, **9** (1972), 235–256.
- [6] C. C. HEYDE and E. SENETA, Notes on “Estimation theory for growth and immigration rates in a multiplicative process”, *J. Appl. Probability*, **11** (1974), 572–577.
- [7] J. JACOD and A. N. SHIRYAEV, *Limit theorems for stochastic processes*, Springer-Verlag, Berlin, 2003.
- [8] N. KAPLAN, The multitype Galton-Watson process with immigration, *Ann. Probability*, **1** (1973), 947–953.
- [9] L. A. KLIMKO and P. I. NELSON, On conditional least squares estimation for stochastic processes, *Ann. Statist.*, **6** (1978), 629–642.
- [10] C. J. MODE, *Multitype branching processes. Theory and applications*, American Elsevier Publishing Co., New York, 1971.

- [11] F. NEDÉNYI, A CUSUM type change detection test based on martingale differences, *arXiv* (2014), arXiv: [1407.4412](#).
- [12] P. I. NELSON, A note on strong consistency of least squares estimators in regression models with martingale difference errors, *Ann. Statist.*, **8** (1980), 1057–1064.
- [13] M. P. QUINE and P. DURHAM, Estimation for multitype branching processes, *J. Appl. Probability*, **14** (1977), 829–835.
- [14] G. SZŰCS, Ergodic properties of subcritical multitype Galton–Watson processes, *arXiv* (2014), arXiv: [1402.5539](#).
- [15] C. Z. WEI and J. WINNICKI, Estimation of the means in the branching process with immigration, *Ann. Statist.*, **18** (1990), 1757–1773.
- [16] J. WINNICKI, Estimation of the variances in the branching process with immigration, *Probab. Theory Related Fields*, **88** (1991), 77–106.
- [17] N. M. YANEV and S. TCHOUKOVA-DANTCHEVA, On the statistics of branching processes with immigration, *C. R. Acad. Bulgare Sci.*, **33** (1980), 469–471.

F. NEDÉNYI, Bolyai Institute, University of Szeged, Aradi vértanúk tere 1, H-6720 Szeged, Hungary; *e-mail*: : nfanni@math.u-szeged.hu