An implicit function approach to constrained optimization with applications to asymptotic expansions

Robert J. Boik
Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717-2400, USA

Received 17 July 2006
Available online 19 January 2007

Abstract
In this article, an unconstrained Taylor series expansion is constructed for scalar-valued functions of vector-valued arguments that are subject to nonlinear equality constraints. The expansion is made possible by first reparameterizing the constrained argument in terms of identified and implicit parameters and then expanding the function solely in terms of the identified parameters. Matrix expressions are given for the derivatives of the function with respect to the identified parameters. The expansion is employed to construct an unconstrained Newton algorithm for optimizing the function subject to constraints.

Parameters in statistical models often are estimated by solving statistical estimating equations. It is shown how the unconstrained Newton algorithm can be employed to solve constrained estimating equations. Also, the unconstraining Taylor series is adapted to construct Edgeworth expansions of scalar functions of the constrained estimators. The Edgeworth expansion is illustrated on maximum likelihood estimators in an exploratory factor analysis model in which an oblique rotation is applied after Kaiser row-normalization of the factor loading matrix. A simulation study illustrates the superiority of the two-term Edgeworth approximation compared to the asymptotic normal approximation when sampling from multivariate normal or nonnormal distributions.

© 2007 Elsevier Inc. All rights reserved.

AMS 2000 subject classification: 41A29; 49K30; 62E20; 62F30; 62H25

Keywords: Bordered determinantal criterion; Bordered Hessian; Edgeworth expansions; Estimating functions; Factor analysis; Lagrange multipliers; Matrix derivatives; Second derivative test; Singular value decomposition; Taylor series expansions

E-mail address: rjboik@math.montana.edu.
URL: http://www.math.montana.edu/~rjboik.
1. Introduction

Let \( \mathcal{L}(\theta) \) be a scalar-valued differentiable function of the \( k \)-vector \( \theta \), where \( \theta \in \Theta \subseteq \mathbb{R}^k \). The components of \( \theta \) are called the nominal parameters. Suppose that it is of interest to maximize or minimize \( \mathcal{L} \) with respect to \( \theta \), subject to \( g(\theta) = 0 \), where \( g \) is a \( q \times 1 \) vector-valued differentiable function of \( \theta \). Denote the constrained parameter space by \( \Theta_g = \{ \theta; \theta \in \Theta, g(\theta) = 0 \} \) and denote the interior points of \( \Theta_g \) as \( \hat{\Theta}_g \). In some simple cases, \( \theta \) can be partitioned as \( \theta = (\theta_1', \theta_2')' \), such that \( g(\theta) = 0 \) is satisfied if and only if \( \theta_2 = h(\theta_1) \), where \( h \) is an explicit function. In these cases, \( \mathcal{L} \) can be written as a function of \( \theta_1 \) alone and can be optimized without constraints. If \( g(\theta) = 0 \) cannot be solved explicitly, then it is conventional to optimize \( \mathcal{L} \) by employing Lagrange multipliers. For this approach, the interior critical points are those that satisfy

\[
\begin{align*}
(a) & \quad \theta \in \hat{\Theta}_g \quad \text{and} \quad \frac{\partial \mathcal{L}(\theta)}{\partial \theta} - \frac{\partial g(\theta)}{\partial \theta} \zeta = 0, \\
(b) & \quad \frac{\partial \mathcal{L}(\theta)}{\partial \theta} = 0,
\end{align*}
\]

where \( \zeta \) is a \( q \)-vector of Lagrange multipliers. Denote the vector space generated by the columns of a matrix, \( M \), by \( \mathcal{R}(M) \) and denote the null space, i.e., the kernel, of \( M \) by \( \mathcal{N}(M) \). Then, condition (b) also can be written as

\[
\frac{\partial \mathcal{L}(\theta)}{\partial \theta} \in \mathcal{R}
\left( \frac{\partial g(\theta)}{\partial \theta} \right) \quad \text{and as} \quad G' \left( \frac{\partial \mathcal{L}(\theta)}{\partial \theta} \right) = 0,
\]

where \( G \) is any full column-rank matrix that satisfies

\[
\mathcal{R}(G) = \mathcal{N} \left( \frac{\partial g(\theta)}{\partial \theta'} \right).
\]

The Lagrange equations are solved using an iterative algorithm.

Let \( S \) be an open set and denote an open neighborhood of the point \( x \) in \( S \) by \( \hat{N}_S(x) \). In Section 3 of this article, a local parameterization for \( \theta \in \hat{N}_{\hat{\Theta}_g}(\theta_0) \) is proposed, where \( \theta_0 \) is an arbitrary point in \( \hat{\Theta}_g \). The parameterization transforms \( \theta \) to \( (\eta, \tau) \), where \( \tau \) is an identified parameter and \( \eta \) is an implicit function of \( \tau \). The local parameterization is employed to construct an unconstrained Taylor series expansion of \( \mathcal{L}(\theta) \) around \( \theta = \theta_0 \) to arbitrary order. An unconstrained Newton algorithm to solve the constrained optimization problem follows from the Taylor series expansion and is described in Section 4. A second derivative test for local extrema also is given in Section 4.

In Section 5, the results from Section 3 and Section 4 are applied to the problem of estimating parameters in a statistical model. An estimating function, \( U(\theta; Y) \), is a \( k \times 1 \) vector-valued function of \( \theta \) and \( Y \), where \( Y \) is a matrix of observable random variables. An unconstrained estimator of \( \theta \) is obtained as the solution to the estimating equation \( U(\theta; Y) = 0 \). See [32,33] for details about estimating equations. If \( \theta \) is subject to \( g(\theta) = 0 \), then the Lagrange multiplier approach for computing the constrained estimator is to simultaneously solve

\[
\begin{align*}
(a) & \quad \theta \in \hat{\Theta}_g \quad \text{and} \quad \frac{\partial U(\theta; Y)}{\partial \theta} - \frac{\partial g(\theta)}{\partial \theta} \zeta = 0, \\
(b) & \quad \frac{\partial U(\theta; Y)}{\partial \theta} = 0,
\end{align*}
\]

where \( \zeta \) is a \( q \)-vector of Lagrange multipliers. In Section 5, it is shown how the unconstrained Newton algorithm in Section 4 can be employed as an alternative approach to solve the constrained estimating equation. Also, the unconstrained Taylor series expansion is employed in Section 5 to
construct an Edgeworth expansion for the distribution of scalar-valued functions of the constrained estimator. The proposed approach is illustrated in Section 6 by obtaining a two-term Edgeworth expansion for the distribution of parameter estimators in the exploratory factor analysis constrained estimator. The proposed approach is illustrated in Section 6 by obtaining a two-term Edgeworth expansion for the distribution of parameter estimators in the exploratory factor analysis model.

Notational conventions and definitions are given in Section 2. The definitions are consistent with those used by [8, Chapter 1; 13]. For completeness, product and chain rules for matrix-valued functions of matrix arguments as well as various Kronecker product identities are given in Appendix A.

2. Notation

Suppose that \( Z \) is an \( a \times b \) matrix-valued function of the \( k \)-vector \( \theta \). In this article, derivatives are arranged as follows:

\[
\frac{\partial Z}{\partial \theta} = \frac{\partial}{\partial \theta} \otimes Z = \sum_{i=1}^{k} e_{i}^{k'} \otimes \frac{\partial Z}{\partial \theta_{i}} \quad \text{and} \quad \frac{\partial^{2} Z}{\partial \theta' \otimes \partial \theta'} = \sum_{i=1}^{k} e_{i}^{k'} \otimes \frac{\partial Z}{\partial \theta_{i}'}, \tag{2}
\]

where \( A \otimes B = \{a_{ij}B\} \) is the right Kronecker product, and \( e_{i}^{k} \) is the \( i \)th column of the identity matrix \( I_{k} \). To simplify the presentation, derivatives of \( Z \) with respect to \( \theta \) are denoted as follows:

\[
D_{Z; \theta}^{(1)} \overset{\text{def}}{=} \frac{\partial Z}{\partial \theta}, \quad D_{Z; \theta'}^{(1)} \overset{\text{def}}{=} \frac{\partial Z}{\partial \theta'}, \quad D_{Z; \theta, \theta'}^{(2)} \overset{\text{def}}{=} \frac{\partial^{2} Z}{\partial \theta \otimes \partial \theta'}, \quad D_{Z; \theta, \theta'}^{(3)} \overset{\text{def}}{=} \frac{\partial^{3} Z}{\partial \theta \otimes \partial \theta' \otimes \partial \theta'},
\]

and so forth, where \( \theta_{o} \) is a point in the domain of \( Z \). If follows from (2) that

\[
D_{Z; \theta', \theta}^{(2)} = \sum_{i=1}^{k} \sum_{j=1}^{k} \left[ e_{i}^{k} \otimes e_{j}^{k'} \otimes \frac{\partial^{2} Z}{\partial \theta_{i} \partial \theta_{j}} \right],
\]

\[
D_{Z; \theta', \theta_{o}, \theta'}^{(3)} = \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{s=1}^{k} \left[ e_{i}^{k} \otimes e_{j}^{k'} \otimes e_{s}^{k'} \otimes \frac{\partial^{3} Z}{\partial \theta_{i} \partial \theta_{j} \partial \theta_{s}} \right],
\]

and so forth. Note that \( e_{i}^{k} \otimes e_{j}^{k'} = e_{j}^{k'} \otimes e_{i}^{k} \). Accordingly,

\[
D_{Z; \theta', \theta}^{(2)} = D_{Z; \theta', \theta}^{(2)}, \quad D_{Z; \theta', \theta_{o}, \theta'}^{(3)} = D_{Z; \theta', \theta_{o}, \theta'}^{(3)} = D_{Z; \theta, \theta', \theta}^{(3)},
\]

and so forth, provided that the required partial derivatives are continuous.

Extensions of elementary vectors are used to simplify expressions for derivatives and other quantities. An elementary matrix, for example, is constructed by adjoining a subset of consecutive columns of an identity matrix. Specifically, let \( f \) be a \( t \times 1 \) vector of positive integers that sum to \( k \) and denote the \( i \)th entry of \( f \) by \( f_{i} \). The elementary matrix \( E_{i; f} \), with dimension \( k \times f_{i} \), together
with two additional extensions of elementary vectors are defined below:

\[
E_{i,f} \overset{\text{def}}{=} \left( 0_{f_i \times a_i} \ I_{f_i} \ 0_{f_i \times (k-a_i-f_i)} \right)', \quad L_{k,21} \overset{\text{def}}{=} \sum_{i=1}^{k} \left( e_{i}^{k} \otimes e_{i}^{k} \right) e_{i}^{k'},
\]

and \(L_{k,22} \overset{\text{def}}{=} \sum_{i=1}^{k} \left( e_{i}^{k} \otimes e_{i}^{k} \right) \left( e_{i}^{k} \otimes e_{i}^{k} \right)' \) where \(a_1 = 0, \ a_i = \sum_{j=1}^{i-1} f_j \) if \(i > 1\).

and \(0_{f_i \times 0}\) is an empty matrix; i.e., \((0_{f_i \times 0} \ I_{f_i}) = (I_{f_i} \ 0_{f_i \times 0}) = I_{f_i}\). If \(f = I_k\), then \(E_{i,f}\) simplifies to \(e_{i}^{k}\). The matrix \(L_{k,22}\) also can be written as \(L_{k,21} L_{k,21}'\) and as \(\text{Diag}(\text{vec}(I_k))\), where the vec operator stacks the columns of a matrix and \(\text{Diag}(a)\) is a diagonal matrix whose \(i\)th diagonal component is the \(i\)th component of the vector \(a\).

