WORKING PAPER

NEAREST COMOMENT ESTIMATION WITH UNOBSERVED FACTORS

Kris Boudt
Dries Cornilly
Tim Verdonck

April 2019
2019/970

GHENT UNIVERSITY

Department of Economics

Faculty of Economics and Business Administration
Campus Tweekerken, St.-Pietersplein 5, 9000 Ghent - BELGIUM

Nearest comoment estimation with unobserved factors

Kris Boudt
Department of Economics, Universiteit Gent
Department of Business, Vrije Universiteit Brussel
School of Business and Economics, Vrije Universiteit Amsterdam
and
Dries Cornilly*
Department of Business, Vrije Universiteit Brussel
Department of Mathematics, KU Leuven
and
Tim Verdonck
Department of Mathematics, Universiteit Antwerpen
Department of Mathematics, KU Leuven

April 3, 2019

Abstract

We propose a minimum distance estimator for the higher-order comoments of a multivariate distribution exhibiting a lower dimensional latent factor structure. We derive the influence function of the proposed estimator and prove its consistency and asymptotic normality. The simulation study confirms the large gains in accuracy compared to the traditional sample comoments. The empirical usefulness of the novel framework is shown in applications to portfolio allocation under non-Gaussian objective functions and to the extraction of factor loadings in a dataset with mental ability scores.

Keywords: Higher-order multivariate moments; latent factor model; minimum distance estimation; risk assessment; structural equation modelling.

JEL: C100; C130; C510

*Corresponding author: Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium. (email: dries.cornilly@kuleuven.be, tel: +32 16 37 77 82)

1 Introduction

Skewness and kurtosis are widely used statistics to describe the non-normality of distributions. In multivariate analysis, they help to characterize non-linear dependence and to identify latent factors (Mooijaart (1985), Bonhomme & Robin (2009)). The assumption of a lower dimensional latent factor structure is widespread in multivariate analysis of economic and financial time series; see for example Bai & Ng (2013), Ross (1976) and Fan et al. (2017). In this regard, as mentioned by Ghalanos et al. (2015), an unresolved research question is how to exploit the latent factor structure for higher-order comoment estimation. Better estimates could improve any application that takes these comoments as input. Such applications include the identification of factor loadings in a noisy independent component model (Bonhomme & Robin (2009)), portfolio allocation based on non-Gaussian objective functions (Briec et al. (2007), Harvey et al. (2010) and Boudt et al. (2013)), risk measurement (Zangari (1996) and Stoyanov et al. (2013)) and factor analysis (Mooijaart (1985)).

To solve the problem of exploiting latent factor models for improved estimation of higher-order comoments, we propose a minimum distance estimation approach that involves exploiting the non-linearities in the higher-order comoments under a latent factor model. Based on the sample comoments, the nearest comoment estimates are taken as the structured comoment matrices closest to the sample comoments in a weighted quadratic loss while respecting the structure of a latent factor model. We derive the influence function of the proposed estimator and prove consistency and asymptotic normality based on the theory of minimum distance estimation (Newey & McFadden (1994)).

An advantage of our approach is that there is no pre-estimation of the factors or factor loadings involved, which is usually required when estimating a latent factor model (see e.g., Bai (2003) and Chen et al. (2018)). Moreover, our aim of exploiting the latent factor structure does not require the dimension to grow, as would be the case when estimating the factors (Lawley & Maxwell (1963)). A computational disadvantage of the proposed estimator is the use of a weight matrix of dimension equal to the number of sample moments, thus growing as $O(p^4)$. In the simulations and empirical applications, we therefore consider only settings of up to $p = 10$, which is still highly relevant to asset allocation (Boudt et al. (2013)) and to identifying loadings in factor analysis (Yuan & Chan (2016)). Future advances in computing technology will further enlarge the scope of applications in terms of choice of p.

Since the proposed nearest comoment (NC) estimator depends on higher moments of different orders, it is important to regularize the influence of each moment order onto the final estimates. To do so, we propose to use a ridge-based weight matrix (Yuan & Chan (2016) and Yuan et al. (2017)), together with a bootstrap procedure for selecting the optimal regularization constant.

We illustrate the novel framework in two settings. First, we show the economic gains of the proposed NC estimates of the higher-order comoments for portfolio allocation under non-Gaussian objective functions. Second, we use the proposed methodology for factor extraction in the Holzinger & Swineford (1939) dataset with mental ability scores of seventh- and eight-grade pupils. Our procedure makes use of the non-normality in the latent variables to uniquely identify the matrix of factor loadings. A promax rotation (Hendrickson & White (1964)) then confirms the traditional pattern in the explanatory variables, and we uncover a new latent relation between the 'visual' and 'textual' factors.

The remainder of the paper is organised as follows: Section 2 introduces the theoretical framework. In Section 3 we derive the nearest comoment estimator, and in Section 4 its asymptotic properties are worked out. Section 5 provides the practical guidelines for the method. An extensive simulation study, presented in Section 6, examines all aspects of the NC estimator in different settings. In Section 7 we illustrate the practical usefulness of the proposed methodology for portfolio optimization and factor selection under non-normal distributions. We end the paper with a conclusion and with suggestions for further research.

A supplementary appendix provides more detail about the shape of the influence function in a single-factor model and contains additional simulation results, including a misspecified model. We provide examples of moment expansions that are relevant in economics and finance and explain how the theory changes when the mean is assumed to be known. Finally, we demonstrate the R code for our estimator, which is available publicly in the PerformanceAnalytics package of Peterson & Carl (2018).

2 Framework

As in Mooijaart (1985), the latent factor model is represented semi-parametrically, based on a vector of structural parameters $\boldsymbol{\theta}$. We adopt the same notation for the higher-order comoment matrices as in Jondeau & Rockinger (2006), Martellini & Ziemann (2010) and Boudt et al. (2015), among others.

2.1 Parameters of interest

Consider a p-dimensional random vector $\boldsymbol{X} \in \mathbb{R}^p$ with mean $\boldsymbol{\mu}$ and finite fourth-order moments. The covariance, coskewness and cokurtosis matrices are defined by

$$\begin{aligned} \boldsymbol{\Sigma} &= \mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})'\right], \\ \boldsymbol{\Phi} &= \mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})' \otimes (\boldsymbol{X}-\boldsymbol{\mu})'\right], \\ \boldsymbol{\Psi} &= \mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})' \otimes (\boldsymbol{X}-\boldsymbol{\mu})' \otimes (\boldsymbol{X}-\boldsymbol{\mu})'\right], \end{aligned} \tag{1}$$

where $\otimes$ denotes the Kronecker product. Denote by $\boldsymbol{\sigma} = (\sigma_{\{11\}} \quad \sigma_{\{12\}} \quad \cdots \quad \sigma_{\{pp\}})'$ the vector that stacks the unique covariance elements in order of increasing indices. Here, $\sigma_{\{ij\}}$ equals the covariance between X_i and X_j. Analogously, the vectors $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ contain the unique elements $\phi_{\{ijk\}}$ and $\psi_{\{ijkl\}}$, respectively, in order of increasing indices. The vector $\boldsymbol{\sigma}$ has $p(p+1)/2$ elements, while $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ have $p(p+1)(p+2)/6$ and $p(p+1)(p+2)(p+3)/24$ elements. Note that these vectors have fewer elements than the matrices in (1) due to symmetries in the matrix representation. Combine all unique second, third and fourth-order central moments into the column vector $\boldsymbol{\zeta}$ as

$$\boldsymbol{\zeta} = (\boldsymbol{\sigma}' \quad \boldsymbol{\phi}' \quad \boldsymbol{\psi}')'. \tag{2}$$

2.2 Semi-parametric model

In order to model the moments up to the fourth order of $\boldsymbol{X}$, we employ a semi-parametric model $\mathcal{P}_{\boldsymbol{\theta}} = \{\mathbb{P}_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta} \subset \mathbb{R}^{\kappa}\}$, with $\boldsymbol{\Theta}$ a compact set and κ denoting the dimension of $\boldsymbol{\theta}$. The observed random variable $\boldsymbol{X}$ is defined through the equation

$$\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{F} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\boldsymbol{F}$ are the unobserved factors of dimension $q < p$. The matrix with factor loadings $\boldsymbol{B} \in \mathbb{R}^{p \times q}$ is of full column rank and $\boldsymbol{\varepsilon} \in \mathbb{R}^p$ denotes the idiosyncratic term. We further restrict the structure of the model by assuming that the factors $\boldsymbol{F}$ are independent, have mean zero and unit variance, the idiosyncratic term $\boldsymbol{\varepsilon}$ has independent components and factors and idiosyncratic terms are mutually independent. These assumptions are common in the literature in order to limit the number of free parameters (see e.g., Mooijaart (1985)). Additionally, $\boldsymbol{F}$ and $\boldsymbol{\varepsilon}$ are assumed to have finite eighth-order moments in order to have all necessary moments finite, which is required for the asymptotic analysis in Section 4.

This model is very flexible since it allows up to $p-1$ independent common factors with an additional idiosyncratic term for each variable. Moreover, the model leaves room for correlated but dependent components due to the codependencies implied by the latent factors. However, our assumptions do not allow for higher-order dependence between different variables of $\boldsymbol{\varepsilon}$ and $\boldsymbol{F}$. Doing so would raise the number of parameters too much. We remark that $\boldsymbol{X}$ is not restricted to the observed variable, but may be a transformation of interest, as in Luciani & Veredas (2015) and Barigozzi & Hallin (2016, 2017). Finally, note that under our semi-parametric model, we do not distinguish between two distributions if they have equal moments up to the fourth order. Hence, some of the independence assumptions can be weakened.

Under the semi-parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ in (3), the covariance, coskewness and cokurtosis matrices are

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{\theta}} &= \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{\Delta}, \\
\boldsymbol{\Phi}_{\boldsymbol{\theta}} &= \boldsymbol{B}\boldsymbol{\Phi}_{\boldsymbol{F}}\left(\boldsymbol{B}' \otimes \boldsymbol{B}'\right) + \boldsymbol{\Omega}, \\
\boldsymbol{\Psi}_{\boldsymbol{\theta}} &= \boldsymbol{B}\boldsymbol{\Psi}_{\boldsymbol{F}}\left(\boldsymbol{B}' \otimes \boldsymbol{B}' \otimes \boldsymbol{B}'\right) + \boldsymbol{\Gamma},
\end{aligned} \tag{4}$$

where $\boldsymbol{\Phi}_{\boldsymbol{F}}$ and $\boldsymbol{\Psi}_{\boldsymbol{F}}$ are the coskewness and cokurtosis matrices of the factors. The matrix $\boldsymbol{B}$ contains the factor loadings, $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}$ are the covariance and coskewness matrices of the idiosyncratic term and $\boldsymbol{\Gamma}$ contains the residual cokurtosis elements not explained by the factors. This structure is derived in Boudt et al. (2015) and is provided in the supplementary appendix for completeness. In this representation, the vector $\boldsymbol{\theta}$ with structural parameters equals

$$\boldsymbol{\theta} = (\mathrm{vec}(\boldsymbol{B})' \quad \boldsymbol{\phi}_{\boldsymbol{F}}' \quad \boldsymbol{\psi}_{\boldsymbol{F}}' \quad \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}' \quad \boldsymbol{\phi}_{\boldsymbol{\varepsilon}}' \quad \boldsymbol{\psi}_{\boldsymbol{\varepsilon}}')', \tag{5}$$

Table 1: Number of elements in $\boldsymbol{\zeta}$ and $\boldsymbol{\theta}$ depending on dimensions p and q.

p		2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\#\boldsymbol{\zeta}$		12	31	65	120	203	322	486	705	990	1353	1807	2366	3045	3860
$\#\boldsymbol{\theta}$	$(q = 1)$	10	14	18	22	26	30	34	38	42	46	50	54	58	62
	$(q = 2)$		19	24	29	34	39	44	49	54	59	64	69	74	79
	$(q = 3)$			30	36	42	48	54	60	66	72	78	84	90	96
	$(q = 4)$				43	50	57	64	71	78	85	92	99	106	113

Note: The vector $\boldsymbol{\zeta}$ contains the unique second, third and fourth-order central moments of a random variable $\boldsymbol{X}$ of dimension p. The number of parameters in the multi-factor model equals $\#\boldsymbol{\theta} = p(q+3)+2q$.

where $\boldsymbol{\phi_F}, \boldsymbol{\psi_F} \in \mathbb{R}^q$ are the marginal third and fourth-order central moments of $\boldsymbol{F}$ and $\boldsymbol{\sigma_\varepsilon}, \boldsymbol{\phi_\varepsilon}, \boldsymbol{\psi_\varepsilon} \in \mathbb{R}^p$ the marginal second, third and fourth-order moments of $\boldsymbol{\varepsilon}$. Hence, the number of structural parameters in $\boldsymbol{\theta}$ equals

$$\#\boldsymbol{\theta} = p(q+3) + 2q, \tag{6}$$

which is less than the number of unique comoments up to fourth order if $q < p$.

