Journal of the Royal Statistical Society
Statistical Methodology
Series B

J. R. Statist. Soc. B (2018)
80, *Part* 4, *pp.* 769–791

Modelling non-stationary multivariate time series of counts via common factors

Fangfang Wang

University of Wisconsin, Madison, USA

and Haonan Wang

Colorado State University, Fort Collins, USA

[Received November 2016. Final revision February 2018]

Summary. We develop a new parameter-driven model for multivariate time series of counts. The time series is not necessarily stationary. We model the mean process as the product of modulating factors and unobserved stationary processes. The former characterizes the long-run movement in the data, whereas the latter is responsible for rapid fluctuations and other unknown or unavailable covariates. The unobserved stationary processes evolve independently of the past observed counts and might interact with each other. We express the multivariate unobserved stationary processes as a linear combination of possibly low dimensional factors that govern the contemporaneous and serial correlation within and across the observed counts. Regression coefficients in the modulating factors are estimated via pseudo-maximum-likelihood estimation, and identification of common factor(s) is carried out through eigenanalysis on a positive definite matrix that pertains to the autocovariance of the observed counts at non-zero lags. Theoretical validity of the two-step estimation procedure is documented. In particular, we establish consistency and asymptotic normality of the pseudo-maximum-likelihood estimator in the first step and the convergence rate of the second-step estimator. We also present an exhaustive simulation study to examine the finite sample performance of the estimators, and numerical results corroborate our theoretical findings. Finally, we illustrate the use of the proposed model through an application to the numbers of National Science Foundation fundings awarded to seven research universities from January 2001 to December 2012.

Keywords: Count time series; Eigendecomposition; Factor model; Generalized auto-regressive conditional heteroscedasticity; Generalized linear models; Pseudolikelihood

1. Introduction

Modelling time series of counts has always been challenging, as these data are discrete, non-negative and often overdispersed, in addition to being correlated in the time dimension and the cross-section dimension. In this paper, we propose a new parameter-driven model for multivariate time series of counts, with the aim of dimension reduction. The correlation is assumed to arise from latent factors of possibly low dimension. The approach that is pursued here is rooted in the idea of Zeger (1988) and Lam *et al.* (2011). Zeger (1988) developed an extension of log-linear models for univariate count time series. By conditioning on a stationary latent process, the observed counts are independent of each other and follow a Poisson log-linear regression model. Such models have also been considered by Brännäs and Johansson (1994), Campbell (1994), Chan and Ledolter (1995) and Davis *et al.* (2000), among others. In particular,

Address for correspondence: Haonan Wang, Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA.
E-mail: wanghn@stat.colostate.edu

 1369–7412/18/80769

Davis *et al.* (2000) proposed a practical approach to diagnosing the existence of a latent process. The present work considers a multivariate extension of the parameter-driven models that were discussed in Zeger (1988) and Davis *et al.* (2000), which meanwhile adapts the common factor identification procedure for linear models that was proposed by Lam *et al.* (2011) and Chang *et al.* (2015) to non-linear models with count responses. Because we break the linkage between the conditional mean and conditional probability distribution, this paper offers a unified framework for count time series, which could be adapted to different parametric models, for instance, the negative binomial regression model.

The literature on factor analysis of multivariate count time series is relatively sparse. Jørgensen *et al.* (1999) proposed a multivariate Poisson state space model with one latent factor specified as a gamma Markov process to model daily emergency room visits. Wedel *et al.* (2003) considered a *static* multivariate Poisson factor model for cross-sectional analysis. Jung *et al.* (2011) proposed a dynamic factor model for analysing transaction volumes over a 5-min interval from multiple US stocks. The conditional mean is determined by three latent factors: the common market factor, industry-specific factor and stock-specific factor, and the factors are assumed to follow independent Gaussian auto-regressive AR(1) processes.

The motivation for this study comes from the data set that contains the numbers of US National Science Foundation (NSF) fundings awarded to seven research universities from January 2001 to December 2012. The time series are depicted in Fig. 1 in Section 5, and a strong and stable month of the year effect is revealed. To understand the underlying driving forces that are common to all the institutions, we consider a doubly stochastic model, where the observed counts are Poisson distributed conditionally on the means. The mean process is further modelled as the product of modulating factors and unobserved stationary processes. The former picks up the long-run movement in the data, whereas the latter is responsible for rapid fluctuations and other unknown or unavailable covariates. Exogenous variables enter the mean processes through a multiplicative structure, which allows the observed counts to be non-stationary. The advantage of adopting a multiplicative structure is its ease of interpretation when we need to account for non-stationarity and/or explanatory ability of other variables. Multiplicative models are commonly used for non-negative time series. A notable example is the spline generalized auto-regressive conditional heteroscedasticity (GARCH) model of Engle and Rangel (2008) which modelled high frequency return volatility as a product of a slowly moving component and a stationary unit GARCH process, and the slowly moving component is non-stationary and is a function of macroeconomic and financial variables. The unobserved stationary processes evolve independently of the past observed counts. Distinct from Jørgensen *et al.* (1999) and Jung *et al.* (2011), the unobserved processes in this paper are allowed to interact with each other. Yet, we do not specify a dynamic (state space) model for the unobserved counts. We attempt to extract common factors out of them. The multivariate unobserved stationary processes are therefore expressed as a linear combination of possibly low dimensional factors that govern the contemporaneous and serial correlation within and between the observed counts.

The marginal distribution of observed counts is not specified, however, in that we would like to explore the intertemporal and intratemporal dependence of count time series with little interference. The likelihood function *per se* is not readily available. We therefore consider a two-step procedure to estimate the model parameters. In the first step, a pseudo-maximum-likelihood approach is adopted to estimate the regression coefficients in the modulating factors. Gourieroux *et al.* (1984a) discussed pseudo-maximum-likelihood (PML) based on linear exponential families, quasi-generalized PML and PML based on quadratic exponential families, where the last two cases have the first and second moments involved. Gourieroux *et al.* (1984b) pointed out that, among the normal, Poisson, negative binomial and gamma families, none of them outperforms the others. In this paper, we estimate the parameters by maximizing a Poisson likelihood function in that it requires only the first moment. After obtaining consistent estimators of the regression coefficients, we then calibrate the factor loading matrix through eigenanalysis in the second step. Precisely speaking, we conduct eigendecomposition of the covariance matrices of the count responses that are discounted by the estimated modulating factors. Our analysis relies on the fact that the latent factors preserve the cross-covariance of the observed counts at non-zero lags. We adapt the common factor identification procedure for linear models that was proposed by Lam *et al.* (2011) and Chang *et al.* (2015) to non-linear models with count responses. Our proposed two-step procedure is fast to compute and easy to implement. The optimization in the first step is a convex problem, and its numerical calculation can be carried out by most standard algorithms or packages. In the second step, we formulate the estimation problem of a loading matrix as an eigenproblem, which can be readily solved by existing packages for eigendecomposition.

The rest of the paper is organized as follows. Section 2 introduces a new doubly stochastic model for multivariate time series of counts. The estimation procedure and asymptotic properties are provided in Section 3. We assess the proposed estimation procedure in finite samples in Section 4. Real data analysis is illustrated in Section 5. Discussion regarding the calibration of the asymptotic covariance matrix is included in Appendix A. Notation, proofs and additional discussion and simulation results are given in the on-line supplemental material. The MATLAB code for carrying out our proposed method can be downloaded from `https://github.com/fangfanw/Multi-Count-Time-Series`.

2. A doubly stochastic model with latent factors

Consider a $p\times 1$ vector of time series of counts, $Y_t=(y_{1t},\ldots,y_{pt})^{\mathrm{T}}$. Our objective is to study the dependence, over time, of each component of Y_t and across components. Each univariate time series y_{jt} is modelled as a doubly stochastic process. For this, we introduce an auxiliary vector process $N_t(\cdot)=(N_{1t}(\cdot),\ldots,N_{pt}(\cdot))^{\mathrm{T}}$, where $\{N_{jt}(\cdot), j=1,\ldots,p, t=1,2,\ldots,n\}$ is a sequence of independent Poisson processes of unit intensity. For each j and t, there is a positive (not necessarily stationary) process λ_{jt}, with finite variance and independent of N_t, such that $y_{jt}=N_{jt}(\lambda_{jt})$, which counts the number of events of $N_{jt}(\cdot)$ in the time interval $[0,\lambda_{jt}]$; see Fokianos *et al.* (2009) for more details. In other words, y_{jt} is conditionally Poisson distributed given λ_{jt}, i.e.

$$y_{jt}|\lambda_{jt}\sim\text{Poisson}(\lambda_{jt}), \tag{1}$$

and y_{jt} is conditionally independent of each other given $\Lambda_t=(\lambda_{1t},\ldots,\lambda_{pt})^{\mathrm{T}}$. The process y_{jt} is reminiscent of deformation of a Poisson process; see, for instance, Rydberg and Shephard (2000). The difference is that λ_{jt} is not necessarily non-decreasing. The unobserved p-dimensional process Λ_t preserves the cross-covariance of the observed count Y_t, i.e., for $j\neq j'$,

$$\text{cov}(y_{jt}, y_{j't'})=\text{cov}(\lambda_{jt},\lambda_{j't'}), \qquad \text{for any } t, t'. \tag{2}$$

Yet, the correlation of the observed counts is smaller in magnitude than that of λ_t because $\text{var}(y_{jt})=E(\lambda_{jt})+\text{var}(\lambda_{jt})$.

Enlightened by this discussion, we propose to model the dependence structure through latent factors. In particular, we assume that, conditioning on a latent positive *stationary* process $\epsilon_t=(\epsilon_{1t},\ldots,\epsilon_{pt})^{\mathrm{T}}$, λ_{jt}, $j=1,\ldots,p$, are independent of each other over time and across components, and

$$\mu_{jt}\equiv E(\lambda_{jt}|\epsilon_t)=\alpha_{jt}\epsilon_{jt}, \tag{3}$$

where $\alpha_{jt}=E(\lambda_{jt})$. Evidently, $E(\epsilon_{jt})=1$. Processes satisfying these conditions are given in Section 4. One appealing feature of the multiplicative decomposition (3) is that

$$\operatorname{cov}(y_{jt},y_{j't'})=\alpha_{jt}\alpha_{j't'}\operatorname{cov}(\epsilon_{jt},\epsilon_{j't'}), \tag{4}$$

for either

(a) any t, t' and $j\neq j'$, or
(b) $t\neq t'$ and any j and j'.

In other words, the cross-covariance function of two distinct latent time series $\{\epsilon_{jt}\}$ and $\{\epsilon_{j't}\}$ can be fully recovered from that of $\{\alpha_{jt}^{-1}y_{jt}\}$ and $\{\alpha_{j't}^{-1}y_{j't}\}$, two trend-adjusted series; the auto-covariance matrices of the latent vector process ϵ_t at non-zero lags are equivalent to those of the vector process $(\alpha_{1t}^{-1}y_{1t},\ldots,\alpha_{pt}^{-1}y_{pt})^{\mathrm{T}}$. The implication of result (4) will be discussed further in Section 3.2.

The deterministic processes α_{jt}, $j=1,\ldots,p$, are modulating factors that describe the long-run movement in the count time series. They can be designed to pick up time trends, seasonal effects, business cycles, calendar effect, day of the week effects, diurnal pattern and so forth. In this paper, we link α_{jt} with q-dimensional covariates $x_{n,t}$ in a way that

$$\log(\alpha_{jt})=\beta_j^{\mathrm{T}}x_{n,t}, \qquad j=1,2,\ldots,p, \quad t=1,\ldots,n, \tag{5}$$

where β_j is a $q\times 1$ vector. The subscript n in $x_{n,t}$ indicates that the covariate may depend on the length of the time series. From here onwards we shall suppress this dependence for our convenience of notation.

The latent vector process ϵ_t is introduced to account for possible overdispersion and correlation in the time dimension as well as in the cross-sectional dimension, apart from capturing rapid fluctuations and other unknown or unavailable covariates. A natural step forward is to consider a dynamic model for ϵ_t. This practice, however, would be numerically inefficient or even infeasible if the dimension p is large. This work stands in the middle ground. Instead of modelling the process ϵ_t, we attempt to bring down its dimensionality by assuming that the information that is contained in ϵ_t can be summarized by a few possibly low dimensional factors, i.e.

$$\epsilon_t=Af_t, \tag{6}$$

where $f_t=(f_{1t},\ldots,f_{rt})^{\mathrm{T}}$ is an $r\times 1$ vector of latent factors, and $A=(a_1,\ldots,a_r)$ is a $p\times r$ time invariant factor loading matrix of rank r ($r\leqslant p$). In view of results (4) and (6), the dependence between y_{jt} and $y_{j't}$ ($j\neq j'$) emerges from the common underlying factors f_t, and innovations driving y_{jt} may have an effect on $y_{j't}$ and vice versa, which is measured by the matrix A. The specification (6) has been considered in Jørgensen *et al.* (1999) and Jung *et al.* (2011). In particular, Jørgensen *et al.* (1999) assumed that the conditional means of multivariate Poisson counts are driven by one common gamma Markov process, i.e. $r=1$, A is a column vector of 1s and f_t is a gamma Markov process. Jung *et al.* (2011) modelled the log-conditional-mean of stock transaction volumes as a linear combination of three factors—the sector factor, market factor and idiosyncratic factor—and each factor evolves independently of each other in a fashion of stationary Gaussian auto-regression of order 1. In this paper, r is unknown, and it needs to be estimated from the data. Moreover, we would not impose any structural assumption on f_t, in that we aim to reduce the dimension of the factors instead of modelling the dynamics of the data.