Suppose that \(\theta\) is partitioned into \(w\) sub-vectors, i.e., \(\theta = (\theta_{1}', \theta_{2}', \ldots, \theta_{w}')'\), where \(\theta_i\) is \(k_i \times 1\). Define the \(w\)-vector \(k\) as \(k = (k_1 k_2 \cdots k_w)'\). Then,

\[
D_{Z,\theta}^{(1)} = \sum_{i=1}^{w} \left( E_{i,k} \otimes I_a \right) D_{Z,\theta_{i}}^{(1)}, \quad D_{Z,\theta'}^{(1)} = \sum_{i=1}^{w} D_{Z,\theta'_{i}}^{(1)} \left( E_{i,k} \otimes I_b \right),
\]

and so forth, where \(Z\) is an \(a \times b\) matrix function of \(\theta\). Computational effort can be reduced by taking advantage of the symmetries inherent in matrices of continuous second or higher-order derivatives. For example,

\[
D_{Z,\theta'_{i},\theta_{i}'}^{(2)} = D_{Z,\theta'_{i}',\theta_{i}'}^{(2)} \left( I_{(k_i,k_i)} \otimes I_b \right), \quad D_{Z,\theta_{i},\theta_{i}}^{(2)} = \left( I_{(k_i,k_i)} \otimes I_a \right) D_{Z,\theta_{i},\theta_{i}}^{(2)}\]

\[
D_{Z,\theta'_{i}',\theta_{i}'}^{(3)} = D_{Z,\theta'_{i},\theta_{i}}^{(3)} \left( I_{(k_i k_i, k_i)} \otimes I_b \right), \quad \text{and}
\]

\[
D_{Z,\theta_{i}',\theta_{i}}^{(2)} = \left( I_{(k_i k_i, k_i)} \otimes I_a \right) D_{Z,\theta_{i}',\theta_{i}}^{(3)},
\]

where \(I_{(a,b)}\) is the commutation matrix described in [20].

3. Taylor series expansion under a local parameterization

Let \(\theta_0\) be an arbitrary point in \(\Theta_{\Phi}\). Suppose that \(\mathcal{L}\) and \(g\) satisfy the following conditions:

(a) \(\mathcal{L}\) and \(g\) each are \(m \geq 2\) times continuously differentiable at \(\theta_0\);
(b) and \(D_{g;\theta_0}^{(1)}\) has full row-rank.

Condition (b) implies that the full-rank singular value decomposition (svd) of \(D_{g;\theta_0}^{(1)}\) can be written as \(D_{g;\theta_0}^{(1)} = U_0 D_0 F_0'\), where \(U_0\) is a \(q \times q\) orthogonal matrix, \(D_0\) is a \(q \times q\) diagonal matrix
Theorem 1. Suppose that conditions (a) and (b) in (6) are satisfied. Let \( F_0 \) and \( G_0 \) be fixed semi-orthogonal matrices that satisfy \( D^{(1)}_{g, \theta_0'} = U_0 D_0 F_0' \), and \( \mathcal{R}(G_0) = \mathcal{N}(F_0') \), where \( U_0 D_0 F_0' \) is the svd of \( D^{(1)}_{g, \theta_0} \). Define \( \tau_0 \) as in (7). Then, there exists an open neighborhood, \( \hat{N}_{\mathbb{R}^r}(\tau_0) \), such that \( \tau \in \hat{N}_{\mathbb{R}^r}(\tau_0) \) implies that \( g(\theta) = 0 \) can be solved uniquely for \( \eta \), where \( \theta = F_0 \eta + G_0 \tau \). Furthermore, \( \eta = \eta(\tau) \) is \( m \) times differentiable with respect to \( \tau \) at \( \tau_0 \).

Proof. Note that \( D^{(1)}_{g, \eta'} \) evaluated at \( \theta_0 \) is \( D^{(1)}_{g, \theta_0'} D^{(1)}_{\theta_0, \eta_0} = U_0 D_0 \). By the implicit function theorem [10, Section 10.3], nonsingularity of \( U_0 D_0 \) together with continuity of \( D^{(1)}_{g, \eta'} \) is sufficient to ensure that \( g(F_0 \eta + G_0 \tau) = 0 \) can be solved uniquely for \( \eta \). That \( \eta \) is \( m \) times differentiable at \( \tau_0 \) follows from the extended implicit function theorem [21, Theorem A.3, p. 143].

A Newton algorithm to solve for \( \eta \) given \( \tau \) is readily constructed. Let \( \hat{\eta}_0 \) be an initial guess for \( \eta \) (\( \hat{\eta}_0 = \eta_0 = F_0' \theta_0 \) is suitable), and denote the value of \( \eta \) after the \( i \)th iteration by \( \hat{\eta}_i \). Define \( \hat{\theta}_i \) as \( \hat{\theta}_i = F_0 \hat{\eta}_i + G_0 \tau_0 \). Then, the Newton update to \( \hat{\eta}_i \) is

\[
\hat{\eta}_{i+1} = \hat{\eta}_i - \zeta \left( D^{(1)}_{g, \eta' F_0} \right)^{-1} g(\hat{\theta}_i),
\]

where \( \zeta \in (0, 1) \) is chosen to ensure that \( \| g(\hat{\theta}_{i+1}) \| < \| g(\hat{\theta}_i) \| \).

It follows from Theorem 1 that \( \mathcal{L}(\theta) \) is \( m \) times differentiable with respect to \( \tau \) at \( \tau_0 \) and that \( \mathcal{L}(\theta) \) can be expanded in an \( (m - 1) \)-th order Taylor series around \( \theta = \theta_0 \). Expressions for the Taylor series and for the first four derivatives of \( \mathcal{L} \) with respect to \( \tau \) at \( \tau_0 \) are given in Corollary 3.1. The derivative expressions can be verified by using product and chain rules and Kronecker product identities (Appendix A).

Corollary 3.1. If (a) and (b) in (6) are satisfied, then there exists an open neighborhood, \( \hat{N}_{\hat{\theta}_g}(\theta_0) \), such that

\[
\theta \in \hat{N}_{\hat{\theta}_g}(\theta_0)
\]

\[
\implies \mathcal{L}(\theta) = \mathcal{L}(\theta_0) + \sum_{i=1}^{m-1} \left( \frac{\partial^i \mathcal{L}(\theta)}{\partial \tau^i} \right)_{\tau=\tau_0} (\tau - \tau_0)^{\otimes i} + R_m(\tau, \tau_0),
\]

where \( \tau = G_0' \theta, \tau_0 = G_0' \theta_0, \lim_{\tau \to \tau_0} \frac{R_m(\tau, \tau_0)}{\| \tau - \tau_0 \|^{m-1}} = 0 \),

and \( A^{\otimes r} \) is the Kronecker product of \( A \) with itself \( r \) times; i.e., \( A^{\otimes 3} = A \otimes A \otimes A \). Furthermore, if (a) and (b) in (6) are satisfied at \( \theta \in \hat{\theta}_g \) and \( m \geq 5 \), then the first four derivatives of \( \mathcal{L} \)
with respect to $\tau$ can be computed as follows:

\[
D^{(1)}_{L;\tau} = D^{(1)}_{L;\theta'} D^{(1)}_{\theta;\tau'},
\]

\[
D^{(2)}_{L;\tau',\tau} = D^{(1)}_{L;\theta'} D^{(2)}_{\theta;\tau',\tau'} + D^{(2)}_{L;\theta';\theta'} \left( D^{(1)}_{\theta;\tau} \otimes D^{(1)}_{\theta;\tau'} \right),
\]

\[
D^{(3)}_{L;\tau',\tau',\tau} = D^{(1)}_{L;\theta'} D^{(3)}_{\theta;\tau',\tau',\tau'} + D^{(2)}_{L;\theta';\theta'} \left( D^{(2)}_{\theta;\tau',\tau'} \otimes D^{(1)}_{\theta;\tau'} \right) J_{21,v} + D^{(3)}_{L;\theta';\theta'} \left( D^{(1)}_{\theta;\tau'} \otimes D^{(1)}_{\theta;\tau'} \otimes D^{(1)}_{\theta;\tau'} \right), \quad \text{and}
\]

\[
D^{(4)}_{L;\tau',\tau',\tau',\tau'} = D^{(1)}_{L;\theta'} D^{(4)}_{\theta;\tau',\tau',\tau',\tau'} + D^{(2)}_{L;\theta';\theta'} \left( D^{(2)}_{\theta;\tau',\tau'} \otimes D^{(2)}_{\theta;\tau'} \right) J_{22,v} + D^{(3)}_{L;\theta';\theta'} \left( D^{(1)}_{\theta;\tau'} \otimes D^{(1)}_{\theta;\tau'} \otimes D^{(1)}_{\theta;\tau'} \otimes D^{(1)}_{\theta;\tau'} \right), \quad \text{where}
\]

\[
J_{21,v} = I_{(v,v^2)} + (I_v \otimes 2N_v),
\]

\[
J_{22,v} = I_v^4 + (I_v \otimes I_{(v,v^2)}) (I_v^2 \otimes 2N_v),
\]

\[
J_{31,v} = (I_v \otimes I_{(v,v^2)}) + (I_v^2 \otimes 2N_v) + I_{(v,v^3)},
\]

\[
J_{211,v} = (I_v \otimes I_{(v,v^2)}) + (I_v \otimes 2N_v \otimes I_v) + (I_{(v,v^2)} \otimes I_v) \left( I_v \otimes I_{(v,v^2)} \right) + (I_v^2 \otimes 2N_v),
\]

and $N_v = \frac{1}{2} \left( I_v^2 + I_{(v,v)} \right)$.