Similarly as in Equation (2), the unique covariance, coskewness and cokurtosis elements under the multi-factor model are gathered in the vectors $\boldsymbol{\sigma_\theta}$, $\boldsymbol{\phi_\theta}$ and $\boldsymbol{\psi_\theta}$ and combined to

$$\boldsymbol{\zeta_\theta} = \begin{pmatrix}\boldsymbol{\sigma}'_{\boldsymbol{\theta}} & \boldsymbol{\phi}'_{\boldsymbol{\theta}} & \boldsymbol{\psi}'_{\boldsymbol{\theta}}\end{pmatrix}'. \tag{7}$$

To illustrate the dimension reduction implied by model $\mathcal{P}_{\boldsymbol{\theta}}$, the number of parameters in $\boldsymbol{\theta}$ needed to determine the moments $\boldsymbol{\zeta_\theta}$ is given in Table 1 for different values of p and q.

The asymptotic results of the nearest comoment estimator presented in Section 3 require the vector $\boldsymbol{\theta}$ to be identifiable. Conditions for identifiability are given in Mooijaart (1985), where the following theorem is proved.

Theorem 2.1. *If the factors $\boldsymbol{F}$ in (3) are non-Gaussian, and there are no two factors with the same skewness and kurtosis, then the parameter vector $\boldsymbol{\theta}$ is identifiable up to trivial permutations and sign-changes.*

The remainder of this paper takes the assumption of identifiability for the semi-parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ with true parameter $\boldsymbol{\theta}^\star$ in the interior of $\boldsymbol{\Theta}$, with $\boldsymbol{\Theta}$ a compact subspace of $\mathbb{R}^\kappa$.

3 Nearest comoment estimator

This section introduces the nearest comoment (NC) estimator in full generality. Asymptotic properties of the estimator are derived in Section 4 and practical considerations are discussed in Section 5.

The most intuitive way to estimate the moments in $\boldsymbol{\sigma}$, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$ is by the plug-in sample comoments. Let $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ with $\boldsymbol{x}_i \in \mathbb{R}^p$ be a sample of n independent and identically distributed p-dimensional vectors drawn from the distribution of a random variable $\boldsymbol{X}$ with finite fourth-order moments. Replacing each expected value by a sample average, the sample comoments are defined as

$$\begin{aligned}
\widehat{\sigma}_{s,\{ij\}} &= \frac{1}{n}\sum_{m=1}^{n}\left(x_{mi}-\overline{x}_i\right)\left(x_{mj}-\overline{x}_j\right), \\
\widehat{\phi}_{s,\{ijk\}} &= \frac{1}{n}\sum_{m=1}^{n}\left(x_{mi}-\overline{x}_i\right)\left(x_{mj}-\overline{x}_j\right)\left(x_{mk}-\overline{x}_k\right), \\
\widehat{\psi}_{s,\{ijkl\}} &= \frac{1}{n}\sum_{m=1}^{n}\left(x_{mi}-\overline{x}_i\right)\left(x_{mj}-\overline{x}_j\right)\left(x_{mk}-\overline{x}_k\right)\left(x_{ml}-\overline{x}_l\right),
\end{aligned} \tag{8}$$

for $i, j, k, l = 1, \ldots, p$ with $i \leq j \leq k \leq l$ and $\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{m=1}^{n}\boldsymbol{x}_m$. We gather the elements $\widehat{\sigma}_{s,\{ij\}}, \widehat{\phi}_{s,\{ijk\}}$ and $\widehat{\psi}_{s,\{ijkl\}} (i \leq j \leq k \leq l)$ into the vectors $\widehat{\boldsymbol{\sigma}}_s$, $\widehat{\boldsymbol{\phi}}_s$ and $\widehat{\boldsymbol{\psi}}_s$ to define the sample comoments

$$\widehat{\boldsymbol{\zeta}}_s = \begin{pmatrix} \widehat{\boldsymbol{\sigma}}_s' & \widehat{\boldsymbol{\phi}}_s' & \widehat{\boldsymbol{\psi}}_s' \end{pmatrix}'. \tag{9}$$

The estimator $\widehat{\boldsymbol{\zeta}}_s$ is consistent, as will be shown in Section 4, but it may have a large estimation variance compared to alternative estimators that utilize the structure of the underlying data-generating model in (3). Therefore, we propose the NC estimator that takes into account the structure of this model by finding the comoment matrices under $\mathcal{P}_{\boldsymbol{\theta}}$ that are nearest to the sample moments in terms of a weighted quadratic loss function (Newey & McFadden (1994)); hence the name of the NC estimator.

Formally, the structural parameters $\widehat{\boldsymbol{\theta}}_{\mathrm{nc}}$ minimize a weighted quadratic distance between the first-step estimate $\widehat{\boldsymbol{\zeta}}_s$ and the model moments $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$,

$$\widehat{\boldsymbol{\theta}}_{\mathrm{nc}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\right)' \widehat{\boldsymbol{W}} \left(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\right), \tag{10}$$

with $\widehat{\boldsymbol{W}}$ a positive semi-definite weight matrix converging in probability to the positive semi-definite matrix $\boldsymbol{W}$. The weight matrix is denoted as an estimate since it may depend on the sample, but this does not have to be the case. A data-driven way to select an optimal weight matrix $\widehat{\boldsymbol{W}}$ is introduced in Section 5. The nearest comoment estimator is then directly obtained as

$$\widehat{\boldsymbol{\zeta}}_{\text{nc}} = \boldsymbol{\zeta}_{\widehat{\boldsymbol{\theta}}_{\text{nc}}}. \tag{11}$$

We remark that, in the above formulation, the number of factors q is assumed to be known. Often this is not the case, and therefore we provide a suitable selection criterion in Section 5.

4 Theoretical results

This section describes the theoretical properties of the sample estimator and the proposed NC estimator. We first derive the influence function under a finite fourth-order moment condition on $\boldsymbol{X}$ with distribution function H; see for example Hampel et al. (2011). We then show consistency and derive the asymptotic covariance matrix under finite eighth-order moments (Newey & McFadden (1994)) when $n \to \infty$. The results for the NC estimator are under the additional assumptions of the semi-parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ in (3) and the identifiability constraints in Theorem 2.1.

4.1 Influence function

The influence functions of the sample moments $\widehat{\sigma}_{s,\{ij\}}$, $\widehat{\phi}_{s,\{ijk\}}$ and $\widehat{\psi}_{s,\{ijkl\}}$ are provided in the following theorem.

Theorem 4.1 (Influence function of the sample comoments)**.** *Assume $\boldsymbol{X}$ has finite fourth-order moments. The influence functions of the estimators $\widehat{\sigma}_{s,\{ij\}}$, $\widehat{\phi}_{s,\{ijk\}}$ and $\widehat{\psi}_{s,\{ijkl\}}$ are*

$$\begin{aligned}
\text{IF}(\boldsymbol{x}; \widehat{\sigma}_{s,\{ij\}}, H) &= (x_i - \mu_i)(x_j - \mu_j) - \sigma_{\{ij\}}, \\
\text{IF}(\boldsymbol{x}; \widehat{\phi}_{s,\{ijk\}}, H) &= (x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k) - \phi_{\{ijk\}} - (x_i - \mu_i)\sigma_{\{jk\}} \\
&\quad - (x_j - \mu_j)\sigma_{\{ik\}} - (x_k - \mu_k)\sigma_{\{ij\}}, \\
\text{IF}(\boldsymbol{x}; \widehat{\psi}_{s,\{ijkl\}}, H) &= (x_i - \mu_i)(x_j - \mu_j)(x_k - \mu_k)(x_l - \mu_l) - \psi_{\{ijkl\}} - (x_i - \mu_i)\phi_{\{jkl\}} \\
&\quad - (x_j - \mu_j)\phi_{\{ikl\}} - (x_k - \mu_k)\phi_{\{ijl\}} - (x_l - \mu_l)\phi_{\{ijk\}}.
\end{aligned} \tag{12}$$

All proofs are provided in Appendix A. The influence function for $\widehat{\boldsymbol{\zeta}}_s$ then equals

$$\mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\zeta}}_s, H) = \Big(\mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\sigma}}_s, H)' \quad \mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\phi}}_s, H)' \quad \mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\psi}}_s, H)'\Big)'. \tag{13}$$

The influence function of $\widehat{\boldsymbol{\zeta}}_{\mathrm{nc}}$ is formulated in the following theorem. The chain rule implies that it is a linear transformation of $\widehat{\boldsymbol{\theta}}_{\mathrm{nc}}$, and hence a linear transformation of the influence function of $\widehat{\boldsymbol{\zeta}}_s$.

Theorem 4.2 (Influence function of the NC estimator)**.** *Assume $\boldsymbol{X}$ has finite fourth-order moments. Under the assumptions of the semi-parametric model $\mathcal{P}_{\boldsymbol{\theta}}$ with true parameter $\boldsymbol{\theta}^\star$ and the assumptions of Theorem 2.1, it holds that the influence function of $\widehat{\boldsymbol{\theta}}_{nc}$ is given by*

$$\mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}_{nc}, H) = \left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1} \boldsymbol{G}'\boldsymbol{W}\, \mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\zeta}}_s, H), \tag{14}$$

with $\boldsymbol{G}$ the Jacobian of $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$, evaluated in $\boldsymbol{\theta}^\star$. The influence function of the NC estimator $\widehat{\boldsymbol{\zeta}}_{nc}$ equals

$$\mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\zeta}}_{nc}, H) = \boldsymbol{G}\, \mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}_{nc}, H) = \boldsymbol{G}\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1} \boldsymbol{G}'\boldsymbol{W}\, \mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\zeta}}_s, H). \tag{15}$$

As for the sample moments, the influence function of the NC estimator is unbounded. However, the choice of $\boldsymbol{W}$ determines whether or not the influence function is dampened for certain values of $\boldsymbol{x}$.

4.2 Asymptotic normality

When eighth-order moments of $\boldsymbol{X}$ exist, the sample moments $\widehat{\boldsymbol{\zeta}}_s$ have asymptotic normal distribution and the asymptotic covariance matrix can be estimated consistently.

Theorem 4.3 (Asymptotic normality of the sample comoments)**.** *Assume $\boldsymbol{X}$ has finite eighth-order moments. The sample moments $\widehat{\boldsymbol{\zeta}}_s$ have the asymptotic normal distribution*

$$\sqrt{n}\left(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Xi}\right), \qquad n \to \infty, \tag{16}$$

where the entries in the asymptotic covariance matrix $\boldsymbol{\Xi}$ are

$$\begin{aligned}
&\mathrm{ACov}(\sqrt{n}\widehat{\sigma}_{s,\{ij\}}, \sqrt{n}\widehat{\sigma}_{s,\{uv\}}) = \mu_{ijuv} - \mu_{ij}\mu_{uv}, \\
&\mathrm{ACov}(\sqrt{n}\widehat{\sigma}_{s,\{ij\}}, \sqrt{n}\widehat{\phi}_{s,\{uvw\}}) = \mu_{ijuvw} - \mu_{ij}\mu_{uvw} - \mu_{iju}\mu_{vw} - \mu_{ijv}\mu_{uw} - \mu_{ijw}\mu_{uv},
\end{aligned} \tag{17}$$

with the central moments defined as

$$\mu_{i_1 i_2 \cdots i_r} = \mathbb{E}\left[(X_{i_1} - \mu_{i_1})(X_{i_2} - \mu_{i_2}) \cdots (X_{i_r} - \mu_{i_r})\right]. \tag{18}$$

The other elements in Ξ *are given in Appendix A.*

Remark 4.1. *The asymptotic covariance matrix* Ξ *can be estimated consistently by considering the pseudo-observations* $\boldsymbol{\zeta}_{\boldsymbol{x}_t}, t = 1, \ldots, n$*, with* $\boldsymbol{\zeta}_{\boldsymbol{x}} = (\boldsymbol{\sigma}_{\boldsymbol{x}}' \quad \boldsymbol{\phi}_{\boldsymbol{x}}' \quad \boldsymbol{\psi}_{\boldsymbol{x}}')'$*, and*

$$\begin{aligned}
\boldsymbol{\sigma}_{\boldsymbol{x},\{ij\}} &= (x_i - \overline{x}_i)(x_j - \overline{x}_j) - \widehat{\sigma}_{s,\{ij\}}, \\
\boldsymbol{\phi}_{\boldsymbol{x},\{ijk\}} &= (x_i - \overline{x}_i)(x_j - \overline{x}_j)(x_k - \overline{x}_k) - \widehat{\phi}_{s,\{ijk\}} \\
&\qquad - (x_i - \overline{x}_i)\widehat{\sigma}_{s,\{jk\}} - (x_j - \overline{x}_j)\widehat{\sigma}_{s,\{ik\}} - (x_k - \overline{x}_k)\widehat{\sigma}_{s,\{ij\}}, \\
\boldsymbol{\psi}_{\boldsymbol{x},\{ijkl\}} &= (x_i - \overline{x}_i)(x_j - \overline{x}_j)(x_k - \overline{x}_k)(x_l - \overline{x}_l) - \widehat{\psi}_{s,\{ijkl\}} - (x_i - \overline{x}_i)\widehat{\phi}_{s,\{jkl\}} \\
&\qquad - (x_j - \overline{x}_j)\widehat{\phi}_{s,\{ikl\}} - (x_k - \overline{x}_k)\widehat{\phi}_{s,\{ijl\}} - (x_l - \overline{x}_l)\widehat{\phi}_{s,\{ijk\}}.
\end{aligned} \tag{19}$$

A positive semi-definite consistent estimator of Ξ *is then given by*

$$\widehat{\Xi} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\zeta}_{\boldsymbol{x}_t}\boldsymbol{\zeta}_{\boldsymbol{x}_t}'. \tag{20}$$

The estimator is positive definite when the sample size n *is larger than the number of unique comoments considered, which is the length of the vector* $\boldsymbol{\zeta}$*. Note that the pseudo-observations are constructed according to* (12) *in order to ensure consistency.*

A similar result holds when the mean is known, and hence the sample mean can be replaced by the true mean when estimating the central moments. A typical situation in which the mean is known is the analysis of financial returns at high frequency where the authors assume the mean to be zero (Lee & Mykland (2007)). In the supplementary appendix we discuss how knowing the mean impacts the asymptotic covariance matrix of the resulting sample moments.