This work reaches beyond the existing literature in two ways. While we extend the parameter-driven models of Zeger (1988) and Davis and Wu (2009) for univariate count time series to multivariate cases, we break the linkage between the conditional mean and conditional probability distribution of λ_{jt}. We do not propose a new data-generating process, but rather we offer a unified framework that could be embedded into different parametric models. Note that, with $p=1$ and $\lambda_{jt}=\alpha_{jt}\epsilon_{jt}$, expression (1) together with expression (5) are the well-studied Poisson regression model for univariate Poisson counts; see, for instance, Brännäs and Johansson (1994), Campbell (1994), Chan and Ledolter (1995), Davis *et al.* (2000) and Zeger (1988). Moreover, if λ_{jt} is gamma distributed conditioning on ϵ_t, $\lambda_{jt}|\epsilon_t \sim \text{gamma}(\kappa_j, \kappa_j^{-1}\mu_{jt})$ with $\mu_{jt}=\alpha_{jt}\epsilon_{jt}$ and κ_j is a positive number, then y_{jt} is conditional negative binomial given ϵ_t, i.e. $\text{NB}\{\kappa_j, (\mu_{jt}+\kappa_j)^{-1}\mu_{jt}\}$; this is analogous to the negative binomial regression model of Davis and Wu (2009) if $p=1$. More details will be offered in the next section. Definitions of the aforementioned distributions are provided in the on-line supplemental material.

As opposed to many observation-driven models, $\{Y_t\}$ does not have to be stationary. The framework proposed is applicable to both stationary and non-stationary count responses. The vector time series $\{Y_t\}$ is stationary if and only if $\{\Lambda_t\}$ is, for the reason that $E(y_{jt})=E(\lambda_{jt})$, $\text{var}(y_{jt})=E(\lambda_{jt})+\text{var}(\lambda_{jt})$ and $\text{cov}(y_{jt}, y_{jt'})=\text{cov}(\lambda_{jt}, \lambda_{jt'})$ for $t\neq t'$, together with equation (2).

In the present work, we would like to explore the intertemporal and intratemporal dependence of count time series with little interference. The conditional distribution of λ_{jt} is left unspecified. This is in clear contrast with the work of Jørgensen *et al.* (1999) and Jung *et al.* (2011). Considering that the conditional distribution of y_{jt} is unknown, we shall use a PML approach to estimate the regression coefficients and then calibrate the factor loading matrix by means of eigenanalysis.

3. Latent factors recovery

In this section, we study a two-step procedure for estimating the model parameters. In particular, we first estimate the coefficient vectors β_j, $j=1,\ldots,p$, as described in expression (5), by maximizing a possibly misspecified likelihood function. Next, we formulate the estimation of the loading matrix A in equation (6) as an eigenproblem, which can be easily solved by eigendecomposition.

3.1. *Pseudo-maximum-likelihood estimation of β_j*

Let $B=(\beta_1,\ldots,\beta_p)$ be a $q\times p$ matrix of coefficients, and b be its vectorization, i.e. $b=(\beta_1^{\mathrm{T}},\ldots,\beta_p^{\mathrm{T}})^{\mathrm{T}}\in\mathbb{R}^{pq}$. Further denote by $B_0=(\beta_{01},\ldots,\beta_{0p})$ the true value of B with corresponding vectorization b_0. Recall that, in our proposed doubly stochastic model, the conditional distribution of y_{jt}, $j=1,\ldots,p$, $t=1,\ldots,n$, given ϵ_t, $t=1,\ldots,n$, is

$$E\left\{\prod_{j=1}^{p}\prod_{t=1}^{n}\frac{\exp(-\lambda_{jt})(\lambda_{jt})^{y_{jt}}}{y_{jt}!}\middle|\epsilon_t\right\}.$$

To estimate the regression coefficients, a standard approach is to integrate out the latent vectors ϵ_t, $t=1,\ldots,n$, further to obtain the distribution of y_{jt}, $j=1,\ldots,p$, $t=1,\ldots,n$. This would inevitably require knowledge of the conditional distribution of λ_{jt} given ϵ_t and the joint distribution of $(\epsilon_1,\ldots,\epsilon_n)$, which, as a result, not only makes the model less parsimonious but also increases computational complexity.

In this paper, we are seeking a simple estimation approach that produces consistent estimators regardless of the probability distributions of λ_{jt} and ϵ_t, $j=1,\ldots,p$, $t=1,\ldots,n$. A PML method is adopted. PML estimators are obtained by maximizing a likelihood function that is associated with a family of probability distributions that does not necessarily contain the true distribution. Here, we construct the PML on the basis of the Poisson family in that only the first moment is required. See Gourieroux *et al.* (1984a, b) for further discussion. Essentially, an estimator of b_0, denoted by $\hat{b}$, is a maximizer of the objective function

$$l_n(b) = (pn)^{-1} \sum_{j=1}^{p} \sum_{t=1}^{n} \{y_{jt}\beta_j^{\mathrm{T}} x_t - \exp(\beta_j^{\mathrm{T}} x_t)\}, \tag{7}$$

for $b = (\beta_1^{\mathrm{T}}, \ldots, \beta_p^{\mathrm{T}})^{\mathrm{T}}$ within a compact set $\Theta \subset \mathbb{R}^{pq}$ that contains b_0 as an interior point. Note that $l_n(b)$ is the true rescaled log-likelihood function (up to an additive constant) if y_{jt}, $j = 1, \ldots, p$, follow a Poisson distribution with mean $\exp(\beta_j^{\mathrm{T}} x_t)$ and are independent within each time series and across different time series. Though a similar practice appeared in Gourieroux *et al.* (1984b), a remarkable difference is that Gourieroux *et al.* (1984b) applied PML methods to Poisson models with independent observations.

To establish consistency and asymptotic normality of the resulting estimator, we assume that the latent factor f_t and the regressor x_t satisfy the following assumptions.

Assumption 1. The latent factor f_t is stationary and strongly mixing with mixing coefficient $\alpha(m)$ satisfying $\Sigma_{m=1}^{\infty} \alpha(m)^{\lambda/(\lambda+2)} < \infty$ for some $\lambda > 0$, and $E|f_{jt}|^{2(\lambda+2)} < \infty$ for all j.

Assumption 2. For each $j = 1, \ldots, p$,

(a) the limit of $n^{-1}\Sigma_{t=1}^{n} x_t x_t^{\mathrm{T}} \exp(\beta_{0j}^{\mathrm{T}} x_t)$ exists, and it is invertible, denoted by $\Omega_1^{(j)}$,
(b) the limit of $n^{-1}\Sigma_{t=1}^{n} x_t x_t^{\mathrm{T}} E\{\mathrm{var}(\lambda_{jt}|\epsilon_t)\}$ exists and is denoted by $\Omega_2^{(j)}$,
(c) $n^{-1}\Sigma_{t=1}^{n} x_{t+k} x_t^{\mathrm{T}} \exp(\beta_{0j}^{\mathrm{T}} x_{t+k} + \beta_{0j'}^{\mathrm{T}} x_t)$ converges to $W_k^{(j,j')}$ uniformly in $|k| < n$ as $n \to \infty$, and the limit $W_k^{(j,j')}$ fulfils the property that $\Omega_3^{(j,j')} \doteq \Sigma_{k=-\infty}^{\infty} W_k^{(j,j')} \Sigma_{\epsilon}^{(j,j')}(k) < \infty$ (componentwise) (moreover, both $n^{-1}\Sigma_{t=1}^{k} x_{t-k} x_t^{\mathrm{T}} \exp(\beta_{0j}^{\mathrm{T}} x_{t-k} + \beta_{0j'}^{\mathrm{T}} x_t)$ and $n^{-1}\Sigma_{t=n-k+1}^{n} \{x_{t+k} x_t^{\mathrm{T}} \times \exp(\beta_{0j}^{\mathrm{T}} x_{t+k} + \beta_{0j'}^{\mathrm{T}} x_t)\}$ are uniformly bounded in $k \in (0, n)$ and converge to 0 as $n \to \infty$) and
(d) $\sup_{n \geqslant 1} n^{-1}\Sigma_{t=1}^{n} |x_{ti} x_{tj} x_{tk}| \exp(\beta_j^{\mathrm{T}} x_t) < \infty$ for any $i, j, k \in \{1, \ldots, q\}$ and $b \in \Theta$, where x_{ti} is the ith element of x_t.

Assumption 1 ensures that the process $\epsilon_t = A f_t$ is stationary and strongly mixing; see Lahiri (2013) for the definition of strong mixing for a multivariate process. Assumption 2 is introduced to regulate the limiting behaviour of the Hessian matrix (i.e. assumption 2, part (a)) and the gradient of $l_n(b)$ (i.e. assumption 2, parts (b) and (c)), and to bound the third derivative of $l_n(b)$ (i.e. assumption 2, part (d)). These assumptions are similar to conditions (3)–(5) in Davis *et al.* (2000), except for assumption 2, part (b), that pertains to the conditional distribution of λ_{jt} given ϵ_t. For instance, if $\lambda_{jt}|\epsilon_t \sim \mathrm{gamma}(\kappa_j, \kappa_j^{-1}\alpha_{jt}\epsilon_{jt})$ and κ_j is a positive number, then $\Omega_2^{(j)} = \kappa_j^{-1} E(\epsilon_{j0}^2) W_0^{(j,j)}$. If λ_{jt} is simply $\alpha_{jt}\epsilon_{jt}$, we have that $\mathrm{var}(\lambda_{jt}|\epsilon_t) = 0$ and thus $\Omega_2^{(j)}$ is 0. Thus the matrix $\Omega_2^{(j)}$ could be degenerate. Moreover, note that $\mathrm{var}(y_{jt}) = \exp(\beta_{0j}^{\mathrm{T}} x_t) + E\{\mathrm{var}(\lambda_{jt}|\epsilon_t)\} + \exp(2\beta_{0j}^{\mathrm{T}} x_t) \Sigma_{\epsilon}^{(j,j)}(0)$. As a direct consequence of assumption 2, parts (a)–(c),

$$\lim_{n\to\infty} n^{-1} \sum_{t,t'=1}^{n} \mathrm{cov}(y_{jt}, y_{jt'}) x_t x_t'^{\mathrm{T}} = \Omega_1^{(j)} + \Omega_2^{(j)} + \Omega_3^{(j,j)}$$

and

$$\lim_{n\to\infty} n^{-1} \sum_{t,t'=1}^{n} \mathrm{cov}(y_{jt}, y_{j't'}) x_t x_t'^{\mathrm{T}} = \Omega_3^{(j,j')}$$

for $j \neq j'$ (see the proof of corollary 1).

Let $\Omega_l = \text{diag}(\Omega_l^{(1)}, \Omega_l^{(2)}, \ldots, \Omega_l^{(p)})$ for $l=1,2$, and $\Omega_3 = (\Omega_3^{(j,j')})_{p\times p}$, which are positive definite matrices. We further introduce notation for the matrix norm. Denote by $\|M\|_{\text{F}}$ the Frobenius norm of a matrix M, and by $\|M\|$ the square root of the maximum eigenvalue of $M^{\text{T}}M$. Note that $\|m\|$ is simply the Euclidean norm if m is a vector. A detailed account of the notation appears in section A.1 of the on-line supplementary material. The asymptotic properties of $\hat{b}$ are summarized in the following theorem.

Theorem 1. Suppose that assumptions 1 and 2 hold. Then $\hat{b}$ is consistent for b_0. Further assume that $\max_{1\leqslant t\leqslant n} n^{-1/6}\|x_t\| = o(1)$ and $\max_{1\leqslant j\leqslant p, 1\leqslant t\leqslant n} E(\lambda_{jt}^3|\epsilon_t) = O_p(1)$. Then we have $n^{1/2}(\hat{b}-b_0)$ converges to $N\{0, \Omega_1^{-1} + \Omega_1^{-1}(\Omega_2+\Omega_3)\Omega_1^{-1}\}$ in distribution.