The series expansion in Corollary 3.1 requires that the derivatives of $\theta$ with respect to $\tau$ be known. These derivatives can be written, at an arbitrary point $\theta \in \Theta_{g}$, as

\[
D^{(r)}_{\theta;\tau} = UDF^{(1)}_{g;\theta} + G \quad \text{and} \quad \frac{\partial^r \theta}{(\partial \tau^r)} \otimes_{\theta} = UDF^{(1)}_{g;\theta} \right\} \otimes_{\theta} G, \quad \text{for} \quad r \geq 2,
\]

where $UDF^{(1)}_{g;\theta}$ is the svd of $D^{(1)}_{g;\theta}$, and $\theta = F\eta + G\tau$. The derivatives of $\eta$ with respect to $\tau$ are obtained by solving

\[
\frac{\partial^r g(\theta)}{(\partial \tau^r)} = 0, \quad r = 1, \ldots, m.
\]

For example,

\[
D^{(1)}_{g;\tau} = 0 \implies D^{(1)}_{g;\theta} D^{(1)}_{\theta;\tau} = 0 \implies UDF^{(1)}_{g;\theta} \left( D^{(1)}_{g;\theta} + G \right) = 0
\]

\[
\implies UDF^{(1)}_{g;\theta} = 0 \implies D^{(1)}_{g;\tau} = 0 \implies D^{(1)}_{\theta;\tau} = G.
\]

In practice, it is common that $\theta$ is partitioned as $\theta = (\theta', \theta'')'$, where $\theta'$ is $k_1 \times 1$ and the constraint function $g(\theta)$ depends solely on $\theta_1$. In these cases, it is convenient to partition $\tau$ as $\tau = (\tau_1', \tau_2')'$,
where $\tau_i$ is $v_i \times 1$, $v_1 = k_1 - q$, $v_2 = k_2$, $\theta_1 = F\eta + G\tau_1$, and $\theta_2 = \tau_2$. Expressions for the first four derivatives of $\theta$ with respect to $\tau$ are given in Theorem 2. The expressions rely on the relations in (4) and (5).

**Theorem 2.** Suppose that (a) and (b) in (6) are satisfied for $m \geq 5$ and that $g(\theta)$ is solely a function of $\theta_1$, where $\theta = (\theta_1', \theta_2')'$ and $\theta_1$ is $k_1 \times 1$. Write the svd of $D_{g; \theta_1}^{(1)}$ as $UDF'$, where $F$ is a $k_1 \times q$ semi-orthogonal matrix with rank $q$. Let $G$ be any $k_1 \times v_1$ semi-orthogonal matrix that satisfies $R(G) = N(F')$. Then, the first four derivatives of $\theta$ with respect to $\tau$ can be written as follows:

$$D_{\theta; \tau'}^{(1)} = E_{1,k}D_{\theta_1; \tau_1}^{(1)} E_{1,\tilde{v}}' + E_{2,k}E_{2,\tilde{v}}' = G_*,$$

$$G_* = G \oplus I_{v_2},$$

$$D_{\theta; \tau', \tau_1}^{(2)} = E_{1,k}D_{\theta_1; \tau_1}^{(2)} (E_{1,\tilde{v}} \otimes E_{1,\tilde{v}})' ,$$

$$D_{\theta; \tau', \tau_1}^{(3)} = E_{1,k}D_{\theta_1; \tau_1}^{(3)} (E_{1,\tilde{v}} \otimes E_{1,\tilde{v}})' ,$$

$$D_{\theta; \tau', \tau_1}^{(4)} = E_{1,k}D_{\theta_1; \tau_1}^{(4)} (E_{1,\tilde{v}} \otimes E_{1,\tilde{v}} \otimes E_{1,\tilde{v}})' ,$$

$$D_{\theta_1; \tau_1}^{(1)} = G_* ,$$

$$D_{\theta_1; \tau_1}^{(2)} = -D_{g; \theta_1}^{(1)+} D_{g; \theta_1}^{(2)} (G \otimes G) ,$$

$$D_{\theta_1; \tau_1}^{(3)} = -D_{g; \theta_1}^{(1)+} D_{g; \theta_1}^{(2)} \left( D_{g; \theta_1}^{(2)} \left( D_{\theta_1; \tau_1}^{(2)} \otimes G \right) J_{21,v_1} \right) + D_{g; \theta_1}^{(3)} \left( G \otimes G \otimes G \right) \right)$$

and

$$D_{\theta_1; \tau_1}^{(4)} = -D_{g; \theta_1}^{(1)+} D_{g; \theta_1}^{(2)} \left( D_{g; \theta_1}^{(2)} \left( D_{\theta_1; \tau_1}^{(2)} \otimes D_{\theta_1; \tau_1}^{(2)} \right) J_{22,v_1} \right) + D_{g; \theta_1}^{(2)} \left( D_{\theta_1; \tau_1}^{(2)} \otimes G \right) J_{31,v_1} + D_{g; \theta_1}^{(3)} \left( D_{\theta_1; \tau_1}^{(2)} \otimes G \otimes G \right) J_{211,v_1} + D_{g; \theta_1}^{(4)} \left( G \otimes G \otimes G \otimes G \right) ,$$

where $E_{i,k}$ is defined in (3), $\oplus$ is the direct sum operator, $D_{g; \theta_1}^{(1)+} = FD^{-1}U'$ is the Moore–Penrose inverse of $D_{g; \theta_1}^{(1)}$, $J_{21,v_1}$, $J_{22,v_1}$, $J_{31,v_1}$, and $J_{211,v_1}$ are defined in Corollary 3.1, $\tilde{k} = (k_1 k_2)'$, $\tilde{v} = (v_1 v_2)'$, $v_1 = k_1 - q$, and $v_2 = k_2$.

For example, Magnus and Neudecker [21, p. 138] examined the function $L(\theta) = \theta' A \theta - 3 = 0$, where

$$A = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix} .$$

In this example, $k = k_1 = 2$, $k_2 = 0$, $q = 1$, $v = v_1 = 1$, and $v_2 = 0$. It follows from Corollary 3.1 that $L(\theta)$ can be expanded around $\theta = \theta_0$, for any point, $\theta_0$, that satisfies $\theta_0' A \theta_0 = 3$. With the
aid of Theorem 2, the required derivatives and other quantities can be shown to be

\[
F_0 = A\theta_0 d_0^{-\frac{1}{2}}, \quad d_0 = \theta'_0 A^2 \theta_0, \\
G_0 = (\cos(z) \; \sin(z))', \quad z = \tan^{-1} \left( -F_{0,1}/F_{0,2} \right), \\
\eta(\tau) = -\tau F_0' A \hat{G}_0 \pm \sqrt{d_1}, \\
d_1 = \tau^2 (F_0' \hat{A} G_0)^2 - (F_0 A F_0) \left[ \tau^2 (G_0 A \hat{G}_0) - 3 \right], \\
D^{(1)}_{\mathcal{L}; \tau_0} = 2\theta'_0 G_0, \\
D^{(2)}_{\mathcal{L}; \tau_0, \tau_0} = 2 \left[ 1 - 3(G_0 AG_0) d_0^{-1} \right], \\
D^{(3)}_{\mathcal{L}; \tau'_0, \tau'_0, \tau'_0} = 6(G_0 AG_0) d_0^{-2} \left[ 3(\theta'_0 A^2 G_0) - (\theta'_0 AG_0) d_0 \right], \\
D^{(4)}_{\mathcal{L}; \tau'_0, \tau'_0, \tau'_0} = 6(G_0 AG_0) d_0^{-3} \left[ (G_0 AG_0) d_0^2 + 4(\theta'_0 AG_0)(\theta'_0 A^3 G_0) d_0 \right. \\
- 12(\theta'_0 A^2 G_0)^2 - 3(G_0 AG_0)(\theta'_0 A^3 \theta_0) \right],
\]

(10)

and the solution for \( \eta \) that is closest to \( \eta_0 \) is chosen.

The accuracy of the Taylor series expansion is illustrated in Fig. 1. The upper left plot displays
the constrained parameter space, \( \hat{\Theta}_g \), within which five points have been selected. The remaining
plots display the approximation error for the first through fourth-order Taylor expansion at each of
the five selected points. For each point, the expansion extends from the point, \( \theta_0 \), to the adjacent
squares in the upper left plot. The accuracy of an expansion depends on the location of \( \theta_0 \) in \( \hat{\Theta}_g \)
and on the distance \( \| \tau - \tau_0 \| \), but in all cases the superiority of the higher-order expansions is
evident.

4. Optimization algorithms and second derivative tests

Algorithms to optimize \( \mathcal{L}(\theta) \) subject to \( g(\theta) = 0 \) are readily constructed by employing the
expansion in Corollary 3.1. As in Theorem 2, it is assumed that \( \theta = (\theta'_1 \theta'_2)' \), and that \( g(\theta) \)
depends solely on \( \theta_1 \). Denote the values of \( \tau \) and \( \theta \) after the \( j \)th iteration by \( \hat{\tau}_{\cdot,j} = (\hat{\tau}_{1,j} \hat{\tau}_{2,j})' \)
and \( \hat{\theta}_{\cdot,j} = (\hat{\theta}_{1,j} \hat{\theta}_{2,j})' \), respectively. Write the full-rank svd of \( D^{(1)}_{g; \theta'_i} \) as \( \hat{U}_j \hat{D}_j \hat{F}_j \) and let \( \hat{G}_j \) be a
semi-orthogonal matrix that satisfies \( \hat{R}(\hat{G}_j) = N(\hat{F}_j) \). Using Corollary 3.1 and Theorem 2, the
two-term expansion of \( \mathcal{L}(\theta) \) around \( \theta = \hat{\theta}_{\cdot,j} \) is

\[
\mathcal{L}(\theta) = \mathcal{L}(\hat{\theta}_{\cdot,j}) + D^{(1)}_{\mathcal{L}; \hat{\tau}_{\cdot,j}} (\tau - \hat{\tau}_{\cdot,j}) \\
+ \frac{1}{2} (\tau - \hat{\tau}_{\cdot,j})' D^{(2)}_{\mathcal{L}; \hat{\tau}_{\cdot,j}, \hat{\tau}_{\cdot,j}} (\tau - \hat{\tau}_{\cdot,j}) + o \left( \| \tau - \hat{\tau}_{\cdot,j} \|^2 \right), \quad \text{where}
\]

\[
D^{(1)}_{\mathcal{L}; \tau} = D^{(1)}_{\mathcal{L}; \theta} G_{\ast}, \\
D^{(2)}_{\mathcal{L}; \tau, \tau'} = G_{\ast} D^{(2)}_{\mathcal{L}; \theta, \theta} G_{\ast} - \left( E_{1,\hat{i}} G' \otimes D^{(1)}_{\mathcal{L}; \theta_i} D^{(1)}_{g; \theta'_i} \right) D^{(2)}_{g; \theta_1, \theta'_i} G_{E_{1,\hat{i}}},
\]

(11)
Fig. 1. Accuracy of Taylor series expansion.
and $G_*$ is defined in Theorem 2. The second derivatives in (11) also can be computed as

$$D_{\mathcal{L}; \tau, \tau'}^{(2)} = \text{dvec} \left( D_{\mathcal{L}; \tau, \tau'}^{(2)} \right) \quad \text{and} \quad D_{g; \theta_1, \theta_1'}^{(2)} = \text{dvec} \left( D_{g; \theta_1, \theta_1'}^{(2)} \right),$$

(12)

where $D_{\mathcal{L}; \tau, \tau'}^{(2)}$ is given in Corollary 3.1, and dvec$(M, a, b)$ is the $a \times b$ matrix that satisfies vec$(\text{dvec}(M, a, b)) = \text{vec}(M)$ provided that vec$M$ has dimension $ab \times 1$. A Newton algorithm based on the quadratic approximation is given in Theorem 3. Corollaries 4.1, and 4.2 describe second derivative tests.