The linear relation between the influence function of the sample moments and the influence function of the NC estimator provides insight into the asymptotic covariance matrix of the NC estimator given in the following theorem.

Theorem 4.4 (Asymptotic normality of the NC estimator)**.** *Assume* $\boldsymbol{X}$ *has finite eighth-order moments. Under the conditions of Theorem 4.2, the estimator* $\widehat{\boldsymbol{\theta}}_{nc}$ *has the asymptotic* *distribution*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{nc}-\boldsymbol{\theta}^{\star}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}\boldsymbol{G}'\boldsymbol{W}\Xi\boldsymbol{W}\boldsymbol{G}\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}\right), \quad n\to\infty, \tag{21}$$

with $\boldsymbol{G}$ the Jacobian of $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$ with respect to $\boldsymbol{\theta}$, evaluated in $\boldsymbol{\theta}^{\star}$. The NC estimator $\widehat{\boldsymbol{\zeta}}_{nc}$ has the asymptotic normal distribution given by

$$\sqrt{n}\left(\widehat{\boldsymbol{\zeta}}_{nc}-\boldsymbol{\zeta}_{\boldsymbol{\theta}^{\star}}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\boldsymbol{G}\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}\boldsymbol{G}'\boldsymbol{W}\Xi\boldsymbol{W}\boldsymbol{G}\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}\boldsymbol{G}'\right), \quad n\to\infty. \tag{22}$$

The asymptotic covariance matrix of $\widehat{\boldsymbol{\zeta}}_{\text{nc}}$ depends on the limit $\boldsymbol{W}$ of the weight matrix $\widehat{\boldsymbol{W}}$, the Jacobian $\boldsymbol{G}$ of $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$ evaluated at $\boldsymbol{\theta}^{\star}$ and the asymptotic covariance matrix Ξ of the sample moments $\widehat{\boldsymbol{\zeta}}_{s}$. When $\boldsymbol{W}=\Xi^{-1}$, the asymptotic covariance matrix of the estimator $\widehat{\boldsymbol{\theta}}_{\text{nc}}$ simplifies to $\boldsymbol{G}\left(\boldsymbol{G}'\Xi^{-1}\boldsymbol{G}\right)^{-1}\boldsymbol{G}'$, in which case the corresponding NC estimator attains the lowest asymptotic variance.

Theorem 4.5 (Asymptotic efficiency)**.** *The estimator $\widehat{\boldsymbol{\theta}}_{nc}$ with $\widehat{\boldsymbol{W}} \xrightarrow{p} \Xi^{-1}$ has the lowest asymptotic variance in the class of all estimators*

$$\left\{\widehat{\boldsymbol{\theta}}_{nc}(\widehat{\boldsymbol{W}})|\widehat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W} \text{ is positive semi-definite}\right\}. \tag{23}$$

The following corollary to Theorem 4.4 shows convergence to a chi-squared distribution when an appropriate weight matrix is used.

Corollary 4.1 (Asymptotic distribution of objective values)**.** *Let $\widehat{\boldsymbol{\zeta}}_{nc}$ be the nearest comoment estimator obtained from $\widehat{\boldsymbol{\theta}}_{nc}$, minimizing* (10) *with a weight matrix $\widehat{\boldsymbol{W}} \xrightarrow{p} \Xi^{-1}$. Then it holds that*

$$n\left(\widehat{\boldsymbol{\zeta}}_{s}-\widehat{\boldsymbol{\zeta}}_{nc}\right)'\widehat{\boldsymbol{W}}\left(\widehat{\boldsymbol{\zeta}}_{s}-\widehat{\boldsymbol{\zeta}}_{nc}\right) \xrightarrow{d} \chi^{2}_{\sharp\boldsymbol{\zeta}-\sharp\boldsymbol{\theta}}, \qquad n\to\infty, \tag{24}$$

where the chi-squared distribution has degrees of freedom equal to the number of unique comoments minus the number of model parameters of the nearest comoment estimator.

Note that this corollary requires that the weight matrix converges in probability to the inverse of the asymptotic covariance matrix Ξ. Goodness-of-fit tests based on Corollary 4.1 and adjusted test statistics, as in Yuan & Bentler (2010), are studied through simulations in the supplementary appendix. However, due to the large dimensionality of $\boldsymbol{\zeta}$, these tests might have considerable size distortions even for moderate values of p. It is outside the scope of this paper to propose finite sample corrections to cope with these distortions.

The sandwich standard errors resulting from the diagonal of the asymptotic covariance matrix of $\boldsymbol{\theta}$ in Theorem 4.4 are available to test the model parameters for significance. These standard errors, however, suffer from the same issues as the goodness-of-fit tests due to dependence on $\widehat{\boldsymbol{\Xi}}$. Note that redefining the latent factors by a linear combination of the independent factors is likely to result in a matrix with factor loadings that is easier to interpret. In this case, the standard errors of elements in the new matrix with factor loadings can be obtained by left and right multiplication of the estimated asymptotic covariance matrix by the appropriate transformation matrix.

5 Practical considerations

So far, we have defined the estimator $\widehat{\boldsymbol{\zeta}}_{\mathrm{nc}}$ for second, third and fourth-order multivariate central moments under the assumption of a latent factor model. Results regarding the asymptotic distribution of the estimator and asymptotic efficiency were provided in the previous section. Two ingredients are left in order to generate a fully functioning estimator: a proposal for the weight matrix $\widehat{\boldsymbol{W}}$ and a criterion on which to determine the number of latent factors.

5.1 The weight matrix $\widehat{\boldsymbol{W}}$

The choice of weight matrix directly impacts the influence function (Theorem 4.2) and asymptotic covariance matrix of the NC estimator (Theorem 4.4) and is thus critical. When the sample size is large enough, $\widehat{\boldsymbol{W}}_A = \widehat{\boldsymbol{\Xi}}^{-1}$ is an optimal choice of weight matrix due to the efficiency result in Theorem 4.5. This choice, however, is not always feasible, leading us to consider the alternative weight matrix $\widehat{\boldsymbol{W}}_D = \mathrm{diag}(\widehat{\boldsymbol{\Xi}})^{-1}$, which ignores the off-diagonal elements in $\widehat{\boldsymbol{\Xi}}$.

Recent advances in distribution-free structural equation modelling indicate that regularization of $\widehat{\boldsymbol{\Xi}}$ before inversion typically increases the finite sample efficiency of the estimator and produces a lower mean squared error (MSE). In Arruda & Bentler (2017) this is done by altering the large and small eigenvalues of $\widehat{\boldsymbol{\Xi}}$, but the more popular choice is to use a ridge penalization as in Yuan & Chan (2016) and Yuan et al. (2017). Therefore, we define the ridge weight matrix as

$$\widehat{\boldsymbol{W}}_R(\alpha) = \left[(1-\alpha)\widehat{\boldsymbol{\Xi}} + \alpha\,\mathrm{diag}(\widehat{\boldsymbol{\Xi}})\right]^{-1}, \qquad \alpha \in [0,1]. \tag{25}$$

The regularization parameter α determines the relative importance of the off-diagonal elements in $\widehat{\boldsymbol{\Xi}}$. A strictly positive value insures that the matrix $\widehat{\boldsymbol{W}}_R(\alpha)$ is invertible, even when $\widehat{\boldsymbol{\Xi}}$ is not positive definite. Yuan & Chan (2016) propose a bootstrap procedure to select the optimal value of α minimizing the simulated MSE of the structural parameters of the latent factor model. Our interest lies mainly in estimating the comoment matrices as accurately as possible, for which we propose a bootstrap procedure where the optimal value of α is determined by minimizing a simulated weighted MSE of the NC moment estimates compared to the sample moment estimates. The MSE is weighted such that it does not depend on the different units of the covariance, coskewness and cokurtosis elements. In addition, we correct for the different cardinality of the covariance, coskewness and cokurtosis elements as in Morton & Lim (2009) and Jondeau et al. (2018). Otherwise, the measure would be dominated by the weighted MSE of the cokurtosis estimates when the dimension increases. The proposed weighting matrix for the MSE equals

$$\widehat{\boldsymbol{W}}_{D,w} = \widehat{\boldsymbol{W}}_D \begin{pmatrix} w_\sigma \boldsymbol{I}_{\#\boldsymbol{\sigma}}/\#\boldsymbol{\sigma} & 0 & 0 \\ 0 & w_\phi \boldsymbol{I}_{\#\boldsymbol{\phi}}/\#\boldsymbol{\phi} & 0 \\ 0 & 0 & w_\psi \boldsymbol{I}_{\#\boldsymbol{\psi}}/\#\boldsymbol{\psi} \end{pmatrix}. \tag{26}$$

The coefficients w_σ, w_ϕ and w_ψ determine the relative importance of the covariance, coskewness and cokurtosis elements. In the simulation study and empirical application we set them to one. We mention, however, that the choice of relative importance should ultimately depend on the application at hand. For example, one might consider the weights from a moment approximation to the investors' utility function (Martellini & Ziemann (2010)). A further alternative is to consider multiple target matrices in (25), allowing for more flexibility in the choice of regularization parameters. We leave this for further research and refer to Boudt et al. (2018) for multi-target shrinkage coskewness estimation.

The bootstrap procedure is as follows. Determine M datasets by sampling with replacement from the original observed dataset. Then, for a grid of α-values, calculate for each of the M samples the NC estimates using weight matrix $\widehat{\boldsymbol{W}}_R(\alpha)$. The value of α that produces the lowest simulated weighted MSE over the unique covariance, coskewness and cokurtosis elements with respect to the sample moments of the original dataset is taken as optimal. The simulated weighted MSE is computed as follows

$$\mathrm{wMSE}(\alpha) = \frac{1}{M} \sum_{m=1}^{M} \left\| \widehat{\boldsymbol{W}}_{D,w}^{1/2} \left(\widehat{\boldsymbol{\zeta}}_{\mathrm{nc},m}(\widehat{\boldsymbol{W}}_R(\alpha)) - \widehat{\boldsymbol{\zeta}}_s \right) \right\|^2 . \tag{27}$$

Typically, the cokurtosis estimates have the highest estimation variance in $\widehat{\boldsymbol{\zeta}}_s$. In some cases, ignoring these elements might positively influence the MSE of the other comoments.

Remark 5.1. *If the fourth moments of the factors are not required for identification, then it is possible to construct the weight matrix such that none of the cokurtosis elements has an influence on the objective value. In this case, $\boldsymbol{X}$ only requires finite sixth-order moments for the NC estimator to be asymptotically normal.*

5.2 Determining the number of factors

Increasing the number of latent factors q results in a better fit to the sample moments. Hence, to achieve parsimony, a criterion on which to evaluate the trade-off between model fit and model simplicity is required in order to select the number of latent factors. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are two popular approaches in which the model fit is penalized by the number of model parameters. In our setting, it is natural to define the AIC and BIC by

$$\begin{aligned} \mathrm{AIC}(\widehat{\boldsymbol{\zeta}}_{\mathrm{nc}}(q)) &= n \left(\widehat{\boldsymbol{\zeta}}_s - \widehat{\boldsymbol{\zeta}}_{\mathrm{nc}} \right)' \widehat{\boldsymbol{W}} \left(\widehat{\boldsymbol{\zeta}}_s - \widehat{\boldsymbol{\zeta}}_{\mathrm{nc}} \right) + 2\left(p(q+3) + 2q\right), \\ \mathrm{BIC}(\widehat{\boldsymbol{\zeta}}_{\mathrm{nc}}(q)) &= n \left(\widehat{\boldsymbol{\zeta}}_s - \widehat{\boldsymbol{\zeta}}_{\mathrm{nc}} \right)' \widehat{\boldsymbol{W}} \left(\widehat{\boldsymbol{\zeta}}_s - \widehat{\boldsymbol{\zeta}}_{\mathrm{nc}} \right) + \left(p(q+3) + 2q\right) \log n. \end{aligned} \tag{28}$$

In accordance with the literature, the model with the lowest AIC or BIC is deemed most appropriate, yielding a trade-off between goodness-of-fit and model simplicity. The relevance of these criteria is shown in a simulation study in the supplementary appendix.

Both information criteria provide data-driven ways in which to select the number of factors. As an alternative approach, we can plot the objective value versus the number of factors. This procedure, which is similar to the scree plot in principal component analysis, usually provides a good indication of the number of latent factors one should use. The scree plot flattens when adding new factors fails to substantially decrease the objective value, indicating that the extra factors are not required.