In section A.2 of the on-line supplemental material, we compare theorem 1 with the results of asymptotic normality that were established in Davis *et al.* (2000) and Davis and Wu (2009), and provide further discussion regarding the connection of our proposed method and the existing literature.

Last but not least, we quantify the mean-squared error of $\hat{b}$ as follows.

Corollary 1. Under the assumptions in theorem 1, the mean-squared-error matrix $E(\hat{b}-b_0)(\hat{b}-b_0)^{\text{T}}$ to order n^{-1} is given by

$$E\{(\hat{b}-b_0)(\hat{b}-b_0)^{\text{T}}\} = n^{-2}H_1^{-1}\left\{\sum_{t,t'=1}^{n} \text{cov}(Y_t, Y_{t'}) \otimes x_t x_{t'}^{\text{T}}\right\}H_1^{-1},$$

where $H_1 = -n^{-1}\Sigma_{t=1}^{n} \text{diag}\{\exp(\beta_{01}^{\text{T}}x_t), \ldots, \exp(\beta_{0p}^{\text{T}}x_t)\} \otimes x_t x_t^{\text{T}}$ and $\lim_{n\to\infty} nE\{(\hat{b}-b_0)(\hat{b}-b_0)^{\text{T}}\} = \Omega_1^{-1}(\Omega_1+\Omega_2+\Omega_3)\Omega_1^{-1}$.

3.2. Estimation of A

The contemporaneous dependence and cross-correlation of the counts Y_t are driven by possibly low dimensional latent factors f_t with the magnitude determined by the modulating factor $\exp(\beta_{0j}^{\text{T}}x_t)$, $j=1,\ldots,p$, and the factor loading matrix A. In this subsection, we shall discuss the estimation of A given a consistent estimator of b_0 or B_0.

Let $\Sigma_\epsilon(k) = \text{cov}(\epsilon_{t+k}, \epsilon_t)$ and $\Sigma_f(k) = \text{cov}(f_{t+k}, f_t)$. We immediately have

$$\Sigma_\epsilon(k) = A\Sigma_f(k)A^{\text{T}}, \qquad k=0, \pm 1, \pm 2, \ldots. \tag{8}$$

A major challenge in this paper is that ϵ_t cannot be calibrated from data. To circumvent this problem, we work with $\eta_t = (\eta_{1t}, \ldots, \eta_{pt})^{\text{T}}$, where $\eta_{jt} = y_{jt}/E(y_{jt}) = \exp(-\beta_{0j}^{\text{T}}x_t)y_{jt}$. This is inspired by equation (4) and the fact that the autocovariance matrix is of primary interest. The appeal of the detrended time series η_t is that it preserves the covariance structure of ϵ_t at non-zero lags, albeit $\eta_{jt} \neq \epsilon_{jt}$. To be precise, the following results hold true:

$$\Sigma_\epsilon(k) = \text{cov}(\eta_{t+k}, \eta_t), \qquad \text{for } k\neq 0, \tag{9}$$

$$\text{cov}(\epsilon_{jt}, \epsilon_{j't}) = \text{cov}(\eta_{jt}, \eta_{j't}), \qquad \text{for } j\neq j', \tag{10}$$

which follow from equation (4). Here, result (9) suggests that we can estimate the *covariance* structure of the latent process ϵ_t through η_t. For univariate count time series, i.e. $p=1$, Zeger (1988) and Davis *et al.* (2000) used a similar, but different, idea to model the autocovariance function, which may not be readily generalized for multivariate count time series. Moreover, Davis *et al.* (2000) and Davis and Wu (2009) adopted Pearson residuals to model the auto-correlation of the latent process ϵ_t for univariate count time series. However, Pearson residuals may not be a suitable choice for estimating the auto-covariance function; in fact, the cross-covariance matrices of the Pearson residuals $h_{jt} = \exp(-\beta_{0j}^{\mathrm{T}} x_t/2)\{y_{jt} - \exp(\beta_{0j}^{\mathrm{T}} x_t)\}$ are not identical to $\Sigma_\epsilon(k)$. For instance, the covariance between h_{jt} and $h_{j't}$ for $j \neq j'$ is $\exp(\beta_{0j}^{\mathrm{T}} x_t/2)\exp(\beta_{0j'}^{\mathrm{T}} x_t/2)\mathrm{cov}(\epsilon_{jt}, \epsilon_{j't})$. In this paper, we take a different approach which is based on the following relationship, as a consequence of results (8) and (9):

$$\mathrm{cov}(\eta_{t+k}, \eta_t) = A\Sigma_f(k)A^{\mathrm{T}}, \qquad k \neq 0. \tag{11}$$

Note that $\mathrm{cov}(\eta_t, \eta_t)$ is time dependent, and thus expression (11) does not hold for $k = 0$. Hence, η_{jt} is *not* necessarily stationary, which is a remarkable distinction compared with ϵ_{jt}. As the right-hand side of expression (11) is invariant of t, we write $\Sigma_\eta(k)$ for $\mathrm{cov}(\eta_{t+k}, \eta_t)$ at $k \neq 0$.

Since $\epsilon_t = Af_t = AH(H^{-1}f_t)$ for any invertible matrix H, the loading matrix A is not uniquely determined; however, the linear space that is spanned by the columns of A, denoted by $\mathcal{M}(A)$, is unique. If $\Sigma_f(k)$ is invertible, $\mathcal{M}(A)$ is the orthogonal complement of the linear space that is spanned by the eigenvectors of $\Sigma_\eta(k)^{\mathrm{T}}$ corresponding to the zero eigenvalues. Lam *et al.* (2011) suggested estimating the matrix A by performing eigenanalysis for a non-negative definite matrix that sums $\Sigma_\eta(k)\Sigma_\eta(k)^{\mathrm{T}}$ over different lags k. Define

$$L = \sum_{k=1}^{k_0} \Sigma_\eta(k)\Sigma_\eta(k)^{\mathrm{T}}, \tag{12}$$

where k_0 is a prescribed fixed positive integer. Then $\mathcal{M}(A)$ is also the orthogonal complement of the linear space that is spanned by the eigenvectors of L corresponding to the zero eigenvalues under the following assumption.

Assumption 3. The $r \times r$ matrix $\Sigma_f(k)$ is invertible, for $k = 1, \ldots, k_0$.

The loading matrix $A = (a_1, \ldots, a_r)$ is made up of the orthonormal eigenvectors corresponding to the positive eigenvalues of L. Lam *et al.* (2011) and Chang *et al.* (2015) proposed to extract the factor loading matrix A by plugging-in estimates of $\Sigma_\eta(k)$ in equation (12). In their theoretical development, the following assumption is required to ensure the uniqueness of the loading matrix up to sign changes.

Assumption 4. The eigenvalues of the $p \times p$ matrix L, $\lambda_1, \ldots, \lambda_p$, satisfy $\lambda_1 > \ldots > \lambda_r > \lambda_{r+1} = \ldots = \lambda_p = 0$.

The columns of A are arranged such that the associated eigenvalues are in descending order, i.e. a_j is an orthonormal eigenvector associated with λ_j, $j = 1, \ldots, r$. We further propose to use $\hat{\eta}_t = (\hat{\eta}_{1t}, \ldots, \hat{\eta}_{pt})^{\mathrm{T}}$ with $\hat{\eta}_{jt} = \exp(-\hat{\beta}_j^{\mathrm{T}} x_t)y_{jt}$ to derive the sample counterpart of $\Sigma_\eta(k)$, $\hat{\Sigma}_\eta(k) = n^{-1}\Sigma_{t=1}^{n-k}(\hat{\eta}_{t+k} - \bar{\eta})(\hat{\eta}_t - \bar{\eta})^{\mathrm{T}}$ for $k > 0$, where $\bar{\eta} = n^{-1}\Sigma_{t=1}^{n}\hat{\eta}_t$. A sample version of L can be obtained by substituting $\hat{\Sigma}_\eta(k)$ for $\Sigma_\eta(k)$:

$$\hat{L} = \sum_{k=1}^{k_0} \hat{\Sigma}_\eta(k)\hat{\Sigma}_\eta(k)^{\mathrm{T}}. \tag{13}$$

We shall show next that the relationships (9) and (10) hold asymptotically after we substitute the PML estimators from the first step for the unknown parameters. Moreover, the validity of the estimator $\hat{L}$ is justified. Some further assumptions are introduced first.

Assumption 5.

(a) $\sup_{t \geqslant 1} \|x_t\|$ is finite;
(b) $\sup_{j=1,\ldots,p, t \geqslant 1} E(\lambda_{jt}^4)$ is finite.

Lemma 1. Suppose that assumptions 1, 2 and 5 hold. Then we have

$$\mathrm{cov}(\eta_{t+k}, \eta_t) = \mathrm{cov}(\hat{\eta}_{t+k}, \hat{\eta}_t) + O(n^{-1/2}), \qquad \text{for } k \neq 0, \tag{14}$$

$$\mathrm{cov}(\eta_{jt}, \eta_{j't}) = \mathrm{cov}(\hat{\eta}_{jt}, \hat{\eta}_{j't}) + O(n^{-1/2}), \qquad \text{for } j \neq j'. \tag{15}$$

Lemma 2. Suppose that assumptions 1, 2 and 5 hold. For $k = 1, 2, \ldots, k_0$, we have

$$\|\hat{\Sigma}_\eta(k) - \Sigma_\epsilon(k)\|_\mathrm{F} = O_p(n^{-1/2}). \tag{16}$$

As an application of lemma 2, we have $\|\hat{L} - L\| = O_p(n^{-1/2})$. Let $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ be the eigenvalues of $\hat{L}$ in descending order, and $\hat{a}_1, \ldots, \hat{a}_p$ be the corresponding orthonormal eigenvectors. An estimator of the loading matrix A is given by $\hat{A} = (\hat{a}_1, \ldots, \hat{a}_r)$. As suggested in Lam *et al.* (2011), we replace some $\hat{a}_j$ with $-\hat{a}_j$ so that the direction matches a_j. The consistency of $\hat{A}$ is established in the following theorem.

Theorem 2. Suppose that r is known and fixed, and assumptions 1–5 hold. Then we have

$$\|\hat{A} - A\| = O_p(n^{-1/2}). \tag{17}$$

Moreover, $|\hat{\lambda}_j - \lambda_j| = O_p(n^{-1/2})$ for $j = 1, 2, \ldots, r$, and $|\hat{\lambda}_j| = O_p(n^{-1})$ for $j = r+1, \ldots, p$.

Note that $E(A^\mathrm{T}\eta_t | f_t) = f_t$. We shall use $\hat{f}_t = \hat{A}^\mathrm{T}\hat{\eta}_t$ as a proxy for the latent factors f_t. We further have the following result.

Theorem 3. Under the assumptions of theorem 2, we have $\|\hat{A}\hat{f}_t - \epsilon_t\| = O_p(1)$.

This result coincides with theorem 2.2 of Chang *et al.* (2015) when p is fixed. Although f_t is not fully recovered, $\hat{f}_t$ tracks the serial correlation and cross-correlation of f_t at non-zero lags. This is made precise in the following theorem.

Theorem 4. Under the assumptions of theorem 2, we have $\|\hat{\Sigma}_f(k) - \Sigma_f(k)\| = O_p(n^{-1/2})$ for $k = 1, 2, \ldots, k_0$, where $\hat{\Sigma}_f(k) = n^{-1}\Sigma_{t=1}^{n-k}(\hat{f}_{t+k} - \bar{f})(\hat{f}_t - \bar{f})^\mathrm{T}$ and $\bar{f} = n^{-1}\Sigma_{t=1}^{n}\hat{f}_t$.

The discussion so far assumes that r is known, but that is not so in practice. To obtain a consistent estimator of r, we adopt the ridge-type ratio estimator (Chang *et al.*, 2015; Xia *et al.*, 2015). In particular, an estimator of r can be obtained through an optimization problem, i.e.

$$\hat{r} = \arg \min_{j=1,\ldots,p-1} \frac{\hat{\lambda}_{j+1} + C_n}{\hat{\lambda}_j + C_n}, \tag{18}$$

where C_n is a positive constant. As will be shown in the following theorem, $\hat{r}$ is a consistent estimator of r.

Theorem 5. Suppose that assumptions 1–5 hold. For $\hat{r}$ defined in equation (18), $P(\hat{r} \neq r) \to 0$ as $n \to \infty$, provided that $C_n = o(1)$ and $C_n^{-1} = o(n)$.