**Theorem 3.** If $\hat{\theta} \in \hat{\Theta}_g$ and conditions (a) and (b) in (6) hold at $\hat{\theta}$, then a necessary condition for $\hat{\theta}$ to be a restricted optimizer of $\mathcal{L}(\theta)$ is

$$\left( \hat{G} \oplus I_{v_2} \right) D_{\mathcal{L}; \hat{\theta}}^{(1)} = 0,$$

where $\hat{G}$ is a full column-rank matrix that satisfies $\mathcal{R}(\hat{G}) = \mathcal{N}(\hat{F})$ and $\hat{U}D\hat{F}^\dagger$ is the svd of $D_{g; \hat{\theta}_1}^{(1)}$.

A solution to the first-order equation can be obtained as follows. Denote the value of $\hat{\theta}$ after the $j$,th iteration by $\hat{\theta}_{., j}$, write the svd of $D_{g; \hat{\theta}_1}^{(1)}$ as $U_j D_j F_j$, and let $G_j$ be a semi-orthogonal matrix that satisfies $\mathcal{R}(G_j) = \mathcal{N}(F'_j)$. Then, the Newton update to $\hat{\theta}_{., j}$ is

$$\hat{\theta}_{., j+1} = \left( \hat{F} \hat{\eta}_{j+1} + \hat{G}_j \hat{\tau}_{1,j+1} \right) \frac{\hat{\tau}_{2,j+1}}{\hat{\tau}_{2,j+1}}$$

where

$$\hat{\tau}_{., j+1} = \hat{\tau}_{., j} - \left( D_{\mathcal{L}; \hat{\tau}_{., j}, \hat{\tau}_{., j}}^{(2)} + \alpha I_v \right)^{-1} D_{\mathcal{L}; \hat{\tau}_{., j}}^{(1)},$$

$\hat{\eta}_{j+1}$ is the solution to $g(\hat{F} \hat{\eta}_{j+1} + \hat{G}_j \hat{\tau}_{1,j+1}) = 0$, $\alpha$ is an adjustable parameter that can be chosen to ensure that $\mathcal{L}(\hat{\theta}_{., j+1})$ is increasing (or decreasing) in $j$, and the remaining terms are defined in (11). The vector $\hat{\eta}_{j+1}$ can be computed using the algorithm in (8) after substituting $\theta_1$ for $\theta$ and $\hat{F}_j$ for $F_0$.

**Corollary 4.1.** Suppose that $\hat{\theta} \in \hat{\Theta}_g$, $\hat{\theta}$ satisfies the first-order condition in Theorem 3, and that (a) and (b) in (6) hold at $\hat{\theta}$. Then, a necessary condition for $\hat{\theta}$ to be a local maximizer (minimizer) is that the Hessian, $D_{\mathcal{L}; \hat{\tau}, \hat{\tau}'}^{(2)}$ be negative (positive) semi-definite, where

$$D_{\mathcal{L}; \hat{\tau}, \hat{\tau}'}^{(2)} = \hat{G}_*^{(2)} \left[ D_{\mathcal{L}; \hat{\tau}, \hat{\tau}'}^{(2)} - E_1 k \left( I_{k_1} \otimes \hat{\zeta} \right) D_{g; \hat{\theta}_1, \hat{\theta}_1'}^{(2)} E_1 k \right] \hat{G}_*^{(2)}, \quad \hat{\zeta}' = D_{\mathcal{L}; \hat{\theta}_1}^{(1)} D_{g; \hat{\theta}_1}^{(1)+},$$

and $\hat{G}_* = \hat{G} \oplus I_{v_2}$. Also, a sufficient condition for $\hat{\theta}$ to be a local maximizer (minimizer) is that the Hessian be negative (positive) definite. If $g(\theta)$ is a function of the entire vector $\theta$, then the Hessian simplifies to

$$D_{\mathcal{L}; \hat{\tau}, \hat{\tau}'}^{(2)} = \hat{G} \left[ D_{\mathcal{L}; \hat{\tau}, \hat{\tau}'}^{(2)} - \left( I_k \otimes \hat{\zeta} \right) D_{g; \hat{\theta}_1, \hat{\theta}_1'}^{(2)} \right] \hat{G} \quad \text{where} \quad \hat{\zeta}' = D_{\mathcal{L}; \hat{\theta}_1}^{(1)} D_{g; \hat{\theta}_1}^{(1)+}.
Note that the quantity in brackets \([\cdots]\) in Corollary 4.1 is the upper left-hand submatrix of the bordered Hessian, \(D^{(2)}_{L(\theta)-\zeta g(\theta);\omega,\omega'}\), evaluated at \(\theta = \hat{\theta}\) and \(\zeta = \hat{\zeta}\), where
\[
D^{(2)}_{L(\theta)-\zeta g(\theta);\omega,\omega'} = \begin{pmatrix}
\hat{\zeta}' \left[ L(\theta)-\zeta g(\theta) \right] & -D^{(1)}_{g;\theta'} \\
\hat{\theta} \otimes \hat{\theta}' & -D^{(1)}_{g;\theta'} \\
-D^{(1)}_{g;\theta'} & 0
\end{pmatrix}
\text{ and } \omega = \begin{pmatrix} \theta \\ \zeta \end{pmatrix}.
\] (13)

The bordered determinantal criterion is an alternative statement of the necessary (sufficient) condition in Corollary 4.1. This criterion requires one to determine the signs of the \(v = k - q\) leading principal minors of the bordered Hessian matrix. Details are given in [19, pp. 290–291, 31, Appendix C]. The criterion in Corollary 4.1 is, arguably, more straightforward. Proofs of the bordered determinantal criterion were given in [21, pp. 136–138, 12, pp. 96–100]. A criterion similar to that in Corollary 4.1 was given by Im [14], except that Im required that the first \(q\) columns of \(D^{(1)}_{g;\theta'}\) be linearly independent.

**Corollary 4.2.** If the constraints on \(\theta\) are linear, then the Hessian simplifies to
\[
D^{(2)}_{L;\hat{\theta},\hat{\theta}} = G^* D^{(2)}_{L;\hat{\theta},\hat{\theta}} G^* \text{ or to } D^{(2)}_{L;\hat{\theta},\hat{\theta}} = G' D^{(2)}_{L;\hat{\theta},\hat{\theta}} G,
\]
if \(\theta\) is not partitioned, where \(G^*\) is defined in Corollary 4.1.

Note that if \(D^{(2)}_{L;\hat{\theta},\hat{\theta}}\) is negative (positive) definite and constraints are linear, then the Hessian also is negative (positive) definite. Accordingly, negative (positive) definiteness of \(D^{(2)}_{L;\hat{\theta},\hat{\theta}}\) is sufficient for \(\hat{\theta}\) to be a local maximizer (minimizer) if constraints are linear.

As an application of Theorem 3, consider optimizing \(L(\theta) = \theta' \theta\) subject to \(g(\theta) = \theta' A \theta - 3 = 0\), where \(A\) is given in (9). Of course, an iterative algorithm is not needed for this problem. The constrained maximum and minimum of \(L\) correspond to the eigenvalues of \(3A^{-1}\) and the constrained optimizers are the scaled eigenvectors. Nonetheless, the optimizers can be computed by the Newton algorithm in Theorem 3. In this example, neither \(\theta\) nor \(\tau\) are partitioned, and the Newton update is
\[
\hat{\theta}_{j+1} = \hat{F} \hat{\theta}_j + \hat{G} \hat{\tau}_{j+1}, \quad \text{where}
\]
\[
\hat{\tau}_{j+1} = \hat{\tau}_j - \frac{(\hat{\theta}_j' A^2 \hat{\theta}_j) \hat{\theta}_j' \hat{G} \hat{G}_j}{(\hat{\theta}_j' A^2 \hat{\theta}_j) - 3(\hat{G}_j' A \hat{G}_j)},
\]
and \(\hat{\theta}_{j+1}\) can be computed as a function \(\hat{\tau}_{j+1}\) as in (10). Each of the 10 points marked by the circles and squares in the upper left plot in Fig. 1 were used as starting points. In all cases, the algorithm converged to the nearest optimizer (1 1)', (−1 − 1)', (\(\sqrt{3} - \sqrt{3}\))', or (−\(\sqrt{3}\) \(\sqrt{3}\))' with at least six decimal places of accuracy in five or fewer iterations. The values of the Hessian, evaluated at the solutions are
\[
D^{(2)}_{L;\hat{\theta},\hat{\theta}} = 2 \left( 1 - 3 \frac{\hat{G}' A \hat{G}}{\hat{\theta}' A^2 \hat{\theta}} \right) = \begin{cases} 
1/3 & \text{if } \hat{\theta} \text{ is a minimizer, and} \\
-4 & \text{if } \hat{\theta} \text{ is a maximizer.}
\end{cases}
\]

The Newton update in Theorem 3 is similar to the update based on the conventional Lagrange multiplier approach. Accordingly, the ease of implementation and the performance of the proposed
optimization algorithm are similar to those of Lagrange-based algorithms. Nonetheless, the proposed approach does have advantages. First, the proposed approach yields an easily implemented second derivative test (see Corollary 4.1). Second, and more importantly, the proposed approach yields simple matrix expressions for quantities required in higher-order expansions (see Section 3). These matrix expressions enjoy a reduced-dimension property compared to those based on Lagrange multipliers because the dimension of the parameter space is reduced rather than increased by the constraints. The proposed approach is especially useful for obtaining asymptotic expansions of statistical distributions based on solving estimating equations.

5. Edgeworth expansions for functions of parameter estimators

In this section, it is assumed that a statistical model for $Y_1, \ldots, Y_N$ has been proposed, where $\{Y_i\}_{i=1}^N$ are independently distributed random $p$-vectors whose cumulative distribution functions are continuous. The statistical model depends on a $k$-vector of unknown parameters, $\theta$, where $\theta$ is constrained by $g(\theta) = 0$. It is of interest to estimate $\theta$ and to approximate the distribution of the estimator.

In some applications, say type A, the estimator is obtained as a constrained maximizer of the log likelihood function or as a constrained minimizer of a discrepancy function. In these applications, the objective function is denoted as $L(\theta; Y, X)$, where $L$ has magnitude $O_p(N)$, $Y$ is the $N \times p$ matrix whose $i$th row is $Y_i$ and $X$ is an $N \times c$ matrix of known constants whose $i$th row is $x_i$.

The estimator, $\hat{\theta}$, is obtained as a solution to the estimating equation

$$D_{L, \tau}^{(1)} = 0 \quad \text{where} \quad D_{L, \tau}^{(1)} = D_{\theta, \tau}^{(1)} D_{\rho, \tau}^{(1)}.$$

In other applications, say type B, an objective function $L$ does not exist and the estimator is obtained as a solution to

$$U(\theta; Y, X) - D_{L, \tau}^{(1)} \zeta = 0 \quad \text{subject to} \quad g(\theta) = 0,$$

where

$$U(\theta; Y, X) = \sum_{i=1}^N \psi(\theta; Y_i, x_i),$$

$\psi$ is a known $k \times 1$ vector-valued function (e.g., see [26]), and $\zeta$ is a $q$-vector of Lagrange multipliers.