6 Simulation study

In this section we perform a Monte Carlo study to show the improvements of the NC estimator over the sample comoments in terms of the MSE of the estimated multivariate higher-order moments. The bootstrap procedure in selecting an optimal value for the regularization constant in $\widehat{\boldsymbol{W}}_R$ is examined. Additionally, we use simulations to show the accuracy in terms of MSE when estimating the matrix with factor loadings. The set up is based on (3), calibrated in line with the two empirical applications presented in Section 7. The present section is divided according to the setting, each time introducing the data-generating process and proceeding to display and discuss the relevant results.

In both settings, the sample moments and a PC-based estimator serve as benchmarks by which to evaluate the gains in precision achieved by the proposed NC methodology. The PC-based approach estimates the comoment matrices under the assumption that the first q principal components are observed factors. Treating these scores as observed factors, moment estimates are obtained as in Boudt et al. (2015). We emphasize that the factors are not treated as independent when computing $\boldsymbol{\Phi}_{\boldsymbol{F}}$ and $\boldsymbol{\Psi}_{\boldsymbol{F}}$ contrary to our model assumptions. We denote this approach by PC-FM and set the number of principal components equal to the true number of factors.

The supplementary appendix includes results of the goodness-of-fit tests under both simulation settings and illustrates the performance of the AIC and BIC criteria for determining the number of latent factors. Moreover, we study the influence of higher-order dependence in the idiosyncratic term and consider the case in which p grows with n.

6.1 Setting calibrated on hedge fund returns

Set up. As a first set up, we calibrate a simulation model on the weekly returns of ten common hedge fund indices, in line with our application as presented in Section 7.1. We use the NC estimator to obtain the loadings of a latent factor model as data-generating process; see (3). Factors and idiosyncratic terms are estimated as skewed and heavy-tailed, and hence we expect the Normal-Inverse Gaussian (NIG) distribution to be a good fit. The parameters of the NIG are then determined using the method of moments, as given in Karlis (2002), with the moments provided by the NC estimator. This way, data-generating processes are obtained for dimensions $p = 3, 5$ and 10, which we use for the subsequent study. All parameter values are reported in the supplementary appendix.

Performance results. We generate $M = 1000$ samples for dimensions $p = 3, 5$ and 10. We consider the sample sizes $n = 250, 500$ and 1000 and the cases of a single latent factor and two latent factors. The number of factors in the NC estimator is determined by the AIC criterion when $p = 3, 5$ and by the BIC criterion when $p = 10$. The performance is measured by the MSE of the comoment vectors $\boldsymbol{\sigma}$, $\boldsymbol{\phi}$ and $\boldsymbol{\psi}$, divided by the respective number of elements in each vector. In addition, we present the weighted MSE of $\boldsymbol{W}_D^{1/2}\boldsymbol{\zeta}$, similar to (27), jointly over all moments.

Table 2 confirms the advantage conferred by the proposed NC estimator with the data-driven selection of the number of factors and the regularization parameter, with gains between 15% and 50% on the weighted combination of all moments in $\boldsymbol{\zeta}$. The largest impact is due to the significantly lower MSE when estimating coskewness and cokurtosis elements. For the covariance elements, the results are better than PC-FM and only slightly worse than the sample moments. We observe that the relative efficiency increases with the dimension, with gains up to 75% in dimension $p = 10$ for the cokurtosis elements. We further observe the largest reductions in MSE when the sample size is small, indicating that the finite sample improvements are larger than the asymptotic efficiency gain. Also, results under a single latent factor are better than those under two latent factors. This effect derives from the lower dimension, since there are fewer parameters in $\boldsymbol{\theta}$ to estimate when $q = 1$.

Surprisingly, perhaps, the PC-FM estimator offers no benefits over the sample moments when estimating the coskewness and cokurtosis matrices. An explanation for this poor performance is that the principal components are uncorrelated but are not independent and do not take into account the independence assumption in the idiosyncratic component. Hence, the PC scores do not sufficiently remove the dependence from the idiosyncratic component, resulting in a biased model estimate and dependent idiosyncratic terms.

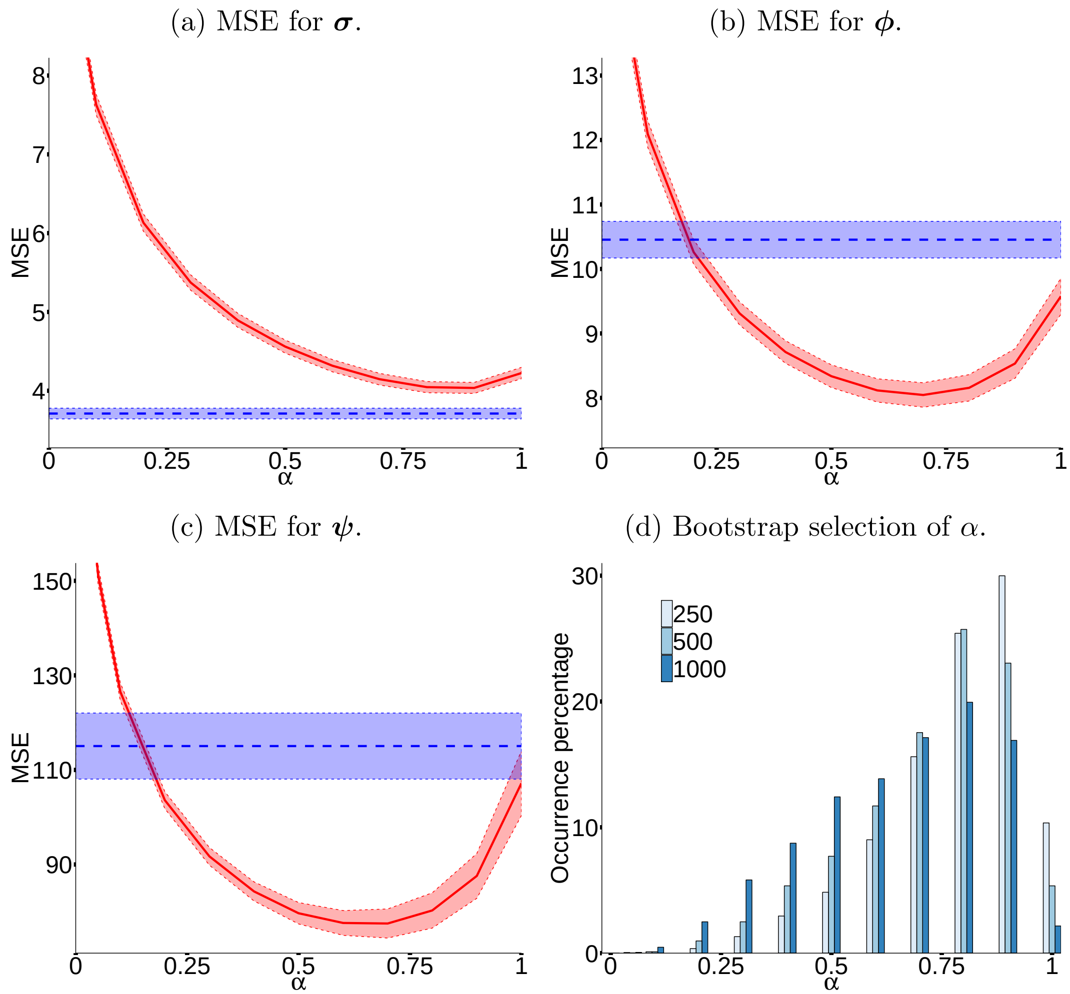
Figure 1a-c illustrates the dependence of the MSE of the NC estimator on the choice of the regularization parameter α when $n = 1000$, $q = 1$ and $p = 5$. The figures reveal a smooth convex relation between the regularization parameter α and the resulting MSE.

Table 2: MSE for the higher-order comoments.

		$p=3$			$p=5$			$p=10$		
		250	500	1000	250	500	1000	250	500	1000
Panel A: $q=1$										
$\boldsymbol{W}_D^{1/2}\boldsymbol{\zeta}$	Sample	67.17 (4.31)	34.09 (2.41)	17.74 (0.81)	91.02 (13.11)	38.94 (2.35)	22.24 (1.58)	82.18 (7.61)	35.19 (1.83)	20.50 (1.32)
	PC-FM	108.83 (6.38)	75.27 (3.27)	62.01 (1.78)	94.59 (9.84)	51.53 (2.83)	38.25 (2.12)	75.18 (6.12)	38.40 (1.46)	28.15 (1.04)
	NC_b	**41.03** (1.59)	**22.57** (0.76)	**11.97** (0.41)	**47.38** (2.40)	**26.57** (1.06)	**15.12** (0.73)	**49.71** (3.32)	**26.51** (1.02)	**14.74** (0.49)
$\boldsymbol{\sigma}$	Sample	**23.36** (0.77)	**11.04** (0.31)	**5.75** (0.16)	**15.00** (0.50)	**7.25** (0.22)	**3.71** (0.11)	**21.02** (0.72)	**10.11** (0.27)	**5.16** (0.14)
	PC-FM	46.27 (1.18)	32.60 (0.60)	27.64 (0.39)	24.17 (0.68)	15.46 (0.33)	11.99 (0.21)	39.39 (0.97)	28.77 (0.46)	23.20 (0.29)
	NC_b	24.38 (0.69)	12.12 (0.33)	6.17 (0.17)	15.76 (0.44)	8.10 (0.23)	4.16 (0.12)	26.48 (0.66)	13.70 (0.34)	7.20 (0.18)
$\boldsymbol{\phi}$	Sample	74.35 (4.83)	37.90 (2.30)	18.93 (0.78)	42.04 (3.24)	20.33 (1.08)	10.61 (0.53)	119.34 (27.42)	41.16 (2.52)	25.80 (2.83)
	PC-FM	73.06 (5.16)	38.18 (2.44)	20.77 (0.86)	39.78 (3.23)	20.08 (1.11)	11.22 (0.54)	119.49 (28.51)	43.87 (2.58)	30.57 (2.93)
	NC_b	**49.56** (2.18)	**27.76** (1.07)	**14.26** (0.47)	**26.47** (1.26)	**15.34** (0.65)	**7.97** (0.31)	**52.60** (3.46)	**26.32** (1.39)	**14.13** (0.51)
$\boldsymbol{\psi}$	Sample	928.65 (112.96)	494.92 (89.99)	249.11 (19.97)	485.40 (89.76)	183.36 (14.07)	119.59 (11.48)	4844.20 (2879.63)	726.04 (102.90)	742.64 (236.44)
	PC-FM	1011.32 (118.17)	567.18 (91.84)	331.69 (21.76)	503.97 (91.72)	200.25 (15.03)	140.24 (12.10)	4999.99 (2970.98)	799.38 (104.23)	825.45 (235.22)
	NC_b	**503.58** (30.36)	**286.01** (12.22)	**169.30** (8.35)	**230.43** (16.38)	**129.38** (6.20)	**77.86** (4.56)	**1135.06** (205.27)	**457.95** (38.34)	**284.38** (19.24)
Panel B: $q=2$										
$\boldsymbol{W}_D^{1/2}\boldsymbol{\zeta}$	Sample	92.00 (5.86)	41.64 (1.64)	23.47 (0.88)	108.58 (12.46)	46.34 (2.44)	25.78 (1.37)	121.53 (12.01)	64.21 (6.24)	39.74 (4.98)
	PC-FM	89.35 (5.82)	41.95 (1.66)	25.29 (0.93)	93.31 (10.61)	42.95 (2.30)	24.64 (1.24)	110.86 (9.49)	66.36 (4.95)	50.15 (4.93)
	NC_b	**63.09** (1.91)	**34.42** (0.99)	**17.41** (0.43)	**62.05** (3.03)	**32.97** (1.46)	**17.53** (0.60)	**81.73** (8.42)	**37.15** (1.45)	**21.07** (0.71)
$\boldsymbol{\sigma}$	Sample	**26.64** (0.87)	**12.72** (0.42)	**6.35** (0.18)	**15.84** (0.52)	**7.69** (0.25)	**3.88** (0.12)	**23.34** (0.75)	**10.77** (0.34)	**5.83** (0.18)
	PC-FM	32.08 (1.01)	17.85 (0.51)	11.52 (0.25)	19.17 (0.62)	10.82 (0.31)	6.94 (0.16)	34.34 (0.84)	21.84 (0.39)	16.97 (0.22)
	NC_b	27.88 (0.77)	13.89 (0.42)	6.72 (0.19)	17.15 (0.46)	8.73 (0.25)	4.38 (0.13)	28.42 (0.70)	13.96 (0.37)	7.37 (0.20)
$\boldsymbol{\phi}$	Sample	98.23 (5.66)	43.44 (1.93)	24.22 (0.89)	46.94 (3.38)	22.19 (1.07)	11.45 (0.47)	146.45 (16.49)	65.25 (5.16)	40.60 (5.78)
	PC-FM	83.62 (5.21)	37.78 (1.62)	21.26 (0.75)	41.70 (3.35)	19.79 (1.00)	10.61 (0.45)	124.12 (14.61)	54.73 (3.94)	38.38 (4.34)
	NC_b	**75.97** (2.55)	**38.13** (1.25)	**19.48** (0.59)	**31.92** (1.23)	**16.74** (0.70)	**8.85** (0.30)	**93.81** (10.12)	**40.29** (3.33)	**23.72** (3.99)
$\boldsymbol{\psi}$	Sample	1235.34 (148.13)	538.85 (36.89)	327.36 (19.14)	502.80 (92.14)	220.17 (16.46)	120.44 (8.95)	4048.28 (1118.81)	1667.36 (345.39)	1742.55 (776.73)
	PC-FM	1089.31 (149.79)	471.17 (28.95)	292.18 (16.76)	477.14 (96.21)	200.90 (15.4)8	118.81 (9.42)	3419.75 (932.53)	1311.43 (220.81)	1326.36 (447.28)
	NC_b	**727.83** (35.21)	**401.46** (18.89)	**216.17** (6.83)	**277.73** (19.85)	**149.84** (8.71)	**83.18** (3.78)	**2152.05** (466.64)	**1024.77** (269.90)	**1317.96** (887.63)

Note: this table shows the average MSE per comoment element of the sample, PC-FM and NC estimators. The NC estimator is provided for α set by the bootstrap (NC_b). The study was conducted using 1000 replications in each setting of dimension $p = 3, 5$ and 10, number of factors $q = 1, 2$ and sample size $n = 250, 500$ and 1000. The lowest MSE per scenario is highlighted in bold. Standard errors are shown in parentheses and all values are to be multiplied by 10^{-4}.