In practice, there are several choices, up to a constant scale, for C_n including $n^{-1}\log(n)$, $n^{-1}\log\{\log(n)\}$, and $n^{-\rho}$ where $0 < \rho < 1$. Xia *et al.* (2015) suggested $C_n = (10n)^{-1}\log(n)$ for selecting the number of factors when modelling the volatility of multivariate time series. As will be seen later, we use a simulation study to assess the performance of $C_n = n^{-1}\log(n)$ and $C_n = n^{-1}\log\{\log(n)\}$ in finite sample.

Denote by $\tilde{A}$ the estimator of A with r estimated from the data, i.e. $\tilde{A} = (\hat{a}_1, \ldots, \hat{a}_{\hat{r}})$, and write $\tilde{f}_t$ for $\tilde{A}^\mathrm{T}\hat{\eta}_t$. Further define $\tilde{\Sigma}_f(k) = n^{-1}\Sigma_{t=1}^{n-k}(\tilde{f}_{t+k} - \bar{f})(\tilde{f}_t - \bar{f})^\mathrm{T}$, where $\bar{f} = n^{-1}\Sigma_{t=1}^{n}\tilde{f}_t$. If $\tilde{r}$

and r disagree, $(\tilde{A}, \tilde{\Sigma}_f(k))$ have different dimensions compared with $(A, \Sigma_f(k))$. To gauge the accuracy of our estimators, we use a variant of discrepancy measure of Pan and Yao (2008) that metricizes the distance between the two-factor loading spaces, $\mathcal{M}(\tilde{A})$ and $\mathcal{M}(A)$:

$$D\{\mathcal{M}(\tilde{A}), \mathcal{M}(A)\} = \sqrt{\left\{1 - \frac{1}{\max(\hat{r}, r)} \mathrm{tr}(\tilde{A}\tilde{A}^{\mathrm{T}} A A^{\mathrm{T}})\right\}}, \tag{19}$$

which was introduced by Chang *et al.* (2015). Here, $D\{\mathcal{M}(\tilde{A}), \mathcal{M}(A)\}$ ranges from 0 to 1, and it is 0 if and only if $\hat{r} = r$ and $\mathcal{M}(\tilde{A}) = \mathcal{M}(A)$. The following theorem can be shown.

Theorem 6. Suppose that assumptions 1–5 hold. For $\hat{r}$ defined in equation (18) with $C_n = o(1)$ and $C_n^{-1} = o(n)$, we have $D\{\mathcal{M}(\tilde{A}), \mathcal{M}(A)\} = O_p(n^{-1/2})$.

4. Simulation study

In this section, we conduct simulation studies to assess the finite sample performance of our proposed two-step procedure as described in Section 3. We further investigate and compare several choices of k_0 and C_n in determining the number of factors.

4.1. Data-generating process

We consider a 5×1 count time series $Y_t = (y_{1t}, y_{2t}, \ldots, y_{5t})^{\mathrm{T}}$ that are conditionally Poisson distributed given a (multivariate) positive process Λ_t. The conditional mean of Λ_t given ϵ_t, denoted by $\mu_t = (\mu_{1t}, \ldots, \mu_{5t})^{\mathrm{T}}$, is determined by

$$\begin{pmatrix} \log(\mu_{1t}) \\ \log(\mu_{2t}) \\ \log(\mu_{3t}) \\ \log(\mu_{4t}) \\ \log(\mu_{5t}) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0.3 & -0.05 \\ 2 & 0.1 & -0.03 \\ 3 & 0.5 & -0.07 \\ 4 & 0.4 & -0.01 \\ 5 & 0.6 & -0.1 \end{pmatrix}}_{B_0^{\mathrm{T}}} \underbrace{\begin{pmatrix} 1 \\ \cos(2\pi t/5) \\ \sin(2\pi t/5) \end{pmatrix}}_{x_t} + \begin{pmatrix} \log(\epsilon_{1t}) \\ \log(\epsilon_{2t}) \\ \log(\epsilon_{3t}) \\ \log(\epsilon_{4t}) \\ \log(\epsilon_{5t}) \end{pmatrix}. \tag{20}$$

The latent process ϵ_t is driven by low dimensional common factors f_t. Here, we generate f_t under two different scenarios.

4.1.1. Scenario 1

We first consider a situation in which only one latent factor presents, i.e. $r = 1$. The factor loading matrix A is $A = 1/\sqrt{5}(1, 1, 1, 1, 1)^{\mathrm{T}}$. Let $h_t = \log(f_t/\sqrt{5})$. We set h_t as a Gaussian auto-regression model of order 1 as in Davis *et al.* (2000) and Davis and Wu (2009), i.e.

$$h_t = \phi_0 + \phi_1 h_{t-1} + z_t, \tag{21}$$

where z_t has a normal distribution with mean 0 and variance σ_z^2, $\phi_0 = -0.285(1 - \phi_1)$ and $\sigma_z^2 = 0.57(1 - \phi_1^2)$. The magnitude of ϕ_1 reflects the strength of intertemporal dependence of the latent process, in that

$$\Sigma_f(k) = 5\{\exp(0.57\phi_1^k) - 1\}, \qquad \text{for } k \geqslant 0, \tag{22}$$

and $\Sigma_f(k) = \Sigma_f(-k)$ for $k < 0$. In the simulation study, ϕ_1 takes values of 0.3 and 0.9 to achieve low and strong serial correlation respectively.

4.1.2. Scenario 2

We consider a model with two common latent factors, i.e. $r=2$, and ϵ_t takes the form of equation (6) with $A=(a_1,a_2)$ where $a_1=(\sqrt{\frac{1}{2}},\sqrt{\frac{1}{2}},0,0,0)^{\mathrm{T}}$ and $a_2=(0,0,\sqrt{\frac{1}{3}},\sqrt{\frac{1}{3}},\sqrt{\frac{1}{3}})^{\mathrm{T}}$. The process $f_t=(f_{1,t},f_{2,t})^{\mathrm{T}}$ is a (positive) bivariate GARCH process: $f_{j,t}=V_{j,t}(\varepsilon_{j,t})^2$, $\varepsilon_{j,t}\sim^{\mathrm{IID}} N(0,1)$ for $j=1,2$ with $\mathrm{corr}(\varepsilon_{1,t},\varepsilon_{2,t})=\rho$ and $\mathrm{corr}(\varepsilon_{1,t},\varepsilon_{2,t'})=0$ for $t\neq t'$, and

$$\begin{pmatrix} V_{1,t} \\ V_{2,t} \end{pmatrix}=\begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}+\begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix}\begin{pmatrix} f_{1,t-1} \\ f_{2,t-1} \end{pmatrix}+\begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix}\begin{pmatrix} V_{1,t-1} \\ V_{2,t-1} \end{pmatrix}, \tag{23}$$

where $(\omega_1,\omega_2)=(0.0580,0.3031)$, $(\alpha_1,\alpha_2)=(0.009,0.025)$ and $(\beta_1,\beta_2)=(0.95,0.80)$. The values of the parameters are chosen in such a way that $A^{\mathrm{T}}A=I_2$, $E(\epsilon_t)=1$ and assumption 1 is fulfilled. With the parameters in equation (23) fixed, the intradependence and interdependence between $f_{1,t}$ and $f_{2,t}$ are fully determined by ρ, the contemporaneous correlation between the Gaussian innovations. This is because

$$\Sigma_f(k)=\begin{pmatrix} \mathrm{cov}(f_{1,k},f_{1,0}) & \mathrm{cov}(f_{1,k},f_{2,0}) \\ \mathrm{cov}(f_{2,k},f_{1,0}) & \mathrm{cov}(f_{2,k},f_{2,0}) \end{pmatrix}, \tag{24}$$

where, letting $\sigma_j^2=(1-\alpha_j-\beta_j)^{-1}\omega_j$,

$$\begin{aligned} \mathrm{var}(f_{j,0})&=2\sigma_j^4\frac{1-(\alpha_j+\beta_j)^2+\alpha_j^2}{1-(\alpha_j+\beta_j)^2-2\alpha_j^2}, && j=1,2, \\ \mathrm{cov}(f_{j,0},f_{j',0})&=2\rho^2\sigma_j^2\sigma_{j'}^2\frac{1-(\alpha_j+\beta_j)(\alpha_{j'}+\beta_{j'})+\alpha_j\alpha_{j'}}{1-(\alpha_j+\beta_j)(\alpha_{j'}+\beta_{j'})-2\alpha_j\alpha_{j'}\rho^2}, && j\neq j', \end{aligned} \tag{25}$$

$$\begin{aligned} \mathrm{cov}(f_{j,k},f_{j,0})&=2\alpha_j\sigma_j^4(\alpha_j+\beta_j)^{k-1}\frac{1-(\alpha_j+\beta_j)\beta_j}{1-(\alpha_j+\beta_j)^2-2\alpha_j^2}, && k>0, \quad j=1,2, \\ \mathrm{cov}(f_{j,k},f_{j',0})&=2\alpha_j\rho^2(\alpha_j+\beta_j)^{k-1}\sigma_j^2\sigma_{j'}^2\frac{1-(\alpha_j+\beta_j)\beta_{j'}}{1-(\alpha_j+\beta_j)(\alpha_{j'}+\beta_{j'})-2\alpha_j\alpha_{j'}\rho^2}, && k>0, \quad j\neq j', \end{aligned} \tag{26}$$

and $\Sigma_f(k)=\Sigma_f(-k)^{\mathrm{T}}$ for $k<0$. Derivation is available in the on-line supplemental material. We consider ρ at two levels: $\rho=0$ and $\rho=0.8$. The former indicates that the two components $f_{j,t}$ are uncorrelated with each other at all leads and lags, whereas the latter reflects a strong cross-correlation between the two processes.

Last, but not least, for each scenario, λ_{jt} is drawn from one of the following two cases:

(a) $\lambda_{jt}=\mu_{jt}$ and
(b) $\lambda_{jt}|\epsilon_t\sim\mathrm{gamma}(\kappa_j,\kappa_j^{-1}\mu_{jt})$.

As a consequence, the conditional distribution of Y_t given ϵ_t is respectively Poisson and negative binomial. In case (b), the parameters κ_j, $j=1,2,\ldots,5$, that control the degree of dispersion, are chosen as 39.06, 7.30, 11.57, 20.85 and 19.58.

For each scenario and each choice of λ_{jt}, we simulate 1000 data sets. To ensure the stationarity of f_t in each iteration, we drop the first 100 observations and keep the remaining n observations. Here, we consider two different sample sizes: $n=150$ and $n=1500$. The first choice is in line with the data size in our empirical analysis in Section 5.

4.2. Simulation results

Simulation results pertaining to scenario 1 are summarized in Table 1 ($\phi_1=0.3$) and Table 2

Table 1. Bias, Bias, sample standard deviation, Std, standard error, SE, and asymptotic standard deviation, Astd, for our proposed estimator of the matrix $B_0 = (B_0^{(kj)})$ in scenario 1 with $\phi_1 = 0.3$

k, j	*Results for Poisson model*				*Results for negative binomial model*			
	Bias	*Std*	*SE*	*Astd*	*Bias*	*Std*	*SE*	*Astd*
$n=150$								
$k=1, j=1$	−0.0148	0.1030	0.0806	0.1054	−0.0132	0.1078	0.0819	0.1068
$k=2, j=1$	0.0071	0.1220	0.1085	0.1246	0.0058	0.1266	0.1124	0.1270
$k=3, j=1$	−0.0060	0.1295	0.1079	0.1240	−0.0059	0.1273	0.1119	0.1265
$k=1, j=2$	−0.0142	0.0949	0.0683	0.0975	−0.0132	0.1044	0.0717	0.1055
$k=2, j=2$	0.0082	0.1103	0.0882	0.1105	0.0077	0.1289	0.0960	0.1243
$k=3, j=2$	−0.0049	0.1104	0.0885	0.1105	−0.0046	0.1239	0.0968	0.1243
$k=1, j=3$	−0.0116	0.0931	0.0640	0.0947	−0.0145	0.0975	0.0651	0.0999
$k=2, j=3$	0.0068	0.1073	0.0894	0.1077	0.0067	0.1178	0.0938	0.1175
$k=3, j=3$	−0.0039	0.1065	0.0874	0.1066	−0.0052	0.1157	0.0916	0.1162
$k=1, j=4$	−0.0115	0.0915	0.0593	0.0935	−0.0111	0.0940	0.0602	0.0964
$k=2, j=4$	0.0057	0.1045	0.0778	0.1047	0.0048	0.1083	0.0809	0.1102
$k=3, j=4$	−0.0060	0.1032	0.0766	0.1041	−0.0046	0.1089	0.0797	0.1096
$k=1, j=5$	−0.0107	0.0911	0.0539	0.0931	−0.0114	0.0960	0.0557	0.0963
$k=2, j=5$	0.0048	0.1050	0.0713	0.1063	0.0033	0.1092	0.0761	0.1125
$k=3, j=5$	−0.0054	0.1048	0.0684	0.1046	−0.0042	0.1116	0.0727	0.1105
$n=1500$								
$k=1, j=1$	−0.0027	0.0340	0.0246	0.0333	−0.0017	0.0350	0.0250	0.0338
$k=2, j=1$	−0.0010	0.0394	0.0313	0.0394	−0.0004	0.0404	0.0324	0.0402
$k=3, j=1$	0.0006	0.0391	0.0310	0.0392	−0.0009	0.0400	0.0321	0.0400
$k=1, j=2$	−0.0013	0.0319	0.0207	0.0308	−0.0025	0.0344	0.0221	0.0334
$k=2, j=2$	−0.0002	0.0345	0.0244	0.0349	−0.0003	0.0394	0.0269	0.0393
$k=3, j=2$	−0.0002	0.0348	0.0245	0.0349	−0.0009	0.0412	0.0270	0.0393
$k=1, j=3$	−0.0017	0.0307	0.0194	0.0299	−0.0012	0.0321	0.0203	0.0316
$k=2, j=3$	−0.0012	0.0334	0.0252	0.0341	−0.0017	0.0359	0.0265	0.0372
$k=3, j=3$	−0.0004	0.0337	0.0245	0.0337	−0.0007	0.0363	0.0260	0.0367
$k=1, j=4$	−0.0017	0.0306	0.0186	0.0296	−0.0017	0.0318	0.0190	0.0305
$k=2, j=4$	−0.0013	0.0320	0.0227	0.0331	−0.0012	0.0336	0.0234	0.0349
$k=3, j=4$	−0.0000	0.0333	0.0224	0.0329	−0.0002	0.0347	0.0231	0.0347
$k=1, j=5$	−0.0014	0.0306	0.0175	0.0294	−0.0020	0.0318	0.0181	0.0305
$k=2, j=5$	−0.0015	0.0328	0.0218	0.0336	−0.0015	0.0343	0.0230	0.0356
$k=3, j=5$	−0.0002	0.0335	0.0209	0.0331	−0.0002	0.0357	0.0222	0.0350