The two types of estimating functions can be treated identically by redefining $D_{L, \theta}^{(1)}$ and $D_{L, \tau}^{(1)}$ for type B functions as

$$D_{L, \theta}^{(1)} \overset{\text{def}}{=} \sum_{i=1}^N \psi(\theta; Y_i, x_i) \quad \text{and} \quad D_{L, \tau}^{(1)} \overset{\text{def}}{=} D_{\theta, \tau}^{(1)} \sum_{i=1}^N \psi(\theta; Y_i, x_i).$$

Quantities such as $D_{L, \theta, \theta}^{(2)}$ and $D_{L, \tau, \tau}^{(2)}$ are then obtained by computing derivatives in the usual manner. In particular, the estimator, $\hat{\theta}$, is obtained as the solution to the estimating equation $D_{L, \tau}^{(1)} = 0$ in type A as well as in type B applications. In either case, the Newton algorithm in Theorem 3 can be used to solve the estimating equation.

The expansion to be developed below requires that the estimating function $D_{L, \tau}^{(1)}$ satisfy certain regularity conditions; e.g., existence of continuous derivatives and existence of finite joint cumulants of the derivatives. Specific validity conditions are described in [28, pp. 80–81] in the
Corollary 5.1. If \( \hat{\tau} \) is a consistent solution to the estimating equation and the regularity conditions are satisfied, then

\[
\frac{1}{\sqrt{n}} D_{L; \tau}^{(1)} = 0 = \frac{1}{\sqrt{n}} D_{L; \tau}^{(1)} + \frac{1}{n} D_{L; \tau, \tau'} \sqrt{n} (\hat{\tau} - \tau)
\]

\[
+ \frac{1}{2n^2} D_{L; \tau, \tau', \tau'} \left[ \sqrt{n} (\hat{\tau} - \tau) \right] \otimes 2
\]

\[
+ \frac{1}{6n^2} D_{L; \tau, \tau', \tau'} \left[ \sqrt{n} (\hat{\tau} - \tau) \right] \otimes 3 + O_p \left( n^{-\frac{1}{2}} \right),
\]

where

\[
D_{L; \tau}^{(1)} = \begin{cases}
D_{\theta, \tau}^{(1)} D_{L, \theta}^{(1)} & \text{in type A cases}, \\
D_{\theta, \tau}^{(1)} \sum_{i=1}^{N} \psi(\theta; Y_i, x_i) & \text{in type B cases},
\end{cases}
\]

\[
D_{L; \tau, \tau', \tau'}^{(3)} = \text{dvec} \left( D_{L; \tau, \tau', \tau'; \nu, \nu^2} \right),
\]

\[
D_{L; \tau, \tau', \tau'}^{(4)} = \text{dvec} (D_{L; \tau, \tau', \tau'; \nu, \nu^3}),
\]

\( D_{L; \tau, \tau'}^{(2)} \) is given in (11), the dvec operator is defined in (12), \( n = N - r \), and the remaining terms are defined in Corollary 3.1 and Theorem 2.

An expansion for \( \sqrt{n} (\hat{\tau} - \tau) \) can be constructed by solving the equation in Corollary 5.1 for \( \sqrt{n} (\hat{\tau} - \tau) \). A solution can be obtained in the following two steps.

First, define \( \{K_i\}_{i=1}^{4} \) and \( \{Z_i\}_{i=1}^{3} \) as

\[
K_1 \equiv \frac{1}{n} \mathbb{E} \left( D_{L; \tau}^{(1)} \right), \quad K_2 \equiv \frac{1}{n} \mathbb{E} \left( D_{L; \tau, \tau'}^{(2)} \right), \quad K_3 \equiv \frac{1}{n} \mathbb{E} \left( D_{L; \tau, \tau'}^{(3)} \right),
\]

\[
K_4 \equiv \frac{1}{n} \mathbb{E} \left( D_{L; \tau, \tau', \tau'}^{(4)} \right), \quad Z_1 \equiv \sqrt{n} \left( \frac{1}{n} D_{L; \tau}^{(1)} - K_1 \right),
\]

\[
Z_2 \equiv \sqrt{n} \left( \frac{1}{n} D_{L; \tau, \tau'}^{(2)} - K_2 \right) \quad \text{and} \quad Z_3 \equiv \sqrt{n} \left( \frac{1}{n} D_{L; \tau, \tau'}^{(3)} - K_3 \right). \tag{14}
\]

It follows from the regularity conditions that \( Z_i = O_p(1) \) for all \( i \), \( K_i = O(1) \) for \( i \geq 2 \), and \( K_2 \) is nonsingular whenever \( \theta \) is identified. An estimating function is said to be unbiased if \( K_1 = 0 \) [11]. Derivatives of log likelihood functions possess this property. Nonetheless, if \( K_1 = O(n^{-1}) \), then the expansions to be developed are still valid.
Second, use (14) to express the derivatives of $\mathcal{L}$ as functions of $K_i$ and $Z_i$; e.g., $D_{\mathcal{L},\tau',\tau}^{(2)} = \sqrt{n}Z_2 + nK_2$ and then write $\sqrt{n} (\hat{\tau} - \tau)$ as $\sqrt{n} (\hat{\tau} - \tau) = \hat{\delta}_1 + n^{-\frac{1}{2}}\hat{\delta}_2 + n^{-\frac{1}{2}}\hat{\delta}_3 + O_p(n^{-3/2})$, where $\hat{\delta}_i$ for $i = 1, 2, 3$ are $O_p(1)$ quantities whose values are to be determined. Making these substitutions in Corollary 5.1 and collecting terms of like order reveals that

$$0 = \left[ Z_1 + K_2\hat{\delta}_1 \right] + \frac{1}{\sqrt{n}} \left[ nK_1 + K_2\hat{\delta}_2 + Z_2\hat{\delta}_1 + \frac{1}{2} K_3 (\hat{\delta}_1 \otimes \hat{\delta}_1) \right]$$

$$+ \frac{1}{n} \left[ K_2\hat{\delta}_3 + Z_2\hat{\delta}_2 + K_3 (\hat{\delta}_1 \otimes \hat{\delta}_2) + \frac{1}{2} Z_3 (\hat{\delta}_1 \otimes \hat{\delta}_1) + \frac{1}{6} K_4 (\hat{\delta}_1 \otimes \hat{\delta}_1 \otimes \hat{\delta}_1) \right]$$

$$+ O_p \left( n^{-\frac{3}{2}} \right).$$

(15)

The right-hand side of (15) is 0 if and only if the $O_p(n^{-1/2})$ subexpression is zero for $i = 0, 1, 2$. Accordingly,

$$\hat{\delta}_1 = -K_2^{-1}Z_1,$$

$$\hat{\delta}_2 = -K_2^{-1} \left[ nK_1 + Z_2\hat{\delta}_1 + \frac{1}{2} K_3 (\hat{\delta}_1 \otimes \hat{\delta}_1) \right]$$

and

$$\hat{\delta}_3 = -K_2^{-1} \left[ Z_2\hat{\delta}_2 + K_3 (\hat{\delta}_1 \otimes \hat{\delta}_2) + \frac{1}{2} Z_3 (\hat{\delta}_1 \otimes \hat{\delta}_1) + \frac{1}{6} K_4 (\hat{\delta}_1 \otimes \hat{\delta}_1 \otimes \hat{\delta}_1) \right].$$

(16)

An investigator may be interested in a specific component of $\theta$ or, more generally, in a vector-valued function of $\theta$, say $F(\theta)$. A Taylor series expansion for the estimator of $F$ in terms of $(Z_i, K_i)$ is readily constructed using (16). The expansion is summarized in Theorem 4. The asymptotic distribution of $\hat{F}$ is given in Corollary 5.2. The two term Edgeworth expansion for the distribution of $\sqrt{n}(\hat{F} - F)$ in the special case where $F$ is scalar-valued is given in Corollary 5.3.

**Theorem 4.** Suppose that $F = F(\theta)$ has continuous derivatives up to order four. Define $\hat{F}$ as $\hat{F} \equiv F(\hat{\theta})$, where $\hat{\theta} = \hat{\theta}_0 + \hat{\theta}_1$, and $\hat{\theta}$ is a consistent solution to the estimating equation. If the regularity conditions are satisfied, then

$$\sqrt{n}(\hat{F} - F) = Q_1 + \frac{1}{\sqrt{n}} Q_2 + \frac{1}{n} Q_3 + O_p \left( n^{-\frac{3}{2}} \right)$$

where

$$Q_1 = D^{(1)}_{F,\theta} \hat{\theta}_1,$$

$$Q_2 = D^{(1)}_{F,\theta} \hat{\theta}_2 + \frac{1}{2} D^{(2)}_{F,\theta} \hat{\theta}_1 \otimes \hat{\theta}_1,$$

$$Q_3 = D^{(3)}_{F,\theta} \hat{\theta}_3 + D^{(2)}_{F,\theta} \hat{\theta}_2 \otimes \hat{\theta}_1 + \frac{1}{6} D^{(3)}_{F,\theta} \hat{\theta}_2 \otimes \hat{\theta}_1 \otimes \hat{\theta}_1,$$

$$D^{(1)}_{F,\theta} = D^{(1)}_{F,\theta} \hat{\theta}_1 \otimes \hat{\theta}_1,$$

$$D^{(2)}_{F,\theta} = D^{(2)}_{F,\theta} \hat{\theta}_1 \otimes \hat{\theta}_1 \otimes \hat{\theta}_1,$$

$$D^{(3)}_{F,\theta} = D^{(3)}_{F,\theta} \hat{\theta}_2 \otimes \hat{\theta}_1 \otimes \hat{\theta}_1 \otimes \hat{\theta}_1,$$

where $\hat{\theta}_i$ for $i = 1, 2, 3$ are defined in (16), and $J_{21,v}$ is defined in Corollary 3.1.