Figure 1: MSE of the NC estimator as a function of α.



Note: Figures (a), (b) and (c) show the MSE of the NC estimator (full line) as a function of the regularization parameters α in combination with the MSE of the sample moments (dashed line). The values and pointwise 95% confidence bands are based on 10 000 replications. The dimension is $p = 5$ with a single latent factor $q = 1$ and sample size $n = 1000$. All values are to be multiplied by 10^{-4}. Figure (d) shows the percentage of times each α is selected by the bootstrap procedure in the same setting for sample sizes $n = 250, 500$ and 1000.

For the MSE of the covariance estimates, any value except for very low α results in an acceptable increase in MSE. Hence, if the aim is solely to estimate the covariance matrix, the NC estimator is not always recommended. The percentage of times each value of α is selected in this scenario is shown in Figure 1d. The probability mass of the sample distribution moves slowly to the left when the sample size grows and the estimate of $\widehat{\boldsymbol{\Xi}}$ increases in accuracy.

The blow up of the MSE at $\alpha = 0$ in Figure 1a-c is due to an increase in bias in pursuit of the lowest possible estimation variance. The bootstrap procedure counters this effect by selecting α such as to minimize a weighted MSE instead of the estimation variance, hence optimally balancing bias and variance.

6.2 Setting calibrated on mental ability scores

Set up. In structural equation modelling, the assumption of a latent factor model explaining the observations is common. Instead of assuming independence of the factors, the aim is to obtain a sparse and interpretable matrix with factor loadings $\boldsymbol{B}$. This can be achieved under (3) by transforming the factors linearly to yield correlated factors for which the loadings are sparse. Similarly to Yuan & Chan (2016) and Arruda & Bentler (2017), we consider a multi-factor model of form (3) and define the data-generating process by

$$\boldsymbol{X} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\varepsilon} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2}\boldsymbol{z} + \boldsymbol{\varepsilon}, \tag{29}$$

where

$$\boldsymbol{\Lambda}' = \begin{pmatrix} 0.7 & 0.8 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.7 & 0.8 & 0.9 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0.8 & 0.9 \end{pmatrix} \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \begin{pmatrix} 1 & 0.3 & 0.4 \\ 0.3 & 1 & 0.5 \\ 0.4 & 0.5 & 1 \end{pmatrix}. \tag{30}$$

The factors $\boldsymbol{z}$ are independent and distributed as standardized chi-squared random variables with degrees of freedom of 1, 1.5 and 2. The matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2}$ is symmetric and has the property $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2} = \boldsymbol{\Sigma}_{\boldsymbol{\xi}}$. Finally, the idiosyncratic term $\boldsymbol{\varepsilon}$ has mean zero and consists of independent Gaussian variables with variances such that the variances of $\boldsymbol{X}$ are equal to one. We simulate 1000 times from this distribution with sample sizes $n = 300$, 500 and 1000.

Performance results. Our interest also lies in evaluating the estimation accuracy of the factor loadings $\boldsymbol{B} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2}$. This parameter matrix is identifiable using the NC estimator with either the covariance and coskewness elements or with all unique moments up to the fourth order; see Theorem 2.1. Hence, in this set up it makes sense to compare the accuracy of the NC estimators when excluding or including the cokurtosis elements. The sample moments and PC-FM approach do not provide an identifiable matrix with factor

Table 3: MSE for the matrix with factor loadings and higher-order comoments.

		$(\boldsymbol{\sigma} \quad \boldsymbol{\phi})$			$(\boldsymbol{\sigma} \quad \boldsymbol{\phi} \quad \boldsymbol{\psi})$		
		300	500	1000	300	500	1000
$\boldsymbol{B}$	NC_b	**97.24** (2.85)	**74.94** (1.09)	**60.60** (0.65)	**128.31** (9.27)	**84.03** (2.22)	**62.06** (0.71)
$\boldsymbol{W}_D^{1/2}\boldsymbol{\zeta}$	Sample	67.65 (2.37)	50.73 (1.66)	34.60 (0.76)	114.63 (8.61)	85.51 (6.81)	52.46 (2.26)
	PC-FM	70.81 (2.44)	55.32 (1.72)	40.36 (0.80)	120.25 (8.98)	92.78 (7.15)	60.68 (2.43)
	NC_b	**39.87** (1.36)	**32.17** (1.25)	**22.59** (0.49)	**55.57** (4.54)	**42.20** (3.95)	**22.70** (1.09)
$\boldsymbol{\sigma}$	Sample	333.21 (6.86)	298.16 (5.26)	268.61 (3.45)	333.21 (6.86)	298.16 (5.26)	268.61 (3.45)
	PC-FM	444.82 (7.94)	410.79 (6.14)	383.01 (4.08)	444.82 (7.94)	410.79 (6.14)	383.01 (4.08)
	NC_b	**254.74** (5.71)	**234.42** (4.47)	**221.07** (2.97)	**239.14** (5.25)	**209.80** (4.05)	**197.22** (2.79)
$\boldsymbol{\phi}$	Sample	1317.61 (74.11)	935.91 (37.02)	599.98 (16.04)	1317.61 (74.11)	935.91 (37.02)	599.98 (16.04)
	PC-FM	1276.70 (71.12)	918.91 (35.42)	604.35 (15.26)	1276.70 (71.12)	918.91 (35.42)	604.35 (15.26)
	NC_b	**776.57** (32.59)	**604.68** (26.23)	**382.15** (10.27)	**870.03** (42.05)	**612.97** (26.41)	**365.31** (9.73)
$\boldsymbol{\psi}$	Sample				39028.20 (6131.91)	25987.45 (2083.52)	15891.41 (690.39)
	PC-FM				39134.87 (5794.49)	26789.15 (1987.23)	17230.54 (665.36)
	NC_b				**20756.10** (2241.73)	**14253.04** (1350.64)	**7507.80** (270.20)

Note: this table shows the average MSE per comoment element of the sample, PC-FM and NC estimators. The NC estimator is provided for α set by the bootstrap (NC_b). The study was conducted using 1000 replications for sample sizes $n = 300, 500$ and 1000. The lowest MSE per scenario is highlighted in bold. In addition, the average MSE per factor loading is given. Standard errors are provided in parentheses and all values are to be multiplied by 10^{-4}.

loadings, and thus no MSE values are reported. Results are provided in Table 3 and show that including the cokurtosis elements increases the MSE when estimating $\boldsymbol{B}$. When sample size increases, the difference in MSE between the two methods becomes less pronounced and is negligible for sample size 1000. Another observation is that the coskewness elements are better estimated when the cokurtosis elements are excluded when $n = 300$ and $n = 500$. However, for $n = 1000$, including the cokurtosis elements yields better results. As in the previous simulations, the NC estimator has gains up to 55% in accuracy compared to the sample estimator when estimating the higher-order comoments.

7 Empirical applications

The NC estimates of the higher-order comoments of non-normal random variables can be used to improve such operations as dynamic portfolio allocation and factor extraction. In this section, we find that the proposed NC estimator offers a significant economic advantage over the sample comoments in multiple settings of portfolio allocation under non-Gaussian objective functions. In addition, we extract the latent factors in the Holzinger & Swineford (1939) dataset, which is often used as an example in the literature on structural equation modelling; see Yuan & Chan (2016) for a recent example.

7.1 Optimization of a portfolio of hedge fund indices

In this section, we analyse the usefulness of the proposed NC estimator in dynamic portfolio allocation. The data consist of weekly returns of the five main HFRX indices for the period January 2, 2004 to December 29, 2017. These are the equity hedge, event-driven, macro/CTA, relative value arbitrage and global hedge fund indices and are investible through tracker funds constructed by HFR Asset Management, LLC.

To account for potential time variation of the comoments, we follow the industry practice of using rolling five-year samples. Hence, the most recent 260 weekly returns are used to determine the comoments each week. In the NC estimator, the number of factors and the ridge parameter are re-evaluated annually. We consider three settings. In the first, the aim is to optimize the portfolio in order to achieve the lowest adjusted Value-at-Risk (VaR) at a 95% level; in the second, however, we aim to maximize the expected utility of an investor with constant relative risk aversion (CRRA) $\gamma = 15$. In addition, we consider mean-variance-skewness-kurtosis (MVSK) optimal portfolios as proposed in Briec et al. (2007) and implemented in Cornilly & Boudt (2019). Their definitions are recalled in the supplementary appendix. All portfolios assume total investment and no short selling, and we set the estimated mean equal to zero when maximizing the expected utility, as is common in the literature.

The out-of-sample returns are compared to those of the equally weighted portfolio and of the three optimized portfolios based on the sample comoment estimators. All portfolios are evaluated using several out-of-sample performance measures. For skewness and kurtosis

Table 4: Out-of-sample performance of the portfolios.

		min. VaR		max. EU		MVSK	
	EW	S	NC_b	S	NC_b	S	NC_b
Ann. geometric mean (%)	2.46	1.66	2.93	1.90	3.00	2.55	**3.03**
Ann. standard deviation (10^{-2})	3.55	**2.96**	2.98	2.98	3.06	3.10	3.01
Skewness (10^{-8})	-10.78	-6.27	-3.48	-6.36	**−3.44**	-6.01	-4.94
Standardized skewness	-0.91	-0.91	-0.49	-0.90	**−0.45**	-0.76	-0.68
Kurtosis (10^{-9})	3.15	1.67	1.66	1.71	1.86	1.87	**1.58**
Excess kurtosis	2.41	2.88	2.70	2.87	2.79	2.48	**2.21**
95% VaR (%)	0.85	0.68	**0.61**	0.68	0.64	0.68	0.65
MUG over EW (bp.)	0	-51.35	72.89	-27.96	76.70	30.19	**81.44**
Break-even transaction costs ($)			7.76		**8.45**	1.51	2.86

Note: this table shows out-of-sample performance measures for the various portfolios: annualized geometric mean and standard deviation, skewness, standardized skewness, kurtosis, excess kurtosis and 95% Value-at-Risk (VaR). In addition, we provide the annualized monetary utility gain (MUG) and break-even transaction costs (dollar per $1000 traded). The equally weighted portfolio is denoted EW. The minimum 95%-VaR (min. VaR) portfolios are based on the sample moments (S) or the NC estimator (NC_b). The same two estimators are used to construct the maximum expected utility (max. EU) portfolios with risk aversion parameter $\gamma = 15$ and the mean-variance-skewness-kurtosis efficient (MVSK) portfolios. The best portfolio for each statistic is highlighted in bold.

we report the central moments as used in this paper, as well as the more traditional standardized definitions. In addition, we report the 95% VaR and Monetary Utility Gain (MUG) with respect to the equally weighted portfolio. As in Ang & Bekaert (2002) and Martellini & Ziemann (2010), the MUG equals the additional annual percentage return required by investors in the benchmark portfolio, rendering these investors indifferent to a change in investment strategy. In order to measure the relevance of such economic gains, we also report the break-even transaction costs for which a CRRA investor would be indifferent with respect to choosing between the optimized portfolio and the equally weighted portfolio. This measure is denoted in dollars per $1000 traded.

Table 4 reports these summary statistics for the various optimized portfolios and the equally weighted portfolio. Overall, the optimized portfolios have a lower standard deviation, larger skewness and a lower kurtosis compared to the equally weighted portfolio. Especially remarkable is the fact that these improvements are relative not only to the equally weighted portfolio, but also to the three different optimized portfolios based on the sample comoments, which clearly demonstrates the performance advantage conferred by the NC estimator. In terms of return, the NC estimator delivers better performance, while the sample optimized portfolios are mostly worse.

We further find that the MUG values of the optimized portfolios are all positive if the moments are estimated using the NC estimator, which represents an advantage over the equally weighted portfolio. The values are highly economically relevant, ranging from 72 basis points to 81 basis points annually. Surprisingly, a CRRA investor prefers the equally weighted portfolio over both the sample optimized maximum utility and minimum VaR portfolios, indicating an amplification of measurement error when utilizing the sample moments. We verified that different values for γ do not affect the general conclusions presented in this section. To conclude, the break-even transaction costs with respect to the equally weighted portfolio range from \$2.9 to \$8.5 per \$1000 traded in the portfolios with positive MUG, indicating that even with transaction costs included, there is an incentive for a CRRA investor to invest in the NC optimized portfolios instead of the equally weighted one.