($\phi_1 = 0.9$). To evaluate the performance of parameter estimation of B_0 (or its vectorization b_0), we compute the bias, Bias, and standard deviation, Std, for each component of $\hat{b}$ based on 1000 parameter estimates. As a comparison, we report the true asymptotic standard deviations, Astd, based on theorem 1 with $\Sigma_f(k)$ given in expression (22) and its estimates SE as discussed in Appendix A. Normality is further checked in Figs S.1 and S.2 of the on-line supplemental material. For brevity, we show only the normal quantile plots for $\phi_1 = 0.9$ and $n = 1500$. The sample-based standard deviations Std are very close to the Astds for both Poisson and negative binomial distributions. Considering also the normal quantile plots, the simulation results justify theorem 1. The reported SE slightly underestimates Astd in some circumstances in scenario 1. This is expected as discussed in Appendix A. The interdependence level affects the estimation in finite samples, especially in small samples. But the effect becomes less pronounced when the sample size grows bigger. Further note that the complexity of data introduces an extra layer of uncertainty. The negative binomial distribution yields larger Astd and SE than the Poisson case does.

Table 2. Bias, Bias, sample standard deviation, Std, standard error, SE, and asymptotic standard deviation, Astd, for our proposed estimator of the matrix $B_0 = (B_0^{(kj)})$ in scenario 1 with $\phi_1 = 0.9$

k, j	*Results for Poisson model*				*Results for negative binomial model*			
	Bias	*Std*	*SE*	*Astd*	*Bias*	*Std*	*SE*	*Astd*
$n = 150$								
$k=1,\ j=1$	−0.0521	0.2706	0.1429	0.2940	−0.0509	0.2695	0.1456	0.2945
$k=2,\ j=1$	0.0008	0.0856	0.0876	0.0836	0.0023	0.0866	0.0908	0.0872
$k=3,\ j=1$	−0.0038	0.0863	0.0877	0.0830	0.0015	0.0882	0.0908	0.0866
$k=1,\ j=2$	−0.0498	0.2647	0.1371	0.2912	−0.0497	0.2673	0.1391	0.2940
$k=2,\ j=2$	0.0011	0.0632	0.0593	0.0618	0.0018	0.0819	0.0716	0.0840
$k=3,\ j=2$	−0.0003	0.0640	0.0607	0.0618	−0.0009	0.0846	0.0713	0.0840
$k=1,\ j=3$	−0.0470	0.2634	0.1341	0.2903	−0.0491	0.2650	0.1355	0.2920
$k=2,\ j=3$	0.0008	0.0505	0.0486	0.0525	0.0015	0.0692	0.0576	0.0704
$k=3,\ j=3$	−0.0003	0.0524	0.0485	0.0519	−0.0012	0.0697	0.0564	0.0695
$k=1,\ j=4$	−0.0473	0.2640	0.1325	0.2899	−0.0498	0.2643	0.1327	0.2909
$k=2,\ j=4$	−0.0010	0.0459	0.0395	0.0480	0.0001	0.0581	0.0456	0.0590
$k=3,\ j=4$	−0.0020	0.0486	0.0399	0.0477	−0.0028	0.0581	0.0462	0.0587
$k=1,\ j=5$	−0.0473	0.2640	0.1303	0.2898	−0.0488	0.2653	0.1306	0.2908
$k=2,\ j=5$	−0.0009	0.0451	0.0345	0.0469	−0.0010	0.0561	0.0429	0.0596
$k=3,\ j=5$	−0.0011	0.0470	0.0337	0.0464	−0.0027	0.0573	0.0406	0.0587
$n = 1500$								
$k=1,\ j=1$	−0.0021	0.0927	0.0535	0.0930	−0.0043	0.0922	0.0541	0.0931
$k=2,\ j=1$	0.0010	0.0257	0.0263	0.0264	−0.0010	0.0287	0.0270	0.0276
$k=3,\ j=1$	0.0009	0.0271	0.0261	0.0262	0.0003	0.0269	0.0269	0.0274
$k=1,\ j=2$	−0.0035	0.0911	0.0518	0.0921	−0.0038	0.0921	0.0522	0.0930
$k=2,\ j=2$	−0.0011	0.0194	0.0172	0.0196	0.0005	0.0263	0.0205	0.0266
$k=3,\ j=2$	0.0006	0.0193	0.0173	0.0195	0.0024	0.0267	0.0204	0.0266
$k=1,\ j=3$	−0.0035	0.0909	0.0513	0.0918	−0.0035	0.0916	0.0514	0.0924
$k=2,\ j=3$	−0.0008	0.0165	0.0139	0.0166	−0.0013	0.0215	0.0166	0.0223
$k=3,\ j=3$	0.0009	0.0163	0.0138	0.0164	0.0016	0.0218	0.0165	0.0220
$k=1,\ j=4$	−0.0036	0.0911	0.0510	0.0917	−0.0037	0.0914	0.0511	0.0920
$k=2,\ j=4$	−0.0005	0.0142	0.0114	0.0152	−0.0007	0.0189	0.0129	0.0187
$k=3,\ j=4$	0.0011	0.0152	0.0113	0.0151	0.0013	0.0182	0.0131	0.0186
$k=1,\ j=5$	−0.0036	0.0908	0.0506	0.0916	−0.0034	0.0914	0.0507	0.0920
$k=2,\ j=5$	−0.0011	0.0143	0.0099	0.0148	−0.0016	0.0182	0.0126	0.0189
$k=3,\ j=5$	0.0011	0.0147	0.0095	0.0147	0.0007	0.0184	0.0122	0.0185

Simulation results for scenario 2 are presented in Tables A and B and Figs S.8 and S.9 of the on-line supplemental material. In contrast, the SE and Astd that are associated with the two-factor cases are in general relatively big and converge slowly. It is largely due to the complex structure of the bivariate GARCH specification, which results in cross-temporal dependence, as revealed by equations (25) and (26). Overall, similar patterns emerge. The numerical results are in accordance with theorem 1. It is worth noting that the SE and Astd that are reported in Table A, where the two latent factors are uncorrelated at all leads and lags, are close to the results in Table B that deals with correlated latent factors, regardless of the distribution of λ_{jt}. This holds true for both $n = 150$ and $n = 1500$.

After obtaining the estimator of B_0, we then proceed to estimate the number of factors, r, by using the estimator that is defined in equation (18). According to theorem 5, we consider two choices of C_n, $C_n = n^{-1}\log(n)$ and $C_n = n^{-1}\log\{\log(n)\}$. The simulation results are benchmarked with those obtained by using $C_n = 0$, as it yields the estimator of Lam and Yao (2012).

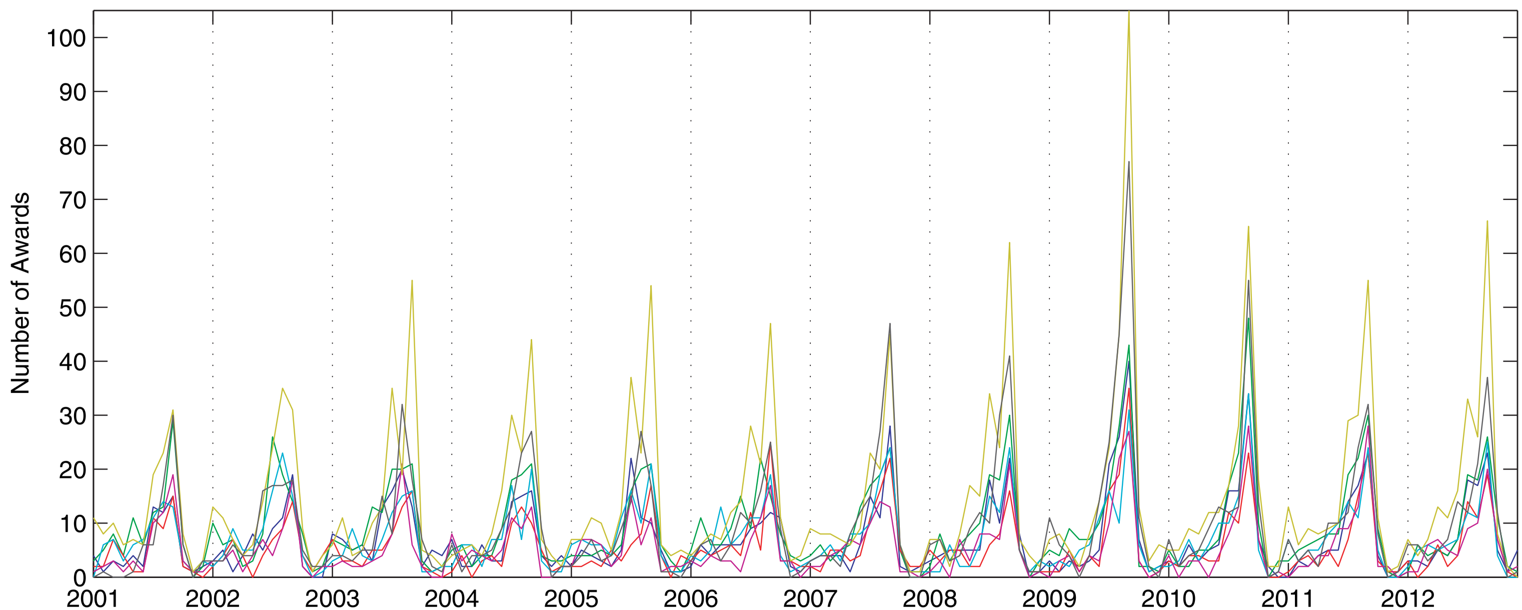


Fig. 1. Time series plot of the monthly NSF awards received by seven institutions from January 2001 to December 2012 (institutions 1–4 are private research universities, and the others are public; the data and a detailed description are available from http://selfsynchronize.com/hayne/sna/): ——, institution 1; ——, institution 2; ——, institution 3; ——, institution 4; ——, institution 5; ——, institution 6; ——, institution 7

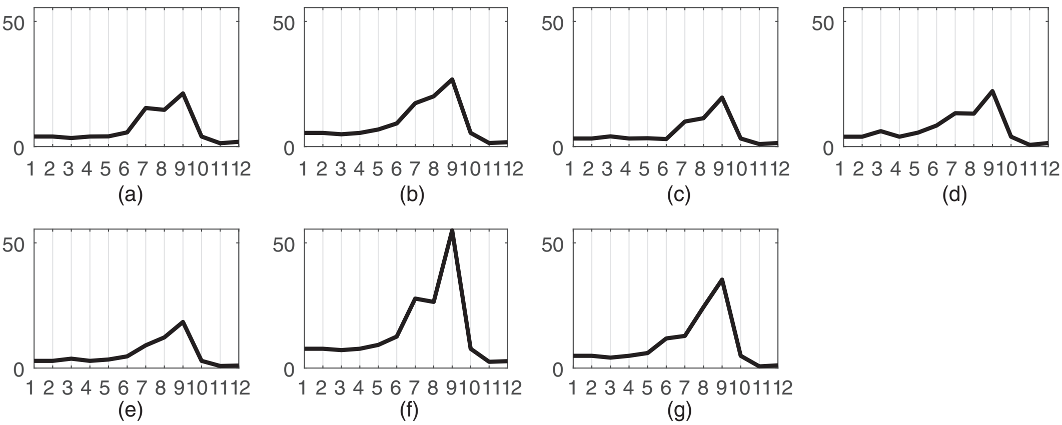


Fig. 2. Time series plot of the estimated modulating factors $\exp(\hat{\beta}_j^{\mathrm{T}} x_t)$ for each institution over 1 year, i.e. $t=1,2,\ldots,12$: (a) institution 1; (b) institution 2; (c) institution 3; (d) institution 4; (e) institution 5; (f) institution 6; (g) institution 7

The eigenvalues $\hat{\lambda}_j$ are extracted from $\hat{L}$ defined in equation (13). We first examine $\hat{L}$ as an estimator of L for some different values of k_0. Figs S.3 and S.10 in the on-line supplemental material depict the relationship between the scaled norm $k_0^{-1}\|L-\hat{L}\|$ and k_0 for two levels of interdependence (i.e. $\phi_1=0.3$ and $\phi_1=0.9$ in scenario 1) and two levels of cross-dependence (i.e. $\rho=0$ and $\rho=0.8$ in scenario 2) respectively. The broken curve corresponds to a sample size of $n=150$, and the full curve is for $n=1500$. The norm $\|L-\hat{L}\|$ drops rapidly as n increases, which is consistent with lemma 2 and equation (S.25) in the on-line supplemental material.