**Corollary 5.2.** Under the conditions of Theorem 4, the asymptotic distribution of $\sqrt{n}(\hat{F} - F)$ is $N(0, \Sigma_\beta)$, where

$$\Sigma_\beta = E(Q_1'Q_1) = D^{(1)}_{F,\theta} G_\alpha K_2^{-1} E(Z_iZ_i') K_2^{-1} G_\alpha D^{(1)}_{F,\theta},$$

where $\Sigma_\beta$ is the covariance matrix of $\beta$. This approximation is valid for large samples and can be used to construct asymptotic confidence intervals for $\beta$.
Then the asymptotic distribution of \( \sqrt{n}(\hat{\theta} - \theta) \) is \( N(0, \Sigma_{\theta}) \), where

\[
\Sigma_{\theta} = G_s K_2^{-1} E(Z_i Z'_i) K_2^{-1} G_s.
\]

**Corollary 5.3.** Suppose that \( \beta \) is a scalar-valued function of \( \theta \). Denote the pdf and cdf of \( Z = \sqrt{n}(\hat{\beta} - \beta)/\sigma_\beta \) by \( f_n(z) \) and \( F_n(z) \), respectively, where \( \sigma_\beta^2 \) is the asymptotic variance of \( \sqrt{n}(\hat{\beta} - \beta) \). Then, under the conditions of Theorem 4, \( f_n(z) \) and \( F_n(z) \) can be expanded as follows:

\[
f_n(z) = f(z) \left[ 1 + \frac{\omega_1 H_1(z)}{\sqrt{n} \sigma_\beta} + \frac{\omega_2 H_2(z)}{2n \sigma_\beta^2} + \frac{\omega_3 H_3(z)}{6 \sqrt{n} \sigma_\beta^3} + \frac{\omega_4 H_4(z)}{24n \sigma_\beta^4} + \frac{\omega_5 H_5(z)}{72n \sigma_\beta^5} \right] + O \left( n^{-\frac{3}{2}} \right)
\]

and

\[
F_n(z) = F(z) - f(z) \left[ \frac{\omega_1 H_1(z)}{\sqrt{n} \sigma_\beta} + \frac{\omega_2 H_2(z)}{2n \sigma_\beta^2} + \frac{\omega_3 H_3(z)}{6 \sqrt{n} \sigma_\beta^3} + \frac{\omega_4 H_4(z)}{24n \sigma_\beta^4} + \frac{\omega_5 H_5(z)}{72n \sigma_\beta^5} \right] + O \left( n^{-\frac{3}{2}} \right),
\]

where \( f(z) \) is the standard normal pdf, \( F(z) \) is the standard normal cdf, \( H_i(z) \) is the Hermite polynomial of order \( i \), \( \sigma_\beta^2 = \text{D}(1)_{\beta; \gamma} K_2^{-1} E(Z_i Z'_i) K_2^{-1} \text{D}(1)_{\beta; \gamma} \), and \( \{\omega_i\}_{i=1}^4 \) are \( O(1) \) cumulant functions that can be computed as

\[
\begin{align*}
\omega_1 &= E(Q_2), \\
\omega_2 &= E(Q_2^2) + 2\sqrt{n}E(Q_1 Q_2) + 2E(Q_1 Q_3), \\
\omega_3 &= \sqrt{n}E(Q_3) + 3E(Q_1^2 Q_2) - 3E(Q_1^2)E(Q_2) \quad \text{and} \\
\omega_4 &= 4\omega_1 \omega_3 + n \left[ E(Q_1^3) - 3E(Q_1^2)^2 \right] \\
&\quad + \sqrt{n} \left[ 4E(Q_1^3 Q_2) - 12E(Q_1^2)E(Q_1 Q_2) - 4E(Q_1^3)E(Q_2) \right] \\
&\quad + 4E(Q_1^3 Q_3) + 6E(Q_1^2 Q_2^2) - 6E(Q_1^2)E(Q_2^2) - 12E(Q_1^3)E(Q_1 Q_3) \\
&\quad - 12E(Q_2)E(Q_1^2 Q_2) + 12E(Q_2^3)E(Q_2)^2.
\end{align*}
\]

If \( \mathcal{L} \) is a log likelihood function, then the expansions in Corollaries 5.2 and 5.3 can be simplified. Denote the average Fisher information in the unrestricted model by \( \mathbf{I}_\theta \); i.e.,

\[
\mathbf{I}_\theta = n^{-1} E \left( \mathbf{D}_{\mathcal{L}; \theta}^{(1)} \mathbf{D}_{\mathcal{L}; \theta}^{(1)} \right). \tag{17}
\]

Recall that under the usual regularity conditions,

\[
E \left( \mathbf{D}_{\mathcal{L}; \theta}^{(1)} \right) = 0 \quad \text{and} \quad E \left( \mathbf{D}_{\mathcal{L}; \theta}^{(2)} \right) = -E \left( \mathbf{D}_{\mathcal{L}; \theta}^{(1)} \mathbf{D}_{\mathcal{L}; \theta}^{(1)} \right).
\]

It follows that the asymptotic covariance matrices of \( \sqrt{n}(\hat{\theta} - \theta) \) and \( \sqrt{n}(\hat{\beta} - \beta) \) simplify to

\[
\Sigma_{\theta} = G_s \mathbf{I}_\tau^{-1} G_s^\prime \quad \text{and} \quad \Sigma_{\beta} = \mathbf{D}_{\beta; \theta}^{(1)} \Sigma_\theta \mathbf{D}_{\beta; \theta}^{(1)},
\]

where \( \mathbf{I}_\tau = G_s' \mathbf{I}_\theta G_s \), \( \tag{18} \)

and \( G_s \) is defined in Theorem 2. Expressions for the \( Z_i \) and \( K_i \) terms in (14) also are simplified if \( \mathcal{L} \) is a log likelihood function and the regularity conditions are satisfied. The simplified expressions are listed in Appendix B.
The conventional approach to computing $\Sigma_\theta$ in likelihood-based models is based on the bordered information matrix, namely

$$\mathbf{I}_{\theta,g} = -n^{-1} \mathbb{E} \left( \mathbf{D}^{(2)}_{\theta} - \mathbf{Z}' \mathbf{g}(\theta) : \omega, \omega' \right) |_{\mathbf{z} = 0} = \begin{pmatrix} \mathbf{I}_\theta & n^{-1} \mathbf{D}^{(1)'}_{g; \theta'} \\ n^{-1} \mathbf{D}^{(1)}_{g; \theta} & 0 \end{pmatrix},$$

(19)

where the bordered Hessian, $\mathbf{D}^{(2)}_{\theta} - \mathbf{Z}' \mathbf{g}(\theta) : \omega, \omega'$, is given in (13). Aitchison and Silvey [1] showed that if $\mathbf{I}_\theta$ is invertible, then the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta} - \theta)$ is the $k \times k$ matrix in the upper left-hand corner of $\mathbf{I}_{\theta,g}^{-1}$ and that the covariance matrix can be computed as

$$\Sigma_\theta = \mathbf{I}_\theta^{-1} - \mathbf{I}_\theta^{-1} \mathbf{D}^{(1)'}_{g; \theta'} \left( \mathbf{D}^{(1)}_{g; \theta'} \mathbf{I}_\theta^{-1} \mathbf{D}^{(1)'}_{g; \theta} \right)^{-1} \mathbf{D}^{(1)'}_{g; \theta} \mathbf{I}_\theta^{-1}.$$

Application of Lemma 1 in Khatri [18] reveals that the Aitchison and Silvey covariance matrix is identical to that in (18).

If one or more parameters are not identified, then $\mathbf{I}_\theta$ is not invertible. Nonetheless, Aitchison and Silvey’s result still holds, provided that $\mathbf{I}_{\theta,g}$ in (19) is invertible. Silvey [29,30] showed that in the case of singular $\mathbf{I}_\theta$, but nonsingular $\mathbf{I}_{\theta,g}$, the asymptotic covariance matrix still can be obtained without inverting the $(k + q) \times (k + q)$ bordered information matrix. Arrange the $q$ constraints so that $\mathbf{g}(\theta) = [\mathbf{g}_1(\theta)', \mathbf{g}_2(\theta)']'$, where the first $q_1$ constraints, $\mathbf{g}_1(\theta) = \mathbf{0}$, ensure identifiability of the parameters and the remaining $q_2 = q - q_1$ constraints, $\mathbf{g}_2(\theta) = \mathbf{0}$, impose restrictions on the parameter space. Silvey showed that the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta} - \theta)$ is the $k \times k$ matrix in the upper left-hand corner of $\mathbf{I}_{\theta,g+}^{-1}$, and that the covariance matrix can be computed as

$$\Sigma_\theta = \mathbf{I}_{\theta+}^{-1} - \mathbf{I}_{\theta+}^{-1} \mathbf{D}^{(1)'}_{g; \theta'} \left( \mathbf{D}^{(1)}_{g; \theta'} \mathbf{I}_{\theta+}^{-1} \mathbf{D}^{(1)'}_{g; \theta} \right)^{-1} \mathbf{D}^{(1)'}_{g; \theta} \mathbf{I}_{\theta+}^{-1},$$

where $\mathbf{I}_{\theta,g+}$ is $\mathbf{I}_{\theta,g}$ in (19) in which $\mathbf{I}_\theta$ has been replaced by $\mathbf{I}_{\theta+} = \mathbf{I}_\theta + \mathbf{D}^{(1)'}_{g; \theta'} \mathbf{D}^{(1)}_{g; \theta}$. An alternative proof of Silvey’s result was given by Neuenschwander and Flury [23]. Again, Khatri’s lemma [18] can be used to verify that the Silvey covariance matrix is identical to that in (18). Note that to compute the asymptotic covariance using (18), it is not necessary to distinguish between identification constraints and substantive model constraints. Furthermore, it can be shown that $\mathbf{I}_{\theta,g}$ is invertible if and only if $\mathbf{I}_r$ in (18) is invertible.

6. Application to factor analysis

Let $\mathbf{Y}$ be an $N \times p$ random matrix that can be represented as

$$\mathbf{Y} = \mathbf{X}_1 \mathbf{B} + \mathbf{X}_2 \mathbf{\Gamma}^*_+ + \mathbf{E},$$

(20)

where $\mathbf{X}_1$ is an $N \times v$ matrix of known constants, rank($\mathbf{X}_1$) = $r \leq v$, $\mathbf{B}$ is a $v \times p$ matrix of unknown regression coefficients, $\mathbf{X}_2$ is an $N \times m$ unobservable matrix of random factors whose rows are independently and identically distributed with mean zero and covariance $\mathbf{I}_m$, $\mathbf{\Gamma}^*_+$ is a $p \times m$ matrix of unknown factor loadings, and $\mathbf{E}$ is an $N \times p$ unobservable random matrix distributed independently of $\mathbf{X}_2$. The rows of $\mathbf{E}$ are assumed to be independently and identically distributed with mean zero and diagonal covariance matrix $\mathbf{\Psi}^*_+$. Under these conditions, the first two moments of $\mathbf{Y}$ are $\mathbb{E}(\mathbf{Y}) = \mathbf{X} \mathbf{B}$ and $\text{Var} \left( \text{vec}(\mathbf{Y}) \right) = \mathbf{\Sigma} \otimes \mathbf{I}_N$, where $\mathbf{\Sigma} = \mathbf{\Gamma}^*_+ \mathbf{\Gamma}^*_+ + \mathbf{\Psi}^*_+$.

Often in practice, the scales of the $p$ response variables are arbitrary. In these cases, it is conventional to fit the factor structure to the correlation matrix rather than to the covariance
matrix. Denote the $p \times p$ correlation matrix by $\Lambda$. Let $M$ be a square matrix with positive diagonal components $m_{11}, m_{22}, \ldots$ and define $(M)^a_D$ as
\[
(M)^a_D = \text{Diag}(m_{11}^a, m_{22}^a, \ldots).
\]
Then, the correlation-based model and associated covariance structure are
\[
Y = X_1B + X_2\Gamma'(\Sigma)^{\frac{1}{2}}_D + E \quad \text{and} \quad \text{Var}(\text{vec} \ Y) = \Sigma \otimes I_N,
\]
where $\Sigma = (\Sigma)^{\frac{1}{2}}_D \Lambda (\Sigma)^{\frac{1}{2}}_D$, $\Lambda = \Gamma' + \Psi$, $\Gamma = \Gamma_s (\Sigma)^{-\frac{1}{2}}_D$ and $\Psi = (\Sigma)^{-\frac{1}{2}}_D \Psi_s (\Sigma)^{-\frac{1}{2}}_D$.