7.2 Factor loadings of mental ability scores

The NC estimator is also useful for extracting factor loadings if the data are assumed to have a latent lower dimensional structure. We illustrate this using a study of the classic Holzinger & Swineford (1939) dataset consisting of mental ability scores of seventh- and eighth-grade pupils. We follow Jöreskog (1969) and many subsequent studies in considering a subset of nine variables. The resulting dataset of 301 observations in nine dimensions is made available in `R` through the `lavaan` package of Rosseel (2012). The nine variables are provided in Table 5. The confirmatory factor analysis model that is often proposed consists of three latent variables: a visual factor (variables 1, 2, 3, 9), a textual factor (variables 4, 5, 6) and a speed factor (variables 7, 8, 9).

The traditional approach is to estimate the structural model using maximum likelihood or distribution-free methods, forcing the other factor loadings to be zero. By contrast, the proposed NC estimator leads to an identified matrix with factor loadings, without assuming zero factor loadings. This approach makes it possible to explore interactions that have previously been neglected. As in the simulation study, we consider the NC estimators

Table 5: Variables from the Holzinger and Swineford dataset.

Variable	Meaning	Factor
1	visual perception	visual
2	cubes	visual
3	lozenges	visual
4	paragraph comprehension	textual
5	sentence completion	textual
6	word meaning	textual
7	speeded addition	speed
8	speeded counting of dots	speed
9	speeded discrimination straight and curved capitals	visual and speed

Note: this table shows the nine variables in the classic Jöreskog (1969) dataset of mental ability scores of seventh- and eighth-grade pupils. This is an often studied subset of the dataset in Holzinger & Swineford (1939). We also provide the traditional three latent factors (visual, textual, speed) and report which ones influence each of the variables.

containing all joint moments up to the fourth order, as presented in Section 3, as well as the NC estimator excluding the cokurtosis elements.

A scree plot, not included here, confirms the use of three factors for both estimators. Next, we determine the optimal ridge parameter α using an equally spaced grid from 0.1 to 1 with increments of 0.1 and 250 bootstrap samples per value of α. When cokurtosis elements are excluded, the optimal value is estimated as $\widehat{\alpha} = 0.9$, while the diagonal matrix is optimal when the cokurtosis elements are included. In this application, the estimated factor loadings are of interest. Since the factors are latent, the obtained matrices cannot be interpreted directly. However, a promax rotation (Hendrickson & White (1964)) yields an interpretable structure of factor loadings.

The obtained factor loadings for both estimators are reported in Table 6, together with p-values for a one-sided test with alternative hypothesis of the loading being positive. The loadings significant at a 5% level are highlighted in bold. We clearly recover the traditional structure of the matrix with factor loadings under both estimators. However, in both cases there is an additional significant factor loading: variable six (word meaning) is positively influenced by both the visual and textual latent factors. One explanation could be that the meaning of a word is related to visual images in memory. Further discussion of why the visual factor is present in this context lies outside the scope of the present paper, and we leave this matter to experts in the relevant fields.

Table 6: Factor loadings implied by the nearest comoment estimator.

		$(\boldsymbol{\sigma} \quad \boldsymbol{\phi})$			$(\boldsymbol{\sigma} \quad \boldsymbol{\phi} \quad \boldsymbol{\psi})$		
	Variable	visual	textual	speed	visual	textual	speed
1	visual perception	**0.53** (<0.01)	0.18 (0.07)	0.10 (0.25)	**0.58** (<0.01)	0.20 (0.05)	0.10 (0.26)
2	cubes	**0.71** (<0.01)	-0.06 (0.73)	-0.17 (0.90)	**0.76** (<0.01)	-0.06 (0.73)	-0.23 (0.95)
3	lozenges	**0.72** (<0.01)	-0.13 (0.89)	0.07 (0.30)	**0.71** (<0.01)	-0.09 (0.80)	0.07 (0.33)
4	paragraph comprehension	0.05 (0.32)	**0.91** (<0.01)	-0.01 (0.52)	-0.01 (0.53)	**0.99** (<0.01)	0.00 (0.51)
5	sentence completion	-0.21 (0.97)	**1.11** (<0.01)	0.07 (0.37)	-0.12 (0.85)	**1.04** (<0.01)	0.05 (0.40)
6	word meaning	**0.22** (0.02)	**0.93** (<0.01)	-0.10 (0.68)	**0.19** (0.03)	**0.92** (<0.01)	-0.08 (0.65)
7	speeded addition	-0.17 (0.88)	0.04 (0.38)	**0.71** (<0.01)	-0.22 (0.94)	0.03 (0.42)	**0.83** (<0.01)
8	speeded counting of dots	0.03 (0.42)	-0.03 (0.57)	**0.76** (<0.01)	0.09 (0.28)	-0.06 (0.64)	**0.74** (<0.01)
9	speeded discrimination	**0.33** (<0.01)	-0.01 (0.53)	**0.49** (<0.01)	**0.35** (<0.01)	0.00 (0.49)	**0.50** (<0.01)

Note: this table shows the factor loadings obtained after a promax rotation for two NC estimators. In column $(\boldsymbol{\sigma} \quad \boldsymbol{\phi})$, only the second and third-order moments are used in the NC estimation procedure, while in column $(\boldsymbol{\sigma} \quad \boldsymbol{\phi} \quad \boldsymbol{\psi})$ all moments up to the fourth order are included. The latent factors are named after their traditional interpretation, and we provide p-values for a one-sided test with alternative hypothesis of the loadings being positive.

8 Conclusion

We propose the NC estimator for joint estimation of the covariance, coskewness and cokurtosis matrices under the assumption of a latent factor model. The estimator exploits the resulting structure in the higher-order comoments to improve finite sample and asymptotic estimation accuracy. An advantage of our approach is that no ex ante selection of factors is required, yet the benefits of the factor structure still pertain. Asymptotic normality was proven, and its relation to the sample comoments was shown by means of the influence function. An extensive simulation study with data-generating processes based on the empirical applications showed improvements in MSE of up to 55% over all covariance, coskewness and cokurtosis elements jointly. We have also illustrated the usefulness of the novel framework in dynamic portfolio allocation and factor extraction.

Further prospective work in this regard consists of including dynamic behaviour in the factor model to accommodate time-variation in the conditional comoments, as in Bauwens & Laurent (2005), Bauwens et al. (2006) and Barigozzi & Hallin (2016, 2017). In addition, as the higher-order comoments are very sensitive to influential observations, it would be useful to make the NC estimator more robust to observations that do not follow the same model as the majority of the data.

Acknowledgements

We thank seminar participants at CREST, ETH Zürich and Vrije Universiteit Amsterdam for their valuable comments. We also benefited from fruitful discussions with participants at the CMStatistics, JSM and R/Finance conferences. The authors gratefully acknowledge support from the Research Foundation - Flanders (FWO research grant $G023815N$ and PhD fellowship $1114117N$) and the Internal Funds KU Leuven (project $C16/15/068$). The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the FWO and the Flemish Government - department EWI.

References

Ang, A. & Bekaert, G. (2002), 'International asset allocation with regime shifts', *Review of Financial Studies* **15**(4), 1137–1187.

Arruda, E. H. & Bentler, P. M. (2017), 'A regularized GLS for structural equation modeling', *Structural Equation Modeling: A Multidisciplinary Journal* **24**(5), 657–665.

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**(1), 135–171.

Bai, J. & Ng, S. (2013), 'Principal components estimation and identification of static factors', *Journal of Econometrics* **176**(1), 18–29.

Barigozzi, M. & Hallin, M. (2016), 'Generalized dynamic factor models and volatilities: recovering the market volatility shocks', *Econometrics Journal* **19**(1), C33–C60.

Barigozzi, M. & Hallin, M. (2017), 'Generalized dynamic factor models and volatilities: estimation and forecasting', *Journal of Econometrics* **201**(2), 307–321.

Bauwens, L. & Laurent, S. (2005), 'A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models', *Journal of Business & Economic Statistics* **23**(3), 346–354.

Bauwens, L., Laurent, S. & Rombouts, J. V. (2006), 'Multivariate garch models: a survey', *Journal of Applied Econometrics* **21**(1), 79–109.

Bonhomme, S. & Robin, J.-M. (2009), 'Consistent noisy independent component analysis', *Journal of Econometrics* **149**(1), 12–25.

Boudt, K., Carl, P. & Peterson, B. G. (2013), 'Asset allocation with Conditional Value-at-Risk budgets', *Journal of Risk* **15**(3), 39–68.

Boudt, K., Cornilly, D. & Verdonck, T. (2018), 'A coskewness shrinkage approach to estimating the skewness of a linear combination of random variables', *Journal of Financial Econometrics* (forthcoming).

Boudt, K., Lu, W. & Peeters, B. (2015), 'Higher order comoments of multifactor models and asset allocation', *Finance Research Letters* **13**, 225–233.

Briec, W., Kerstens, K. & Jokung, O. (2007), 'Mean-variance-skewness portfolio performance gauging: a general shortage function and dual approach', *Management Science* **53**(1), 135–149.

Chen, L., Dolado, J. J. & Gonzalo, J. (2018), 'Quantile factor models', (working paper).

Cornilly, D. & Boudt, K. (2019), *mvskPortfolios: Higher Moment Efficient Portfolios*. R package version 0.1.0.
URL: *https://github.com/cdries/mvskPortfolios*

Fan, J., Xue, L. & Yao, J. (2017), 'Sufficient forecasting using factor models', *Journal of Econometrics* **201**(2), 292–306.

Ghalanos, A., Rossi, E. & Urga, G. (2015), 'Independent factor autoregressive conditional density model', *Econometric Reviews* **34**(5), 594–616.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (2011), *Robust Statistics: the Approach based on Influence Functions*, John Wiley and Sons, New York.

Harvey, C. R., Liechty, J. C., Liechty, M. W. & Müller, P. (2010), 'Portfolio selection with higher moments', *Quantitative Finance* **10**(5), 469–485.

Hendrickson, A. E. & White, P. O. (1964), 'Promax: A quick method for rotation to oblique simple structure', *British Journal of Mathematical and Statistical Psychology* **17**(1), 65–70.

Holzinger, K. J. & Swineford, F. (1939), 'A study in factor analysis: the stability of a bi-factor solution.', *Supplementary Educational Monographs* **48**, xi+91.

Jondeau, E., Jurczenko, E. & Rockinger, M. (2018), 'Moment component analysis: An illustration with international stock markets', *Journal of Business & Economic Statistics* **36**(4), 576–598.

Jondeau, E. & Rockinger, M. (2006), 'Optimal portfolio allocation under higher moments', *European Financial Management* **12**(1), 29–55.

Jöreskog, K. G. (1969), 'A general approach to confirmatory maximum likelihood factor analysis', *Psychometrika* **34**(2), 183–202.

Karlis, D. (2002), 'An EM type algorithm for maximum likelihood estimation of the normal–inverse Gaussian distribution', *Statistics & Probability Letters* **57**(1), 43–52.

Lawley, D. & Maxwell, A. (1963), *Factor Analysis as a Statistical Method*, Butterworth, London.

Lee, S. S. & Mykland, P. A. (2007), 'Jumps in financial markets: A new nonparametric test and jump dynamics', *Review of Financial Studies* **21**(6), 2535–2563.

Luciani, M. & Veredas, D. (2015), 'Estimating and forecasting large panels of volatilities with approximate dynamic factor models', *Journal of Forecasting* **34**(3), 163–176.

Martellini, L. & Ziemann, V. (2010), 'Improved estimates of higher-order comoments and implications for portfolio selection', *Review of Financial Studies* **23**(4), 1467–1502.

Mooijaart, A. (1985), 'Factor analysis for non-normal variables', *Psychometrika* **50**(3), 323–342.

Morton, J. & Lim, L.-H. (2009), 'Principal cumulant component analysis', (working paper).

Newey, W. K. & McFadden, D. (1994), 'Large sample estimation and hypothesis testing', *Handbook of Econometrics* **4**, 2111–2245.

Peterson, B. G. & Carl, P. (2018), *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis*. R package version 1.5.2.
URL: *https://github.com/braverock/PerformanceAnalytics*

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *Journal of Economic Theory* **13**(3), 341–360.

Rosseel, Y. (2012), 'lavaan: An R package for structural equation modeling', *Journal of Statistical Software* **48**(2), 1–36.

Serfling, R. J. (2009), *Approximation Theorems of Mathematical Statistics*, Vol. 162, John Wiley & Sons, New York.

Stoyanov, S. V., Rachev, S. T. & Fabozzi, F. J. (2013), 'Sensitivity of portfolio VaR and CVaR to portfolio return characteristics', *Annals of Operations Research* **205**(1), 169–187.

Stuart, A. & Ord, J. K. (1994), *Kendall's Advanced Theory of Statistics*, Edward Arnold, London.

Yuan, K.-H. & Bentler, P. M. (2010), 'Two simple approximations to the distributions of quadratic forms', *British Journal of Mathematical and Statistical Psychology* **63**(2), 273–291.

Yuan, K.-H. & Chan, W. (2016), 'Structural equation modeling with unknown population distributions: Ridge generalized least squares', *Structural Equation Modeling: A Multidisciplinary Journal* **23**(2), 163–179.

Yuan, K.-H., Jiang, G. & Cheng, Y. (2017), 'More efficient parameter estimates for factor analysis of ordinal variables by ridge generalized least squares', *British Journal of Mathematical and Statistical Psychology* **70**(3), 525–564.