The relative frequencies for $\hat{r}$, 1 in scenario 1 and 2 in scenario 2, are calculated for each choice of k_0, and they are illustrated in Figs S.4, S.6, S.11 and S.13 of the supplemental material. Among the three choices of C_n, $C_n=n^{-1}\log(n)$ unanimously outperforms the others. But a large value of k_0 is required in scenario 2 to raise the chance of favouring 2. Moreover, increasing sample size would also reduce selection bias, as shown in Figs S.11 and S.13. On account of the increasing temporal and/or cross-sectional correlations in f_t, the relationships between the series $\hat{\eta}_{jt}$, $j=1,\ldots,5$, become more difficult to pin down numerically. It should be noted that the less satisfactory performance that is observed in Fig. S.13 is not due to the misspecification in the first step. We plot in Figs S.5, S.7, S.12 and S.14 the corresponding relative frequencies with the true B_0 inserted, to assess the sensitivity of the estimator in the second step to the first-step estimation. By carefully examining the graphs, we reach the same conclusion that $C_n=n^{-1}\log(n)$ is preferred over the others, and a larger k_0 is required for more complex data. It appears that the PML estimator in the first step has little effect on the subsequent analysis. This observation also corroborates the theoretical finding in lemma 1 that the vector time series $\hat{\eta}_t$ preserves the temporal and cross-sectional dependence of η_t asymptotically. Because the choice of k_0 relies on the nature of data under consideration, we suggest examining also the eigenvalues $\hat{\lambda}_j$ in practical implementation.

5. Data example

In this section, we implement our proposed method to analyse the numbers of NSF fundings awarded to seven research universities from January 2001 to December 2012. The data were extracted from `www.nsf.gov`. Fig. 1 depicts the numbers of NSF awards that those institutions received every month over the entire time period. Institutions 1–4 are private research universi-

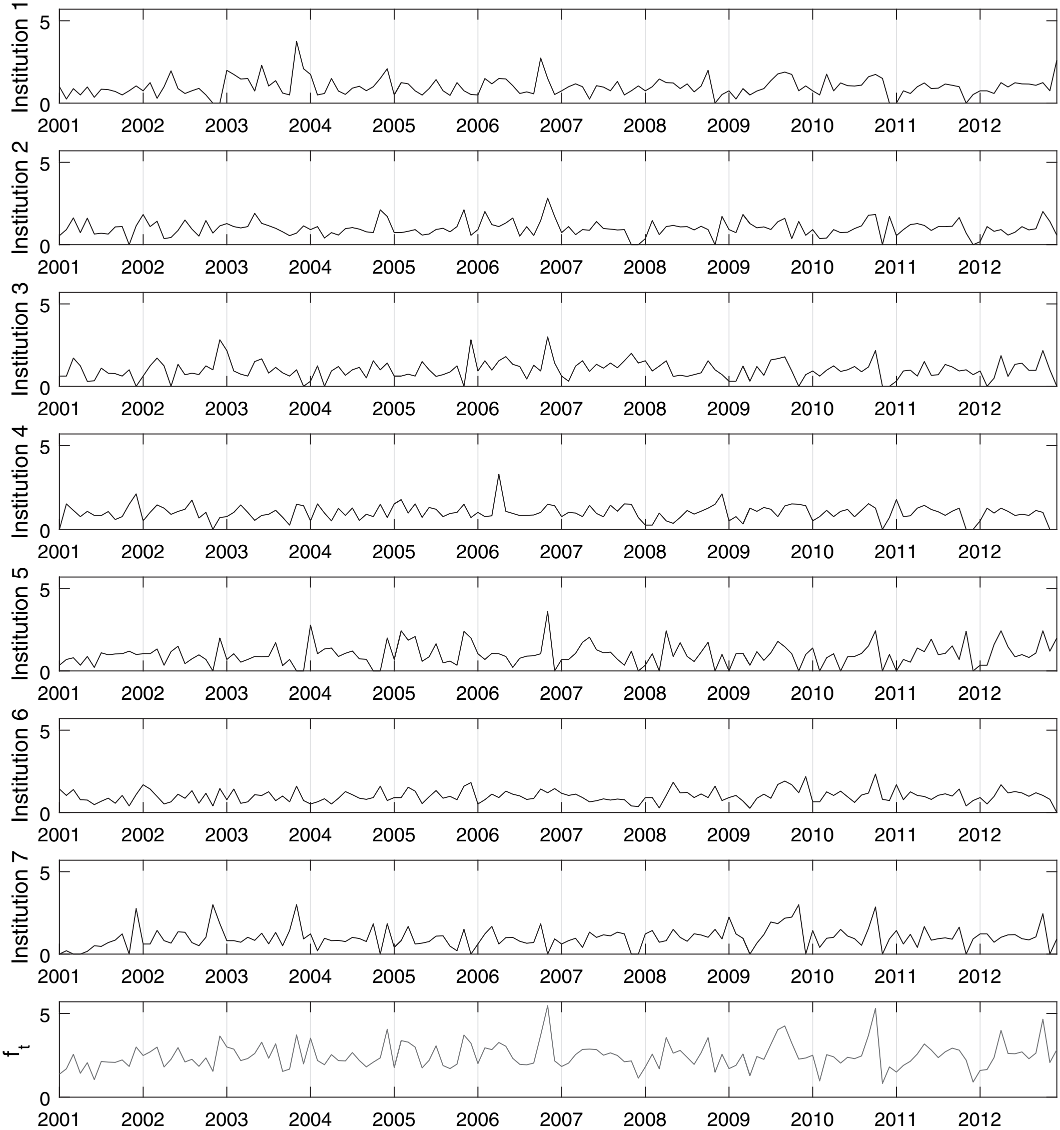


Fig. 3. Time plots of $\hat{\eta}_{j,t}$ and $\tilde{f}_t$: the figure shows the time plots of the seasonally adjusted series $\hat{\eta}_{j,t} = \exp(-\hat{\beta}_j^{\mathrm{T}} x_t) y_{jt}$ for each institution and the estimated latent factor $\tilde{f}_t$ in the bottom panel

ties, whereas the others are public. Table C in the on-line supplemental material reports the mean, the standard deviation and the five-number summary for each of the seven institutions. The data exhibit a strong month of the year effect. More grants were awarded by the NSF in September than at any other time of the year, and there are troughs in January, November and December. In particular, institutions 1, 3, 6 and 7 reached their maximums in September 2009, whereas the peak occurred in September 2010 for institutions 2 and 4. Institution 5 received the most awards in both September 2009 and September 2010. The increase in the number of research grants around 2009 could be attributed to the stimulus package which was effective from February 2009.

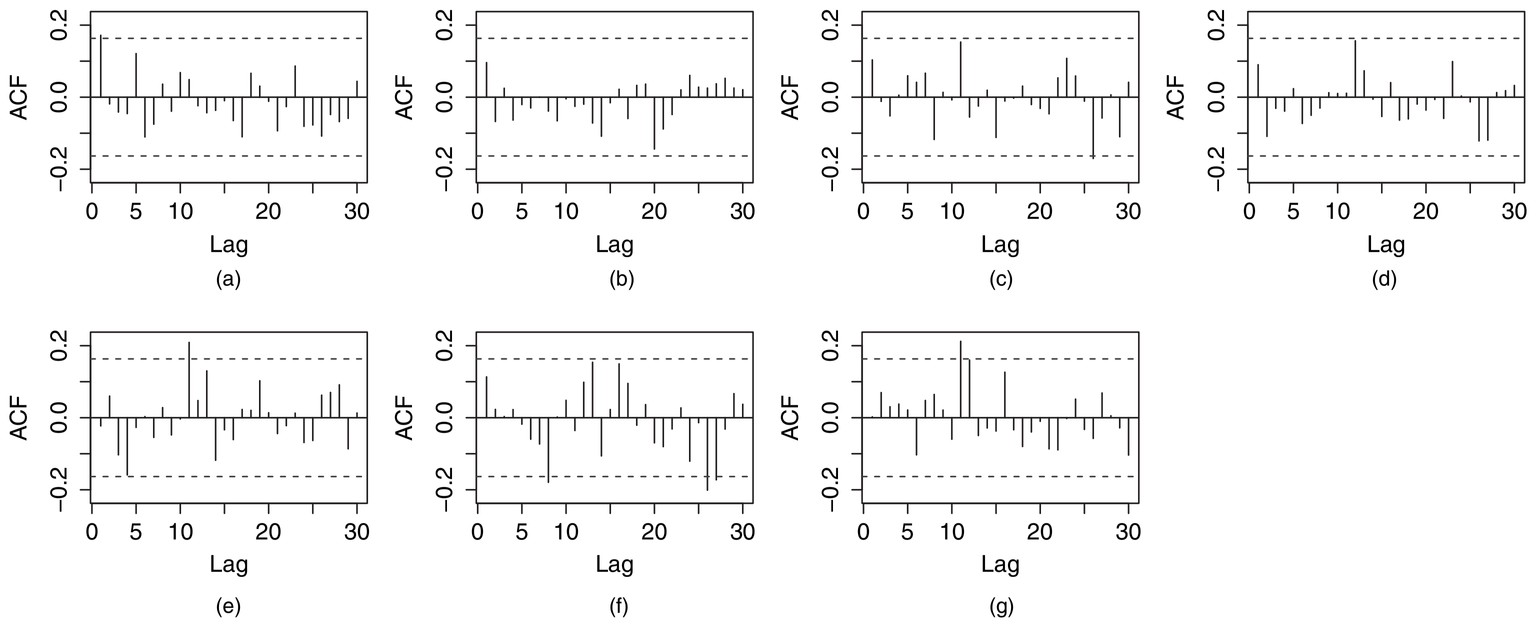


Fig. 4. Correlograms of $\hat{\eta}_{j,t}$ (the figure depicts the correlograms of each seasonally adjusted series $\hat{\eta}_{j,t}$ over the entire sample, i.e. $t = 1, 2, \ldots, 144$): (a) institution 1; (b) institution 2; (c) institution 3; (d) institution 4; (e) institution 5; (f) institution 6; (g) institution 7