The exploratory factor analysis parameters are not identified unless constraints are imposed on $\Gamma$. The lack of identification arises because the structure $X_2 \Gamma' (\Sigma)^{\frac{1}{2}}_D$ also can be written as
\[
X_2 \Gamma' (\Sigma)^{\frac{1}{2}}_D = X_2^s \Xi' (\Sigma)^{\frac{1}{2}}_D \quad \text{where} \quad X_2^s = X_2 T, \quad \Xi = \Gamma (T)^{-1},
\]
$T$ is any nonsingular matrix that satisfies $(T'T)^{\frac{1}{2}}_D = I_m$. In this parameterization, $\Sigma$ becomes
\[
\Sigma = (\Sigma)^{\frac{1}{2}}_D \Lambda (\Sigma)^{\frac{1}{2}}_D \quad \text{where} \quad \Lambda = \Xi \Phi \Xi' + \Psi,
\]
and $\Phi = T'T$ is the $m \times m$ within row correlation matrix for the matrix of rotated factors $X_2^s$. If $T$ is further constrained to be an orthogonal matrix, then the transformation from $X_2$ to $X_2^s$ is an orthogonal rotation. Otherwise, the transformation is called an oblique rotation.

To identify the parameters, $T$ is usually chosen to optimize an objective criterion. In this article, $T$ is chosen to minimize a quartic criterion [6] applied to the factor loadings after Kaiser [17] row-normalization. Define the $p \times p$ diagonal matrix $\Lambda_2$ by $\Lambda_2^{\text{def}} \equiv (\Gamma')^{-1}$. The diagonal components of $\Lambda_2$, namely $\lambda_1^2, \ldots, \lambda_p^2$ are called communalities. The row-normalized rotated factor loading matrix is denoted by $\Gamma_\lambda$ and is defined as
\[
\Gamma_\lambda^{\text{def}} = \Lambda^{-1} \Gamma (T')^{-1},
\]
and satisfies $(\Gamma_\lambda \Phi \Gamma_\lambda')_D = I_p$, where $\Phi = T'T$. The family of quartic criteria applied to $\Gamma_\lambda$ can be written as
\[
Q(\Gamma_\lambda) = (\gamma_\lambda \otimes \gamma_\lambda)' W (\gamma_\lambda \otimes \gamma_\lambda), \quad \text{where} \quad \gamma_\lambda = \text{vec} \ \Gamma_\lambda,
\]
\[
W = \sum_{i=1}^4 a_i W_i, \quad W_1 = I_{m^2 2}, \quad W_2 = I_{mp,3} (\text{vec} I_m \text{vec}' I_m \otimes L_{p,22}) I_{mp,3}',
\]
\[
W_3 = I_{mp,3} (L_{m,22} \otimes \text{vec} I_p \text{vec}' I_p) I_{mp,3}', \quad W_4 = I_{mp,22}.
\]
$W_{mp,3} = I_m \otimes (I_{m,p}) \otimes I_p$, $I_{mp,22}$ is defined in (3), $\text{vec}' M = (\text{vec} M)'$, and the constants $a_1, \ldots, a_4$ are chosen to emphasize different aspects of the rotated loadings. See [6] for details and [24] for a correction.

Browne [5] described explicit parameterizations for the rotation matrix $T$, subject either to $(T'T)^{\frac{1}{2}}_D = I_m$ or to $T'T = I_m$. Implicit parameterizations were described by Boik [4]. Employing
either type of parameterization and choosing \( T \) to be the minimizer of \( Q(\Gamma_\lambda) \) imposes restrictions on \( \Gamma_\lambda \) and \( \Phi \). Denote the \( m(m-1)/2 \times 1 \) vector that contains the distinct off-diagonal components of the correlation matrix \( \Phi \) by \( \phi \). It can be shown [2, Eq. 17; 4] that if \( T \) is constrained to be an orthogonal matrix and \( \Gamma_\lambda \) minimizes \( Q(\Gamma_\lambda) \) in (25), then \( \phi = 0 \) and \( \Gamma_\lambda \) must satisfy

\[
g_1(\gamma_\lambda) = 0, \quad \text{where} \quad g_1(\gamma_\lambda) = A \left( I_{m^2} - I_{(m,m)} \right) (\Gamma'_\lambda \otimes I_m) I_{(m,p)} D_{Q;\gamma_\lambda}^{(1)},
\]

where

\[
a = \sum_{i=2}^{m} \sum_{j=1}^{m-1} e^d_h \left( e^{m}_i \otimes e^{m}_j \right)^\prime, \quad h = \frac{(i-1)(i-2)}{2} + j, \quad d = \frac{m(m-1)}{2}
\]

and

\[
D_{Q;\gamma_\lambda}^{(1)} = 4 \left( \gamma'_\lambda \otimes I_{mp} \right) N_{mp} W (\gamma_\lambda \otimes \gamma'_\lambda).
\]

Expressions for the higher-order derivatives of \( Q(\Gamma_\lambda) \) in (25) with respect to \( \gamma_\lambda \) are given in a supplement that can be downloaded from [http://www.math.montana.edu/~rboik/implicit/](http://www.math.montana.edu/~rboik/implicit/).

It was shown in [4] (also see [15, Eq. 28]) that if the less restrictive constraint, \( (T'T)_D^1 = I_m \), is imposed and \( \Gamma_\lambda \) minimizes \( Q(\Gamma_\lambda) \), then \( \Phi \) and \( \Gamma_\lambda \) must satisfy

\[
g_1(\gamma_\lambda, \phi) = 0, \quad \text{where} \quad g_1(\gamma_\lambda, \phi) = A \left[ I_{m^2} - (I_m \otimes \Phi) L_{m,22} \right] (\Gamma'_\lambda \otimes I_m) I_{(m,p)} D_{Q;\gamma_\lambda}^{(1)},
\]

where

\[
a = \sum_{i=1}^{m} \sum_{j \neq i} e^d_f \left( e^{m}_i \otimes e^{m}_j \right)^\prime, \quad f = (m-1)(i-1) + j - I(j > i),
\]

\[
d = m(m-1), \quad I(j > i) = \begin{cases} 1 & \text{if } j > i, \\ 0 & \text{otherwise}, \end{cases}
\]

\( L_{m,22} \) is defined in (3), and \( D_{Q;\gamma_\lambda}^{(1)} \) is defined in (26).

Jennrich [16] employed the method of Silvey [29] and the restrictions in (26) to obtain the asymptotic covariance matrix of estimators of the orthogonally rotated factor loadings (without row-normalization) under multivariate normality. Ogasawara [24] employed the method of Silvey [29] and restrictions equivalent to those in (27) to obtain the asymptotic covariance matrix of estimators of the oblique rotated factor loadings (with row-normalization) under multivariate normality. In this article, the results in Section 5 are illustrated by obtaining the two-term Edgeworth expansion for the distribution of maximum likelihood parameter estimators (Wishart likelihood function) when sampling from arbitrary distributions having finite cumulants. An oblique rotation of the row-normalized factor loadings is employed. Ogasawara [25] obtained the one-term Edgeworth expansion by an alternative method.

The reparameterized covariance matrix can be written as follows:

\[
\Sigma = \Sigma(\theta) = \left( \Sigma \right)^{\frac{1}{2}} A \left( \Sigma \right)^{\frac{1}{2}} \quad \text{subject to} \quad g(\gamma_\lambda, \phi) = 0, \quad \text{where} \quad g(\gamma_\lambda, \phi) = \begin{pmatrix} g_1(\gamma_\lambda, \phi) \\ g_2(\gamma_\lambda, \phi) \end{pmatrix}, \quad g_2(\gamma_\lambda, \phi) = L_{p,21} \vec{\Gamma_\lambda \Phi \Gamma_\lambda} - I_p,
\]

\[
A = \Xi \Phi \Sigma' + \Psi, \quad \Xi = \Lambda \Gamma_\lambda, \quad \Psi = I_p - \Lambda^2, \quad \theta = (\gamma'_\lambda, \phi') (\lambda'_d, \sigma_d)' \quad (28)
\]

\( \gamma_\lambda = \text{vec} \Gamma_\lambda, \lambda \) is the \( p \)-vector that contains the diagonal components of \( \Lambda \), \( \phi \) contains the distinct correlation coefficients in \( \Phi \), \( \sigma_d \) contains the \( p \) diagonal components of \( (\Sigma)^{\frac{1}{2}} \), \( I_p \) is a \( p \)-vector of
ones, \( g_1(\gamma', \phi) \) is given in (27), and \( L_{p,21} \) is defined in (3). The components of \( \lambda \) are the square roots of the communalities. The constraint \( g_2(\gamma', \phi) = 0 \) is equivalent to \( (\Gamma_2 \Phi \Gamma_2') D = I_p \) and this ensures that \( \Delta \) is a correlation matrix. All constraints in (28) are identification constraints, but this fact plays no role in the expansion. The estimator, \( \hat{\theta} \) is obtained as the maximizer of the Wishart log likelihood function

\[
L(\theta; Y) = -\frac{n}{2} \text{trace} \left( S \Sigma^{-1} \right) - \frac{n}{2} \ln |\Sigma| \quad \text{subject to} \quad g(\gamma', \phi) = 0, \quad \text{(29)}
\]

where \( n = N - r, \quad r = \text{rank}(X_1), \quad S = n^{-1}Y' \left[I_N - X_1(X_1'X_1)^{-1}X_1'\right] Y \)

is the sample covariance matrix, and \((\cdot)^{-1}\) is any generalized inverse.

The constrained log likelihood function can be maximized by employing the Newton algorithm of Theorem 3. Alternatively, a Fisher scoring algorithm can be constructed by replacing \( D_{(2)}^{L; \tau; \tau'} \)

by \( E(D_{(2)}^{L; \tau; \tau'}) = -nI_\tau \), evaluated at the current guess for \( \tau \). The average Fisher information of the identified parameter, \( \tau \), in the factor analysis model is

\[
\bar{I}_\tau = G_\tau' \bar{I}_\theta G_\tau, \quad \text{where} \quad \bar{I}_\theta = \frac{1}{2} D_{\sigma; \theta'}^{(1); p} \left( \Sigma^{-1} \sigma - \Sigma^{-1} \right) D_{\sigma; \theta'}^{(1); p}.
\]

\( G_\tau \) is defined in Theorem 2, and \( \sigma = \text{vec} \Sigma \). Detailed expressions for the required derivatives in the factor analysis model are available in the supplement. These expressions were obtained by employing the results in Appendices A and B.

The Edgeworth expansion in Corollary 5.3, requires computation of the cumulant functions \( \{\omega\}_{i=1}^4 \). These functions can be computed as described by Corollary 5.3. Alternatively, if multivariate normality is assumed, then they can be computed as described by Boik [3, Section 4]. Expressions for these quantities when sampling from normal or nonnormal distributions are given in the supplement.