Zangari, P. (1996), 'A VaR methodology for portfolios that include options', *RiskMetrics Monitor* **1**, 4–12.

A Proofs

Proof of Theorem 4.1. We generalize the proof given in Serfling (2009) and derive the influence function of multivariate central moments of any order.

Consider the functional $\mu_{i_1 i_2 \cdots i_r}(F)$ with $i_j \geq 1, j = 1, \ldots, r$, depending on the r-dimensional distribution F, defined by

$$\mu_{i_1 i_2 \cdots i_r}(F) = \mathbb{E}\left[(X_1 - \mu_1)^{i_1}(X_2 - \mu_2)^{i_2} \cdots (X_r - \mu_r)^{i_r}\right] = \int_{\mathbb{R}^r} \prod_{j=1}^{r} (x_j - \mu_j)^{i_j} dF(\boldsymbol{x}). \quad (31)$$

For a certain $\boldsymbol{x}_0 \in \mathbb{R}^r$, define the distribution $F_\lambda = F + \lambda(\delta_{\boldsymbol{x}_0} - F)$, which has mean $\boldsymbol{\mu}_\lambda = \boldsymbol{\mu} + \lambda(\boldsymbol{x}_0 - \boldsymbol{\mu})$. Then

$$\mu_{i_1 i_2 \cdots i_r}(F_\lambda) = \int_{\mathbb{R}^r} \prod_{j=1}^{r} (x_j - \mu_{\lambda,j})^{i_j} dF(\boldsymbol{x}) + \lambda \int_{\mathbb{R}^r} \prod_{j=1}^{r} (x_j - \mu_{\lambda,j})^{i_j} d(\delta_{\boldsymbol{x}_0} - F)(\boldsymbol{x}) \quad (32)$$

and

$$\begin{aligned} \frac{d\mu_{i_1 i_2 \cdots i_r}(F_\lambda)}{d\lambda} &= -\sum_{k=1}^{r} \left[(x_{0,k} - \mu_k) \int_{\mathbb{R}^r} \frac{1}{(x_k - \mu_{\lambda,k})} \prod_{j=1}^{r} (x_j - \mu_{\lambda,j})^{i_j} dF(\boldsymbol{x})\right] \\ &\quad + \int_{\mathbb{R}^r} \prod_{j=1}^{r} (x_j - \mu_{\lambda,j})^{i_j} d(\delta_{\boldsymbol{x}_0} - F)(\boldsymbol{x}) + \lambda \cdot (\star). \end{aligned} \quad (33)$$

Hence, the influence function of the multivariate central moments equals

$$\begin{aligned} &\mathrm{IF}(\boldsymbol{x}; \mu_{i_1 i_2 \cdots i_r}, F) = \left.\frac{d\mu_{i_1 i_2 \cdots i_r}(F_\lambda)}{d\lambda}\right|_{\lambda=0} \quad (34) \\ &= -\sum_{k=1}^{r} \left[(x_k - \mu_k) \int_{\mathbb{R}^r} \frac{1}{(x_k - \mu_k)} \prod_{j=1}^{r} (x_j - \mu_j)^{i_j} dF(\boldsymbol{x})\right] + \prod_{j=1}^{r} (x_j - \mu_j)^{i_j} - \mu_{i_1 i_2 \cdots i_r}(F). \end{aligned}$$

The functions given in Theorem 4.1 are then obtained by substitution.

Proof of Theorem 4.2. Define the functional $\widehat{\boldsymbol{\zeta}}(F)$, stacking the unique second, third and fourth-order central moment functionals in (31). For some $\boldsymbol{x}_0 \in \mathbb{R}^p$, consider the contaminated distribution $F_\lambda = F + \lambda(\delta_{\boldsymbol{x}_0} - F)$. Define the functional $\boldsymbol{\theta}_\lambda$

$$\boldsymbol{\theta}_\lambda = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\widehat{\boldsymbol{\zeta}}(F_\lambda) - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\right)' \boldsymbol{W}(F_\lambda) \left(\widehat{\boldsymbol{\zeta}}(F_\lambda) - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\right), \tag{35}$$

for some $\lambda \in [0, \varepsilon]$ and functional $\boldsymbol{W}(\cdot)$ satisfying $\widehat{\boldsymbol{W}}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \xrightarrow{p} \boldsymbol{W}(F)$, for $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ a sample of n independent and identically distributed random vectors with distribution F. To obtain an explicit representation of $\boldsymbol{\theta}_\lambda$, we define the function

$$\ell : \boldsymbol{\Theta} \times [0, \varepsilon] \to \mathbb{R}^\kappa : \ell(\boldsymbol{\theta}, \lambda) = \boldsymbol{G}(\boldsymbol{\theta})'\boldsymbol{W}(F_\lambda) \left(\widehat{\boldsymbol{\zeta}}(F_\lambda) - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\right), \tag{36}$$

for some $\varepsilon > 0$, with $\boldsymbol{G}(\boldsymbol{\theta})$ the Jacobian function of $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$. We then note that $\ell(\boldsymbol{\theta}_\lambda, \lambda) = 0$ for all $\lambda \in [0, \varepsilon]$ and $\boldsymbol{\theta}^\star$ satisfies $\ell(\boldsymbol{\theta}^\star, 0) = 0$.

We do a Taylor expansion around the point $(\boldsymbol{\theta}^\star, 0)$ for $\lambda > 0$:

$$\begin{aligned}\ell(\boldsymbol{\theta}_\lambda, \lambda) = \ell(\boldsymbol{\theta}^\star, 0) &+ \left.\frac{\partial \ell}{\partial \boldsymbol{\theta}}\right|_{(\boldsymbol{\theta}^\star,0)} (\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star) + \left.\frac{\partial \ell}{\partial \lambda}\right|_{(\boldsymbol{\theta}^\star,0)} \lambda + \left.\frac{\partial^2 \ell}{\partial \lambda^2}\right|_{(\tilde{\boldsymbol{\theta}},\tilde{\lambda})} \lambda^2 \\ &+ \left.\frac{\partial^2 \ell}{\partial \lambda \partial \boldsymbol{\theta}}\right|_{(\tilde{\boldsymbol{\theta}},\tilde{\lambda})} \lambda\,(\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star) + \left.\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}^2}\right|_{(\tilde{\boldsymbol{\theta}},\tilde{\lambda})} \left((\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star) \otimes (\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star)\right),\end{aligned} \tag{37}$$

where $\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}^2}$ is the $\kappa \times \kappa^2$ matrix containing all second-order derivatives of the vector-valued function ℓ and $(\tilde{\boldsymbol{\theta}}, \tilde{\lambda})$ is a value between $(\boldsymbol{\theta}^\star, 0)$ and $(\boldsymbol{\theta}_\lambda, \lambda)$. It holds that

$$\left.\frac{\partial \ell}{\partial \boldsymbol{\theta}}\right|_{(\boldsymbol{\theta}^\star,0)} = \frac{\partial}{\partial \boldsymbol{\theta}} \left[\boldsymbol{G}(\boldsymbol{\theta})'\boldsymbol{W}(F_\lambda) \left(\widehat{\boldsymbol{\zeta}}(F_\lambda) - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\right)\right] \Bigg|_{(\boldsymbol{\theta}^\star,0)} = -\boldsymbol{G}(\boldsymbol{\theta}^\star)'\boldsymbol{W}(F)\boldsymbol{G}(\boldsymbol{\theta}^\star). \tag{38}$$

The partial derivative of ℓ with respect to λ equals

$$\begin{aligned}\left.\frac{\partial \ell}{\partial \lambda}\right|_{(\boldsymbol{\theta}^\star,0)} &= \frac{\partial}{\partial \lambda} \left[\boldsymbol{G}(\boldsymbol{\theta})'\boldsymbol{W}(F_\lambda) \left(\widehat{\boldsymbol{\zeta}}(F_\lambda) - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\right)\right] \Bigg|_{(\boldsymbol{\theta}^\star,0)} \\ &= \boldsymbol{G}(\boldsymbol{\theta}^\star)' \left[\frac{\partial}{\partial \lambda} \boldsymbol{W}(F_\lambda)\big|_0 \left(\widehat{\boldsymbol{\zeta}}(F) - \boldsymbol{\zeta}_{\boldsymbol{\theta}^\star}\right) + \boldsymbol{W}(F) \frac{\partial}{\partial \lambda} \widehat{\boldsymbol{\zeta}}(F_\lambda)\big|_0\right] \\ &= \boldsymbol{G}(\boldsymbol{\theta}^\star)' \left[0 + \boldsymbol{W}(F)\,\mathrm{IF}(\boldsymbol{x}_0; \widehat{\boldsymbol{\zeta}}, F)\right] = \boldsymbol{G}(\boldsymbol{\theta}^\star)'\boldsymbol{W}(F)\,\mathrm{IF}(\boldsymbol{x}_0; \widehat{\boldsymbol{\zeta}}, F).\end{aligned} \tag{39}$$

Since $\ell(\boldsymbol{\theta}^\star, 0) = \ell(\boldsymbol{\theta}_\lambda, \lambda) = 0$, it holds that

$$\begin{aligned}\frac{\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star}{\lambda} = \left(\boldsymbol{G}(\boldsymbol{\theta}^\star)'\boldsymbol{W}(F)\boldsymbol{G}(\boldsymbol{\theta}^\star)\right)^{-1} &\Bigg[\boldsymbol{G}(\boldsymbol{\theta}^\star)'\boldsymbol{W}(F)\,\mathrm{IF}(\boldsymbol{x}_0; \widehat{\boldsymbol{\zeta}}, F) + \lambda \left.\frac{\partial^2 \ell}{\partial \lambda^2}\right|_{(\tilde{\boldsymbol{\theta}},\tilde{\lambda})} \\ &+ \left.\frac{\partial^2 \ell}{\partial \lambda \partial \boldsymbol{\theta}}\right|_{(\tilde{\boldsymbol{\theta}},\tilde{\lambda})} (\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star) + \frac{1}{\lambda} \left.\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}^2}\right|_{(\tilde{\boldsymbol{\theta}},\tilde{\lambda})} \left((\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star) \otimes (\boldsymbol{\theta}_\lambda - \boldsymbol{\theta}^\star)\right)\Bigg].\end{aligned} \tag{40}$$

Taking the limit $\lambda \to 0$ we obtain the influence function of the estimator $\widehat{\boldsymbol{\theta}}_{\text{nc}}$:

$$\text{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}_{\text{nc}}, F) = \left(\boldsymbol{G}(\boldsymbol{\theta}^\star)'\boldsymbol{W}(F)\boldsymbol{G}(\boldsymbol{\theta}^\star)\right)^{-1} \boldsymbol{G}(\boldsymbol{\theta}^\star)'\boldsymbol{W}(F)\,\text{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\zeta}}_s, F). \tag{41}$$

Since $\boldsymbol{W}(F) = \boldsymbol{W}$, we obtain the expression for the influence function in Theorem 4.2.

Because $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$ is a differentiable function with Jacobian $\boldsymbol{G}(\boldsymbol{\theta}^\star)$ at $\boldsymbol{\theta}^\star$, it follows from the chain rule that $\text{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\zeta}}_{\text{nc}}, F) = \boldsymbol{G}(\boldsymbol{\theta}^\star)\,\text{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}_{\text{nc}}, F)$.

Proof of Theorem 4.3. Establishing asymptotic normality of the sample comoments will be done with the multivariate Lindeberg-Lévy Central Limit Theorem. First, the sample moments will be rewritten in terms of their influence functions. The sample covariance estimator $\widehat{\sigma}_{\{ij\}}$ equals

$$\sqrt{n}\widehat{\sigma}_{\{ij\}} = \frac{1}{\sqrt{n}}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mj}-\mu_j) + \sqrt{n}(\overline{x}_i-\mu_i)(\overline{x}_j-\mu_j). \tag{42}$$

Since $\sqrt{n}(\overline{x}_i-\mu_i) \xrightarrow{d} \mathcal{N}(0, \sigma_{\{ii\}})$ and $(\overline{x}_j-\mu_j) \xrightarrow{a.s.} 0$, it follows by Slutsky's Lemma that

$$\sqrt{n}\widehat{\sigma}_{\{ij\}} = \frac{1}{\sqrt{n}}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mj}-\mu_j) + o_P(1), \tag{43}$$

where $o_P(1)$ is a term converging in probability to zero. Note that this is equivalent to

$$\sqrt{n}\left(\widehat{\sigma}_{\{ij\}} - \sigma_{\{ij\}} - \frac{1}{n}\sum_{i=1}^{n}\text{IF}(\boldsymbol{x}_i; \widehat{\sigma}_{\{ij\}}, F)\right) = o_p(1). \tag{44}$$