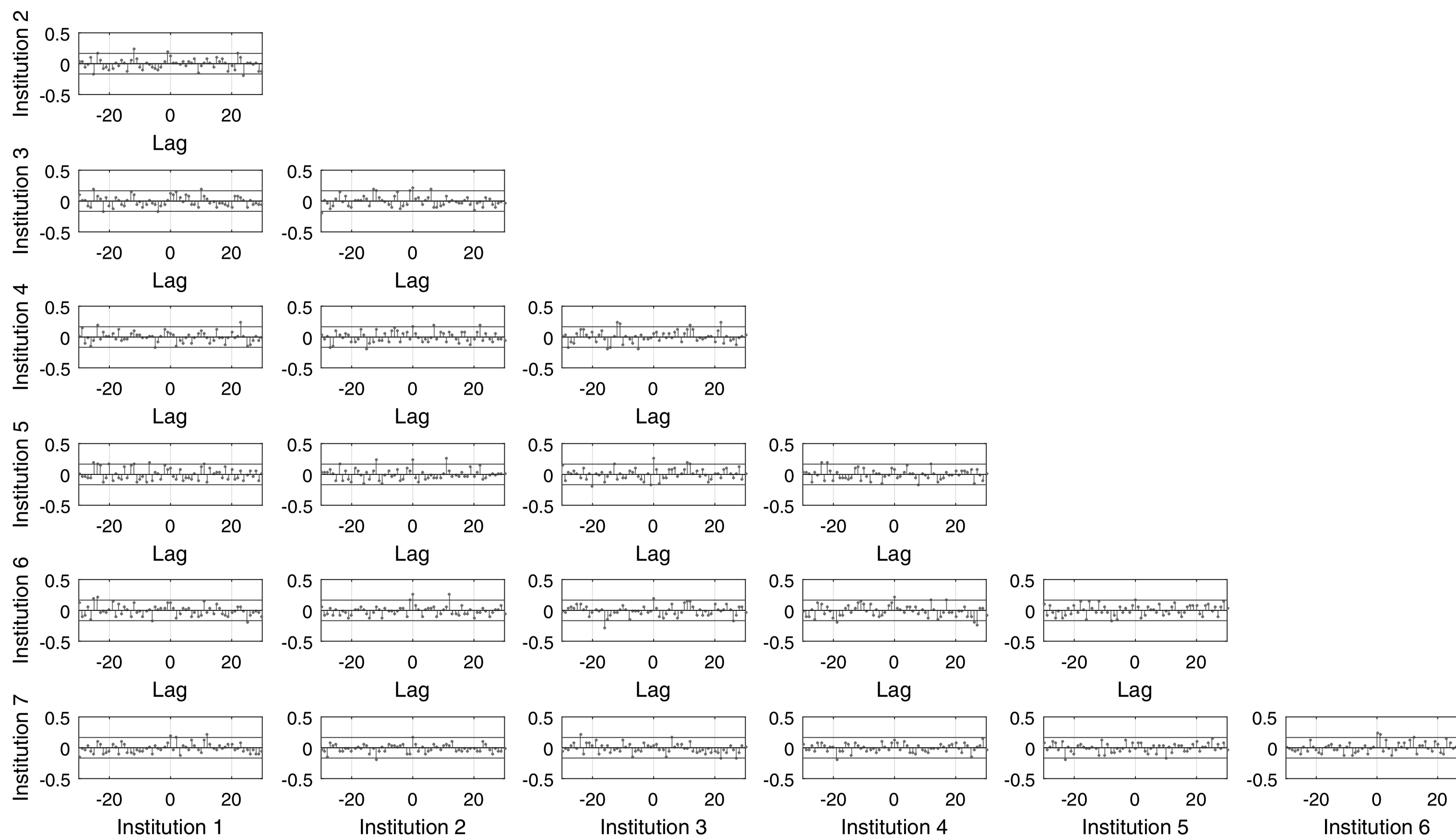
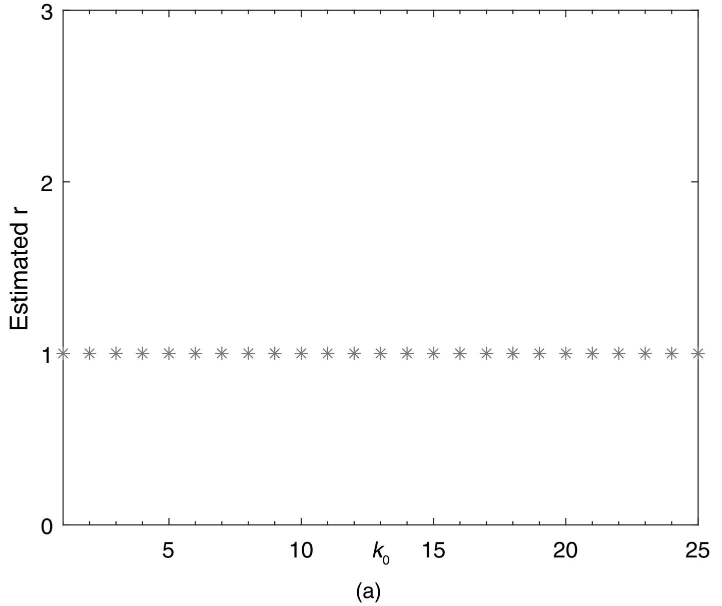


Fig. 5. Cross-correlograms of (prewhitened) $\hat{\eta}_{j,t}$, $j = 1, 2, \ldots, 7$: the figure depicts the cross-correlations between $\hat{\eta}_{j,t}$ $(j \neq 1, 5, 6, 7)$ and prewhitened $\hat{\eta}_{1,t}$, $\hat{\eta}_{5,t}$, $\hat{\eta}_{6,t}$ and $\hat{\eta}_{7,t}$; the plot at row j and column k graphs the cross-correlation between institution $j + 1$ and institution k



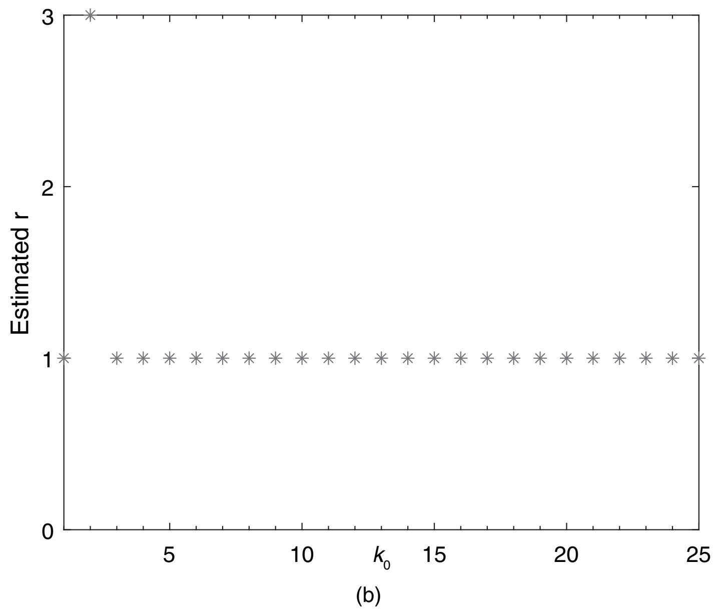


Fig. 6. Estimation of r, the number of the latent factors: here, the estimation of r (vertical axis) is obtained by using equation (18) with the two choices of C_n (a) $C_n = n^{-1}\log(n)$ and (b) $C_n = n^{-1}\log\{\log(n)\}$; the eigenvalues $\hat{\lambda}_j$s are from the matrix $\hat{L}$ defined in equation (13) for k_0 ranging from 1 to 25 (horizontal axis)

We first use dummy variables to model the seasonal variation. We start with 11 dummy variables and the covariate x_t in equation (5) is 12 dimensional including an intercept. The coefficients β_j $(j=1,2,\ldots,7)$ are estimated via the PML approach as described in Section 3.1 and standard errors are calculated by using the approaches that are outlined in Appendix A. It turns out that the amounts of grants received in February, April and October are not significantly

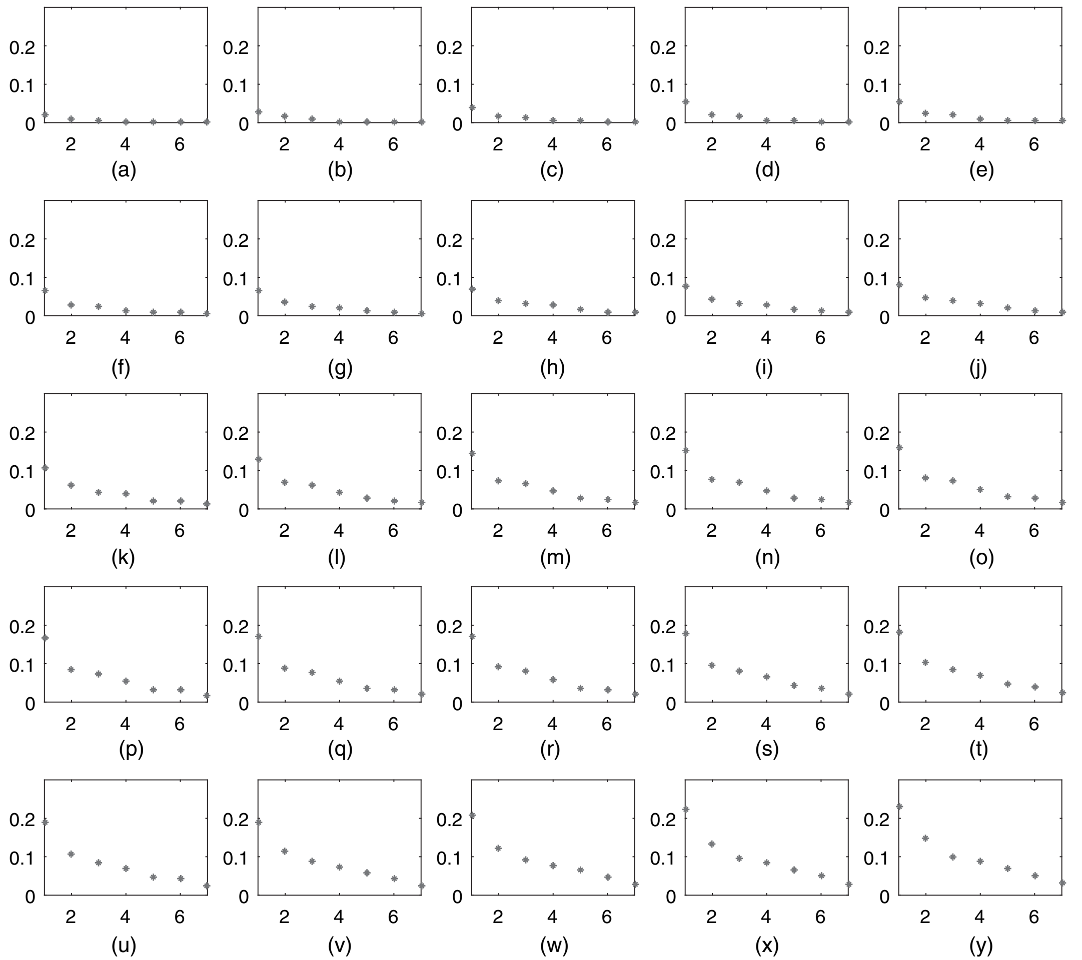


Fig. 7. Eigenvalues of $\hat{L}$: each subplot depicts the seven eigenvalues (in descending order) of the matrix $\hat{L}$ defined in equation (13) for k_0 = (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g) 7, (h) 8, (i) 9, (j) 10, (k) 11, (l) 12, (m) 13, (n) 14, (o) 15, (p) 16, (q) 17, (r) 18, (s) 19, (t) 20, (u) 21, (v) 22, (w) 23, (x) 24, (y) 25

different from those in January by examining their t-values at a 5% level. Hence, we end up with eight dummy variables, and this yields $q=9$ covariates in the modulating factors. Fig. 2 depicts the estimated modulating factors, $\exp(\hat{\beta}_j^{\mathrm{T}} x_t)$, which are also known as seasonality factors, for each institution over 1 year. The seasonally adjusted series $\hat{\eta}_{jt} = \exp(-\hat{\beta}_j^{\mathrm{T}} x_t) y_{jt}$, $j=1,2,\ldots,7$, for the full sample are displayed in Fig. 3. After the regular calendar effect has been removed, other interesting patterns emerge. Many institutions exhibit a peak in 2006 and late 2010.

The auto-correlations for each series are shown in Fig. 4. Among the private universities, apart from institution 1 that has a marginally significant spike at lag 1, the others behave like white noise. The public institutions, however, show moderately significant serial correlation more than 6 months apart: institutions 5 and 7 have a spike at lag 11 whereas institution 6 has significant spikes at lags 8, 26 and 27.

To understand the interinstitution dependence better, we first prewhiten $\hat{\eta}_{1,t}$, $\hat{\eta}_{5,t}$, $\hat{\eta}_{6,t}$ and $\hat{\eta}_{7,t}$ that correspond to institutions 1, 5, 6 and 7, and then compute the cross-correlation along with other series. The cross-correlograms are displayed in Fig. 5, where the plot at row j and column k

graphs the cross-correlation between institution $j+1$ and institution k; more precisely, it reports the correlation between (prewhitened) $\hat{\eta}_{j+1,t}$ and (prewhitened) $\hat{\eta}_{k,t+h}$. Fig. 5 reveals significant interactions between the private institutions, and between the private and public institutions. For instance, institution 1 leads institution 2 by 12 months, but it lags institution 4 by 23 months; institution 2 lags institution 6 by 12 months; institution 3 leads institutions 6 and 7 by 16 and 24 months respectively; it has strong concurrent correlation with institution 5; moreover, there are significant lead-and-lag effects between institutions 2 and 5 and institutions 3 and 4 beyond 10 months. Nonetheless, only weak interactions are observed between the public universities.