A simulation study was conducted to evaluate the accuracy of the Edgeworth expansion. To obtain parameter values for the simulation, an \( m=2 \) factor structure was fit to the \( p=9 \) dimensional correlation matrix taken from [7]. A quartimin rotation (i.e., \( a_1 = 0, \quad a_2 = 1, \quad a_3 = 0, \quad a_4 = -1 \) in (25) was conducted after Kaiser row-normalization. The resulting estimators were taken to be population parameters. Specifically, the population covariance matrix, \( \Sigma \), was taken to be

\[
\Sigma = \Xi \Phi \Xi' + \Psi, \quad \text{where} \quad \Psi = I_9 - \Lambda^2,
\]

\[
\Xi = \begin{pmatrix}
0.700 & 0.052 \\
0.634 & 0.143 \\
0.593 & -0.028 \\
0.059 & 0.860 \\
-0.009 & 0.775 \\
-0.021 & 0.895 \\
0.796 & -0.038 \\
0.515 & -0.015 \\
0.897 & -0.021
\end{pmatrix}, \quad \text{diag}(\Lambda^2) = \begin{pmatrix}
0.539 \\
0.535 \\
0.333 \\
0.808 \\
0.592 \\
0.779 \\
0.597 \\
0.255 \\
0.781
\end{pmatrix}, \quad \text{and} \quad \Phi = \begin{pmatrix}
1 & 0.622 \\
0.622 & 1
\end{pmatrix}.
\]

For each of 100,000 simulation trials, a sample of size \( N = 200 \) was generated in which the components of \( X_2 \) and \( \mathbf{E} \) in (20) were iid random variables having one of four distributions, namely (a) \( N(0, 1) \), (b) \( \gamma_3^2 \) scaled to have mean zero and variance one, (c) lognormal with kurtosis \( \kappa_4 = 5 \) scaled to have mean zero and variance one, and (d) lognormal with kurtosis \( \kappa_4 = 10 \) scaled
Fig. 2. Accuracy of edgeworth expansion: lognormal factors and errors, $\kappa_4 = 5$. The upper four plots display the true pdf (T), and two approximations to the pdf of estimators of two rotated loadings, the between factor correlation, and the communality of the first variable. The two approximations are the asymptotic normal distribution (N) and the two-term Edgeworth expansion (E). The lower four plots display the difference between the true cdf and the two approximations to the cdf. The legends for the lower plots are identical to those in the upper plots.
Fig. 3. Accuracy of edgeworth expansion: lognormal factors and errors, $\kappa_4 = 10$. The format of this figure is identical to that of Fig. 2.
to have mean zero and variance one. For distributions (c) and (d), the random variables were generated as exponentials of scaled standard normal random variables. For each data set, the two factor structure was fit and the factors were rotated after Kaiser normalization. If the maximizer of the likelihood function fell outside of the parameter space, then the sample was discarded and replaced by a new sample. This occurred 5, 49, 15, and 25 times under distributions (a), (b), (c), and (d), respectively. Fig. 2 displays the accuracy of the Edgeworth expansion compared to that of the asymptotic normal approximation for estimators of four functions of the parameters, namely \( \hat{\xi}_{11} \), \( \hat{\xi}_{12} \), \( \hat{\lambda}_1^2 \), and \( \phi \) when sampling from a lognormal distribution having kurtosis \( \kappa_4 = 5 \). In each case, the true pdf was estimated by a kernel density estimator and the true cdf was estimated by the empirical cumulative distribution function. Fig. 2 demonstrates that the accuracy of the asymptotic normal approximation varies from good to poor depending on the parameter function, but the Edgeworth expansion is quite accurate for all four functions. The simulation results based on sampling from the multivariate normal and \( \chi_1^2 \) distributions, conditions (a) and (b), are not displayed because they yielded plots that are essentially identical to those in Fig. 2.

The advantages of the Edgeworth expansion over the asymptotic normal approximation will diminish as the higher-order cumulants of the underlying distribution increase. If these cumulants become too large, then the performance of the Edgeworth expansion for fixed sample size can become too large, then the performance of the Edgeworth expansion for fixed sample size can be worse than that of the normal approximation. Fig. 3 displays the accuracy of the Edgeworth expansion compared to that of the asymptotic normal approximation when sampling from a lognormal distribution having kurtosis \( \kappa_4 = 10 \). For this distribution the eighth-order standardized cumulant is \( \kappa_8 = 293,779 \). Fig. 3 shows that the performance of the Edgeworth expansion is still quite good for estimators of \( \hat{\xi}_{12} \) and \( \phi \), but not for estimators of \( \hat{\xi}_{11} \) and \( \hat{\lambda}_1^2 \).

### Appendix A. Product rules, chain rule, and Kronecker product identities

#### A.1. Product and chain rules

Let \( U \) be an \( a \times b \) matrix function of the \( p \times q \) matrix \( \Theta \) and let \( Z \) be a \( c \times d \) matrix function of \( \Theta \). Then the matrix product and Kronecker product rules are as follows:

\[
D^{(1)}_{UZ; \Theta} = D^{(1)}_{U; \Theta} (I_q \otimes Z) + (I_p \otimes U) D^{(1)}_{Z; \Theta} \quad \text{if} \quad b = c \quad \text{and}
\]

\[
D^{(1)}_{U \otimes Z; \Theta} = \left( D^{(1)}_{U; \Theta} \otimes Z \right) + (I_p \otimes I_{(c,a)}) \left( D^{(1)}_{Z; \Theta} \otimes U \right) (I_q \otimes I_{(b,d)}), \tag{A.1}
\]

where \( I_{(c,a)} \) is the commutation matrix described in [20].

Let \( Z \) be an \( a \times b \) matrix function of the elements of \( X \), where \( X \) is a matrix function of the \( p \times q \) matrix \( \Theta \). Then, the matrix chain rule is as follows:

\[
D^{(1)}_{Z; \Theta} \left( D^{(1)}_{X; \Theta} \otimes I_a \right) \left( I_q \otimes D^{(1)}_{Z;X} \right), \tag{A.2}
\]

where \( X = \text{vec} \ X \). In particular, if \( \Theta \) is a \( k \times 1 \) column vector, \( \theta \), or a \( 1 \times k \) row vector, \( \theta^t \), then

\[
D^{(1)}_{Z; \theta} = \left( D^{(1)}_{X; \theta} \otimes I_a \right) D^{(1)}_{Z;X} = \left( I_k \otimes D^{(1)}_{Z;X} \right) \left( D^{(1)}_{X; \theta} \otimes I_b \right) \quad \text{and}
\]

\[
D^{(1)}_{Z; \theta} = \left( D^{(1)}_{X; \theta} \otimes I_a \right) \left( I_k \otimes D^{(1)}_{Z;X} \right) = D^{(1)}_{Z;X} \left( D^{(1)}_{X; \theta} \otimes I_b \right),
\]

The product rules in (A.1) and the chain rule in (A.2) are verified in [13, Chapter 6].
A.2. Kronecker product identities

Below is a list of identities that are useful in computing derivatives. Identity (a) was established by Roth [27]. Identity (b) was established by Neudecker and Wansbeek [22], and identities (c) and (d) were established by MacRae [20]. Let A, B, C, D, and E be matrices, where A is $p \times q$, B is $q \times r$, C is $r \times s$, and D is $t \times u$. Then,

(a) $\text{vec } ABC = (C' \otimes A)\text{vec } B$,
(b) $\text{vec } (A \otimes C) = (I_q \otimes I_{(p,s)} \otimes I_r)(\text{vec } A \otimes \text{vec } C)$,
(c) $AB = [I_p \otimes \text{vec } B'] (\text{vec } A' \otimes I_r)$,
(d) $AB = [\text{vec } B' \otimes I_p] (I_r \otimes \text{vec } A)$,
(e) $(ABC \otimes D)E = [A \otimes \text{vec } C' \otimes D] (\text{vec } B' \otimes E)$,
(f) $E(ABC \otimes D) = [\text{vec } B' \otimes E] (C \otimes \text{vec } A \otimes D)$,
(g) $(D \otimes ABC)E = [D \otimes \text{vec } C' \otimes A] (E \otimes \text{vec } B)$

$$= I_{(p,t)} [A \otimes \text{vec } C' \otimes D] (\text{vec } B' \otimes I_{(u,s)})E),$$

(h) $E(D \otimes ABC) = [E \otimes \text{vec } B'] (D \otimes \text{vec } A' \otimes C)$

$$= [\text{vec } B' \otimes EI_{(p,t)}] (C \otimes \text{vec } A \otimes D)I_{(u,s)},$$

(i) $[A \otimes \text{vec } C' \otimes D]$

$$= [A \otimes \text{vec } I_s \otimes \text{vec } I_r \otimes D] (I_{(p,s)} \otimes I_{ru}) (\text{vec } C \otimes I_{qrs})$$

where $\text{vec } A' = (\text{vec } A)'$ and $\text{vec } A' = \text{vec}(A')$.

Appendix B. Likelihood-based expressions for $Z_i$ and $K_i$

If $L$ is a log likelihood function and the required regularity conditions are satisfied, then

$$K_1 = 0,$$

$$Z_1 = n^{-1/2} G_* D^{(1)}_{L, \theta},$$

$$K_2 = -\bar{I}_r = -G_* \bar{I}_\theta G_*,$$

$$Z_2 = n^{-1/2} D^{(2)}_{\theta; \tau, \tau'} (I_v \otimes D^{(1)}_{L, \theta}) - \sqrt{n} (\bar{J}_r - \bar{I}_r),$$

$$\bar{J}_r = G_* \bar{I}_\theta G_*,$$

$$J_\theta = -n^{-1} D^{(2)}_{L, \theta, \theta'},$$

$$K_3 = -D^{(2)}_{\theta; \tau, \tau'} (I_v \otimes \bar{I}_\theta G_*) 2N_v + G'_* K_{3, \theta} (G_* \otimes G_*) - G'_* \bar{I}_\theta D^{(2)}_{\theta; \tau, \tau'},$$

$$K_{3, \theta} = n^{-1} E \left(D^{(3)}_{L, \theta, \theta', \theta'} \right),$$

$$Z_3 = n^{-1/2} \left(I_v \otimes D^{(1)}_{L, \theta} \right) D^{(3)}_{\theta; \tau, \tau', \tau'} - D^{(2)}_{\theta; \tau, \tau'} [I_v \otimes \sqrt{n} (\bar{J}_\theta - \bar{I}_\theta) G_*] 2N_v + G'_* Z_{3, \theta} (G_* \otimes G_*) - G'_* \sqrt{n} (\bar{J}_\theta - \bar{I}_\theta) D^{(2)}_{\theta; \tau, \tau'},$$

where $\bar{I}_\theta$ is a log likelihood function and the required regularity conditions are satisfied, then
Appendix C. Supplementary data

Supplementary data associated with this article can be found in the online version at 10.1016/j.jmva.2007.01.005.

References

[4] R.J. Boik, Newton algorithms for analytic rotation: an implicit function approach, Technical Report #9-12-06, Department of Mathematical Sciences, Montana State University, Bozeman, MT.