For the coskewness estimator we write

$$\begin{aligned}
\sqrt{n}\phi_{\{ijk\}} = {} & \frac{1}{\sqrt{n}}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mj}-\mu_j)(x_{mk}-\mu_k) - \sqrt{n}\left(\overline{x}_i-\mu_i\right)\sigma_{\{jk\}} \\
& - \sqrt{n}\left(\overline{x}_j-\mu_j\right)\sigma_{\{ik\}} - \sqrt{n}\left(\overline{x}_k-\mu_k\right)\sigma_{\{ij\}} \\
& - \sqrt{n}\left(\overline{x}_i-\mu_i\right)\left(\frac{1}{n}\sum_{m=1}^{n}(x_{mj}-\mu_j)(x_{mk}-\mu_k) - \sigma_{\{jk\}}\right) \\
& - \sqrt{n}\left(\overline{x}_j-\mu_j\right)\left(\frac{1}{n}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mk}-\mu_k) - \sigma_{\{ik\}}\right) \\
& - \sqrt{n}\left(\overline{x}_k-\mu_k\right)\left(\frac{1}{n}\sum_{m=1}^{n}(x_{mi}-\mu_i)(x_{mj}-\mu_j) - \sigma_{\{ij\}}\right) \\
& + 2\sqrt{n}\left(\overline{x}_i-\mu_i\right)\left(\overline{x}_j-\mu_j\right)\left(\overline{x}_k-\mu_k\right).
\end{aligned} \tag{45}$$

Again, by Slutsky's lemma, the last four terms are $o_P(1)$ and it holds that

$$\sqrt{n}\left(\widehat{\phi}_{\{ijk\}} - \phi_{\{ijk\}} - \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(\boldsymbol{x}_i; \widehat{\phi}_{\{ijk\}}, F)\right) = o_p(1). \tag{46}$$

By similar arguments, the following expression for the cokurtosis estimators is obtained

$$\sqrt{n}\left(\widehat{\psi}_{\{ijk\}} - \psi_{\{ijkl\}} - \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(\boldsymbol{x}_i; \widehat{\psi}_{\{ijkl\}}, F)\right) = o_p(1). \tag{47}$$

Define the function $\mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\zeta}}_s, F) = \Big(\mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\sigma}}_s, F)' \quad \mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\phi}}_s, F)' \quad \mathrm{IF}(\boldsymbol{x}; \widehat{\boldsymbol{\psi}}_s, F)'\Big)'$. The sample moments $\widehat{\boldsymbol{\zeta}}_s$ equal $\widehat{\boldsymbol{\zeta}}_s = \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}(\boldsymbol{x}_i; \widehat{\boldsymbol{\zeta}}_s, F) + o_P(1)$, where $\mathrm{IF}(\boldsymbol{x}_i; \widehat{\boldsymbol{\zeta}}_s, F), i = 1, \ldots, n$ are independent and identically distributed vectors. Hence, by the multivariate Lindeberg-Lévy Central Limit Theorem it holds that $\sqrt{n}(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi})$ as $n \to \infty$, where $\boldsymbol{\Xi} = \mathbb{E}\left[\mathrm{IF}(\boldsymbol{X}; \widehat{\boldsymbol{\zeta}}_s, F)' \,\mathrm{IF}(\boldsymbol{X}; \widehat{\boldsymbol{\zeta}}_s, F)\right]$.

Under the moment conditions, the matrix $\boldsymbol{\Xi}$ exists and has finite elements which are straightforward to compute. Using the summation notation as in Stuart & Ord (1994), they are defined by

$$\begin{aligned}
&\mathrm{ACov}(\sqrt{n}\widehat{\sigma}_{s,\{ij\}}, \sqrt{n}\widehat{\sigma}_{s,\{uv\}}) = \mu_{ijuv} - \mu_{ij}\mu_{uv},\\
&\mathrm{ACov}(\sqrt{n}\widehat{\sigma}_{s,\{ij\}}, \sqrt{n}\widehat{\phi}_{s,\{uvw\}}) = \mu_{ijuvw} - \mu_{ij}\mu_{uvw} - \mu_{iju}\mu_{vw} - \mu_{ijv}\mu_{uw} - \mu_{ijw}\mu_{uv},\\
&\mathrm{ACov}(\sqrt{n}\widehat{\sigma}_{s,\{ij\}}, \sqrt{n}\widehat{\psi}_{s,\{uvwz\}}) = \mu_{ijuvwz} - \mu_{ij}\mu_{uvwz} - \sum_{4}\mu_{vwz}\mu_{iju},\\
&\mathrm{ACov}(\sqrt{n}\widehat{\phi}_{s,\{ijk\}}, \sqrt{n}\widehat{\phi}_{s,\{uvw\}}) = \mu_{ijkuvw} - \mu_{ijk}\mu_{uvw} - \sum_{3}\mu_{ijku}\mu_{vw} - \sum_{3}\mu_{iuvw}\mu_{jk}\\
&\qquad + \mu_{ij}\sum_{3}\mu_{ku}\mu_{vw} + \mu_{ik}\sum_{3}\mu_{ju}\mu_{vw} + \mu_{jk}\sum_{3}\mu_{iu}\mu_{vw},\\
&\mathrm{ACov}(\sqrt{n}\widehat{\phi}_{s,\{ijk\}}, \sqrt{n}\widehat{\psi}_{s,\{uvwz\}}) = \mu_{ijkuvwz} - \mu_{ijk}\mu_{uvwz} - \sum_{4}\mu_{vwz}\mu_{ijku}\\
&\qquad - \sum_{3}\mu_{jk}\mu_{iuvwz} + \mu_{ij}\sum_{4}\mu_{vwz}\mu_{ku} + \mu_{ik}\sum_{4}\mu_{vwz}\mu_{ju} + \mu_{jk}\sum_{4}\mu_{vwz}\mu_{iu},\\
&\mathrm{ACov}(\sqrt{n}\widehat{\psi}_{s,\{ijkl\}}, \sqrt{n}\widehat{\psi}_{s,\{uvwz\}}) = \mu_{ijkluvwz} - \mu_{ijkl}\mu_{uvwz} - \sum_{4}\mu_{ijklu}\mu_{vwz} - \sum_{4}\mu_{iuvwz}\mu_{jkl}\\
&\qquad + \mu_{ijk}\sum_{4}\mu_{vwz}\mu_{lu} + \mu_{ijl}\sum_{4}\mu_{vwz}\mu_{ku} + \mu_{ikl}\sum_{4}\mu_{vwz}\mu_{ju} + \mu_{jkl}\sum_{4}\mu_{vwz}\mu_{iu},
\end{aligned} \tag{48}$$

where the sums are over the different ways of combining the indices in that particular way, for example $\sum_4 \mu_{vwz}\mu_{iju} = \mu_{vwz}\mu_{iju} + \mu_{uwz}\mu_{ijv} + \mu_{uvz}\mu_{ijw} + \mu_{uvw}\mu_{ijz}$, summing the four different ways of multiplying a coskewness element with indices out $\{u, v, w, z\}$ with a coskewness element containing the remaining index together with i and j.

Proof of Theorem 4.4. The proof of asymptotic normality of the NC estimator $\widehat{\boldsymbol{\zeta}}_{\text{nc}}$ consists of two steps. First, asymptotic normality of the estimator $\widehat{\boldsymbol{\theta}}_{\text{nc}}$ is established, according to Theorem 3.2 in Newey & McFadden (1994). Second, asymptotic normality of the NC estimator $\widehat{\boldsymbol{\zeta}}_{\text{nc}}$ follows from the delta-method.

First, the conditions of Theorem 3.2 in Newey & McFadden (1994) are checked. By assumption it holds that $\boldsymbol{\theta}^\star$ lies in the interior of $\boldsymbol{\Theta}$. Also, $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$ is continuously differentiable with respect to $\boldsymbol{\theta}$, as can be seen from their expressions in Boudt et al. (2015). Theorem 4.3 provides asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}_{\boldsymbol{\theta}^\star})$ with mean vector zero and covariance matrix Ξ. The Jacobian function $\boldsymbol{G}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}^\star$ and independent of the sample. We assume that $\boldsymbol{W}$ is such that $\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}$ is nonsingular, with $\boldsymbol{G} = \boldsymbol{G}(\boldsymbol{\theta}^\star)$. It holds that $\widehat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W}$ is a positive semi-definite matrix by the assumptions of the estimator. Finally, when $\widehat{\boldsymbol{\theta}}_{\text{nc}} \xrightarrow{p} \boldsymbol{\theta}^\star$ it follows from Theorem 3.2 in Newey & McFadden (1994) that

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\text{nc}} - \boldsymbol{\theta}^\star\right) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1} \boldsymbol{G}'\boldsymbol{W}\Xi\boldsymbol{W}\boldsymbol{G}\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}\right), \quad n \to \infty. \tag{49}$$

Second, since the Jacobian function $\boldsymbol{G}(\boldsymbol{\theta})$ is continuous at $\boldsymbol{\theta}^\star$, Theorem (4.4) follows from the multivariate Delta-method (see e.g. Theorem A in Section 3.3 of Serfling (2009)).

Note that in the first step it was assumed that $\widehat{\boldsymbol{\theta}}_{\text{nc}} \xrightarrow{p} \boldsymbol{\theta}^\star$. This is not trivial. Under the assumptions of Theorem 2.1 it holds that $\boldsymbol{\theta}^\star$ is identifiable. Hence there exists compact $\boldsymbol{\Theta}$ such that $Q(\boldsymbol{\theta}) = (\boldsymbol{\zeta} - \boldsymbol{\zeta}_{\boldsymbol{\theta}})'\boldsymbol{W}(\boldsymbol{\zeta} - \boldsymbol{\zeta}_{\boldsymbol{\theta}})$ is uniquely maximized at $\boldsymbol{\theta}^\star$. The function $Q(\boldsymbol{\theta})$ is continuous due to continuity of $\boldsymbol{\zeta}_{\boldsymbol{\theta}}$. Due to the moment conditions and compactness of $\boldsymbol{\Theta}$ it also holds that $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}_{\boldsymbol{\theta}}\|] < \infty$. Hence, $(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}_{\boldsymbol{\theta}})'\widehat{\boldsymbol{W}}(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}_{\boldsymbol{\theta}})$ converges uniformly in probability to $Q(\boldsymbol{\theta})$. Thus, by Theorem 2.1 in Newey & McFadden (1994) it follows that $\widehat{\boldsymbol{\theta}}_{\text{nc}} \xrightarrow{p} \boldsymbol{\theta}^\star$.

Proof of Theorem 4.5. The proof is identical to the proof of Theorem 5.2 in Newey & McFadden (1994), but given here for completeness. Let $\boldsymbol{Z}$ be any (mean zero) random vector such that $\Xi = \mathbb{E}[\boldsymbol{Z}\boldsymbol{Z}']$ and let $\boldsymbol{m} = \boldsymbol{G}'\boldsymbol{W}\boldsymbol{Z}$ and $\overline{m} = \boldsymbol{G}'\Xi^{-1}\boldsymbol{Z}$. Then by $\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G} = \mathbb{E}[\boldsymbol{m}\overline{\boldsymbol{m}}']$ and $\boldsymbol{G}'\Xi^{-1}\boldsymbol{G} = \mathbb{E}[\overline{\boldsymbol{m}}\overline{\boldsymbol{m}}']$, it holds that $\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}\boldsymbol{G}'\boldsymbol{W}\Xi\boldsymbol{W}\boldsymbol{G}\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1} - \left(\boldsymbol{G}'\Xi^{-1}\boldsymbol{G}\right)^{-1} = \left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}\mathbb{E}[\boldsymbol{U}\boldsymbol{U}']\left(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}\right)^{-1}$, with $\boldsymbol{U} = \boldsymbol{m} - \mathbb{E}[\boldsymbol{m}\overline{\boldsymbol{m}}']\left(\mathbb{E}[\overline{\boldsymbol{m}}\overline{\boldsymbol{m}}']\right)^{-1}\overline{\boldsymbol{m}}$. Since $\mathbb{E}[\boldsymbol{U}\boldsymbol{U}']$ is positive semi-definite, the difference of the asymptotic variances is positive semi-definite.

Proof of Corollary 4.1. This proof follows the outline of Newey & McFadden (1994). From Theorem 4.3 and Theorem 4.4 it follows that

$$\sqrt{n}\left(\widehat{\boldsymbol{\zeta}}_s - \widehat{\boldsymbol{\zeta}}_{\text{nc}}\right) = \boldsymbol{\Xi}^{1/2}\left(\boldsymbol{I} - \boldsymbol{\Xi}^{-1/2}\boldsymbol{G}'\left(\boldsymbol{G}'\boldsymbol{\Xi}^{-1}\boldsymbol{G}\right)^{-1}\boldsymbol{G}\boldsymbol{\Xi}^{-1/2}\right)\mathcal{U}_n + o_P(1), \tag{50}$$

where $\mathcal{U}_n = \boldsymbol{\Xi}^{-1/2}\sqrt{n}\left(\widehat{\boldsymbol{\zeta}}_s - \boldsymbol{\zeta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Since $\boldsymbol{I} - \boldsymbol{\Xi}^{-1/2}\boldsymbol{G}'\left(\boldsymbol{G}'\boldsymbol{\Xi}^{-1}\boldsymbol{G}\right)^{-1}\boldsymbol{G}\boldsymbol{\Xi}^{-1/2}$ is idempotent of rank $\sharp\boldsymbol{\zeta} - \sharp\boldsymbol{\theta}$, it holds that $n\left(\widehat{\boldsymbol{\zeta}}_s - \widehat{\boldsymbol{\zeta}}_{\text{nc}}\right)'\widehat{\boldsymbol{W}}\left(\widehat{\boldsymbol{\zeta}}_s - \widehat{\boldsymbol{\zeta}}_{\text{nc}}\right) \xrightarrow{d} \chi^2_{\sharp\boldsymbol{\zeta}-\sharp\boldsymbol{\theta}}$, when $n \to \infty$.