Because the matrix $\hat{L}$ that is defined in equation (13) accumulates the interdependence and cross-dependence among the data, we compute $\hat{L}$ with k_0 ranging from 1 to 25 by virtue of Fig. 5. The estimated r, $\hat{r}$, with both $C_n = n^{-1}\log(n)$ and $C_n = n^{-1}\log\{\log(n)\}$, are summarized in Fig. 6. $C_n = n^{-1}\log(n)$ consistently yields $\hat{r}=1$. We also graph the eigenvalues of the matrix $\hat{L}$ in Fig. 7. In view of both Figs 6 and 7, we choose $C_n = n^{-1}\log(n)$ and $k_0 = 15$. So one common factor is identified, i.e. $\hat{r}=1$. The estimated latent factor $\tilde{f}_t$ is plotted in the bottom panel of Fig. 3. It shows peaks in late 2006 and 2010, which are in accordance with the incidences of the NSF FY2006-2011 strategic plan and the stimulus package that was effective from February 2009. The estimated factor loadings across institutions 1–7 are respectively 0.41, 0.32, 0.33, 0.21, 0.65, 0.25 and 0.30. Institution 5 loads heavily on the factor, whereas institutions 2, 3 and 7 load roughly equally on the common factor. In the left-hand panel of Fig. S.15 in the on-line supplemental material, we also plot the serial correlation of $\tilde{f}_t$ up to lag 40. There are two significant spikes at lags 11 and 12, and a marginally significant spike at lag 4. To gain a better insight into the factor, we fit an auto-regressive model of order 13 to the logarithmically transformed series $\log(\tilde{f}_t)$ after examining their auto-correlation, partial auto-correlation and extended auto-correlation. The fitted model is

$$(1-0.2314B^{11}-0.2215B^{13})\{\log(\tilde{f}_t)-0.847\}=a_t, \tag{27}$$

where a_t is white noise with variance 0.089 61. All the parameter estimates are significant at the 5% level. The residuals' auto-correlation plot is depicted in the right-hand panel of Fig. S.15 in the supplemental material. The Box–Ljung statistic for assessing the white noise assumption of residual auto-correlation up to lag 15 is 14.787 with 13 degrees of freedom and a p-value of 0.3208. The auto-regressive polynomial that was obtained from equation (27) has six pairs of complex roots, which indicate cycles of roughly 3 months, 6 months and 2 years.

Acknowledgements

The authors thank the Joint Editor, the Associate Editor and the referees for their helpful comments. The authors also thank Professor Hayne for kindly providing the data. The research of Haonan Wang was partially supported by NSF grants DMS-1521746 and DMS-1737795.

Appendix A: Asymptotic variance calibration

When applying the proposed framework to real data, a practical issue arises—how to compute the standard error of $\hat{b}$. This appendix discusses the calibration of the asymptotic covariance matrix in theorem 1.

Because the conditional distribution of λ_{jt} given ϵ_t is left unspecified, it is, in general, impossible to find a tractable analytic expression for the asymptotic covariance matrix in theorem 1. Now that $E\{\text{var}(\lambda_{jt}|\epsilon_t)\} = \text{var}(y_{jt}) - \exp(\beta_{0j}^{\mathrm{T}}x_t) - \exp(2\beta_{0j}^{\mathrm{T}}x_t)\Sigma_\epsilon^{(j,j)}(0)$, we have

$$\Omega_2^{(j)} = \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} x_t x_t^{\mathrm{T}}\text{var}(y_{jt}) - \Omega_1^{(j)} - W_0^{(j,j)}\Sigma_\epsilon^{(j,j)}(0). \tag{28}$$

Then the diagonal block (j, j) of $\Omega_1+\Omega_2+\Omega_3$ is

$$\Omega_1^{(j)}+\Omega_2^{(j)}+\Omega_3^{(j,j)}=\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} x_t x_t^{\mathrm{T}}\operatorname{var}(y_{jt})+\sum_{k\neq 0} W_k^{(j,j)}\Sigma_\epsilon^{(j,j)}(k), \tag{29}$$

and its off-diagonal block is

$$\Omega_3^{(j,j')}=\sum_{k=-\infty}^{\infty} W_k^{(j,j')}\Sigma_\epsilon^{(j,j')}(k), \qquad j\neq j'. \tag{30}$$

Both equation (29) and expression (30) provide a theoretical basis for estimating the asymptotic covariance.

Define $\hat{\Omega}_1^{(j)}=n^{-1}\Sigma_{t=1}^{n} x_t x_t^{\mathrm{T}}\exp(\hat{\beta}_j^{\mathrm{T}} x_t)$ and

$$\hat{W}_k^{(j,j')}=n^{-1}\sum_{t=\max(1,1-k)}^{\min(n-k,n)} x_{t+k}x_t^{\mathrm{T}}\exp(\hat{\beta}_j^{\mathrm{T}} x_{t+k}+\hat{\beta}_{j'}^{\mathrm{T}} x_t), \qquad k=0, \pm 1, \pm 2, \ldots,$$

for $j, j'=1, \ldots, p$. Because y_{jt} is not stationary, its variance is time dependent. As in Zeger (1988), we use $\{y_{jt}-\exp(\hat{\beta}_j^{\mathrm{T}} x_t)\}^2$ as a proxy for $\operatorname{var}(y_{jt})$. In view of equations (29) and (9), $\Omega_2^{(j)}+\Omega_3^{(j,j)}$ is estimated by

$$\hat{\Sigma}^{(j,j)}=-\hat{\Omega}_1^{(j)}+n^{-1}\sum_{t=1}^{n} x_t x_t^{\mathrm{T}}\{y_{jt}-\exp(\hat{\beta}_j^{\mathrm{T}} x_t)\}^2+\sum_{k\neq 0}\hat{W}_k^{(j,j)}\hat{\Sigma}_\eta^{(j,j)}(k), \tag{31}$$

and $\Omega_3^{(j,j')}$ by $\hat{\Sigma}^{(j,j')}=\Sigma_{|k|<n}\hat{W}_k^{(j,j')}\hat{\Sigma}_\eta^{(j,j')}(k)$ for $j\neq j$, where $\hat{\Sigma}_\eta^{(j,j')}(k)$ refers to the (j, j') element of the matrix $\hat{\Sigma}_\eta(k)$. More precisely,

$$\hat{\Sigma}_\eta(k)=\frac{1}{n}\sum_{t=\max(1,1-k)}^{\min(n-k,n)}(\hat{\eta}_{t+k}-\bar{\eta})(\hat{\eta}_t-\bar{\eta})^{\mathrm{T}}, \qquad k=0, \pm 1, \pm 2, \ldots.$$

Let $\hat{\Sigma}=(\hat{\Sigma}^{(j,j')})$. The matrix $\hat{\Sigma}$ may not be a valid estimator in that $\Omega_2+\Omega_3$ is positive definite but $\hat{\Sigma}$ may not be. To produce a non-negative definite estimator, we write $\hat{\Sigma}=\Gamma\Lambda\Gamma^{\mathrm{T}}$, where Γ is an orthogonal matrix and Λ is diagonal, and replace the negative entries in Λ with 0. The resulting matrix is denoted by Λ^+. A non-negative definite estimator of $\Omega_2+\Omega_3$ is given by $\Gamma\Lambda^+\Gamma^{\mathrm{T}}$. This practice has been adopted by Fan *et al.* (2012). Thus, $\Omega_1^{-1}+\Omega_1^{-1}(\Omega_2+\Omega_3)\Omega_1^{-1}$ can be estimated by using the formula

$$\hat{\Omega}_1^{-1}+\hat{\Omega}_1^{-1}\Gamma\Lambda^+\Gamma^{\mathrm{T}}\hat{\Omega}_1^{-1}, \tag{32}$$

where $\hat{\Omega}_1=\operatorname{diag}(\hat{\Omega}_1^{(1)}, \hat{\Omega}_1^{(2)}, \ldots, \hat{\Omega}_1^{(p)})$.

Note that $\Omega_1^{-1}+\Omega_1^{-1}(\Omega_2+\Omega_3)\Omega_1^{-1}=\Omega_1^{-1}(\Omega_1+\Omega_2+\Omega_3)\Omega_1^{-1}$. An alternative approach is to estimate $\Omega_1^{(j)}+\Omega_2^{(j)}+\Omega_3^{(j,j)}$ as an entity (considering equation (29)) and then to construct a non-negative definite matrix estimator of $\Omega_1+\Omega_2+\Omega_3$ by using the aforementioned decomposition. This approach is not recommended, however, because it seriously underestimates the standard deviations of $\hat{b}$. It does not take into account the positive definiteness of the matrices Ω_1, Ω_2 and Ω_3. By the same token, the estimator (32) also potentially underperformed. Ideally, one should construct non-negative definite estimators of Ω_2 and Ω_3. Nevertheless, this is not feasible because of $\Sigma_\epsilon^{(j,j)}(0)$—see equation (28). Recall that estimation for the variance of ϵ_t that was presented in Zeger (1988), Davis *et al.* (2000) and Davis and Wu (2009) relies on distributional assumptions on ϵ_t, which are not specified in the present context.

References

Brännäs, K. and Johansson, P. (1994) Time series count data regression. *Communs Statist. Theory Meth.*, **23**, 2907–2925.

Campbell, M. J. (1994) Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *J. R. Statist. Soc.* A, **157**, 191–208.

Chan, K. and Ledolter, J. (1995) Monte Carlo EM estimation for time series models involving counts. *J. Am. Statist. Ass.*, **90**, 242–252.

Chang, J., Guo, B. and Yao, Q. (2015) High dimensional stochastic regression with latent factors, endogeneity and nonlinearity. *J. Econmetr.*, **189**, 297–312.

Davis, R. A., Dunsmuir, W. T. and Wang, Y. (2000) On autocorrelation in a Poisson regression model. *Biometrika*, **87**, 491–505.
Davis, R. A. and Wu, R. (2009) A negative binomial model for time series of counts. *Biometrika*, **96**, 735–749.
Engle, R. F. and Rangel, J. G. (2008) The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Rev. Finan. Stud.*, **21**, 1187–1222.
Fan, J., Li, Y. and Yu, K. (2012) Vast volatility matrix estimation using high-frequency data for portfolio selection. *J. Am. Statist. Ass.*, **107**, 412–428.
Fokianos, K., Rahbek, A. and Tjøstheim, D. (2009) Poisson autoregression. *J. Am. Statist. Ass.*, **104**, 1430–1439.
Gourieroux, C., Monfort, A. and Trognon, A. (1984a) Pseudo maximum likelihood methods: theory. *Econometrica*, **52**, 681–700.
Gourieroux, C., Monfort, A. and Trognon, A. (1984b) Pseudo maximum likelihood methods: applications to Poisson models. *Econometrica*, **52**, 701–720.
Jørgensen, B., Lundbye-Christensen, S., Song, P.-K. and Sun, L. (1999) A state space model for multivariate longitudinal count data. *Biometrika*, **86**, 169–181.
Jung, R. C., Liesenfeld, R. and Richard, J.-F. (2011) Dynamic factor models for multivariate count data: an application to stock-market trading activity. *J. Bus. Econ. Statist.*, **29**, 73–85.
Lahiri, S. N. (2013) *Resampling Methods for Dependent Data*. New York: Springer Science and Business Media.
Lam, C. and Yao, Q. (2012) Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.*, **40**, 694–726.
Lam, C., Yao, Q. and Bathia, N. (2011) Estimation of latent factors for high-dimensional time series. *Biometrika*, **98**, 901–918.
Pan, J. and Yao, Q. (2008) Modelling multiple time series via common factors. *Biometrika*, **95**, 365–379.
Rydberg, T. H. and Shephard, N. (2000) A modelling framework for the prices and times of trades made on the New York stock exchange. In *Nonlinear and Nonstationary Signal Processing* (eds W. J. Fitzgerald, R. L. Smith, A. T. Walden and P. C. Young), pp. 217–246. Cambridge: Cambridge University Press.
Wedel, M., Böckenholt, U. and Kamakura, W. A. (2003) Factor models for multivariate count data. *J. Multiv. Anal.*, **87**, 356–369.
Xia, Q., Xu, W. and Zhu, L. (2015) Consistently determining the number of factors in multivariate volatility modelling. *Statist. Sin.*, **25**, 1025–1044.
Zeger, S. L. (1988) A regression model for time series of counts. *Biometrika*, **75**, 621–629.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Supplemental material for “Modeling non-stationary multivariate time series of counts via common factors”’.